

Practical Quantum Chemistry on Near Term Quantum Computers

Alexis Philip Ralli



Thesis submitted for the degree of
Doctor of Philosophy

Primary Supervisor: Prof. Peter Coveney
Secondary Supervisors: Prof. Peter Love,
Prof. Dan Browne

Department of Chemistry
University College London
United Kingdom

April 25, 2023

Declaration

I Alexis Ralli confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Solutions to the time-independent Schrödinger equation for molecular systems allow chemical properties to be studied without the direct need for the material. However, the dimension of this problem grows exponentially with the size of the quantum system under consideration making conventional treatment intractable. Quantum computers can efficiently represent and evolve quantum states. Their use offers a possible way to perform simulations on molecules previously impossible to model. However, given the constraints of current quantum computers even studying small systems is limited by the number of qubits, circuit depth and runtime of a chosen quantum algorithm. The work in this thesis is to explore and provide new tools to make chemical simulation more practical on near-term devices. First, the unitary partitioning measurement reduction strategy is explored. This reduces the runtime of the variational quantum eigensolver algorithm (VQE). We then apply this reduction technique to the contextual subspace method, which approximates a problem by introducing artificial symmetries based on the solution of noncontextual version of the problem that reduces the number of qubits required for simulation. We provide a modification to the original algorithm that makes an exponentially scaling part of the technique quadratic. Finally, we develop the projection-based embedding (PBE) technique to allow chemical systems to be studied using state-of-the-art classical methods in conjunction with quantum computing protocols in a multiscale hierarchy. This allows molecular problems much larger than conventionally studied on quantum hardware to be approached.

Impact Statement

In this thesis, we describe different methods that decrease the resources required to simulate a molecular system on a quantum device. The first technique reduces the overall runtime of the variational quantum eigensolver algorithm, by decreasing the number of measurements required. The second looks at how splitting a problem up into a classical and quantum part can significantly reduce the number of quantum bits (“qubits”) required to study a given problem. Finally, we introduce a chemically intuitive approach that permits a subdomain of a molecule’s electronic structure to be calculated accurately on a quantum device, while the rest of the molecule is described at a cheaper level using density functional theory (DFT) running on a conventional (classical) computer. We demonstrate that our method produces results closer to the current gold standard wavefunction method compared to DFT, for molecules that cannot be simulated fully on current quantum computers. Our algorithm is tunable, so that the size of the quantum simulation can be adjusted to run on available quantum resources, and makes no restriction on the quantum algorithm used to solve the output Hamiltonian. Therefore, our method will continue to enable increasingly large systems to be studied more accurately beyond the noisy intermediate-scale era of quantum computing. We further show that this method is particularly effective when the quantum processor is targeted at a strongly correlated region of the molecule. These three tools should enable quantum computing to achieve useful results sooner than would be possible by full-system quantum simulation. This will facilitate the study of molecules of general chemical and biochemical interest including and not limited to: binding affinities, excited states and chemical reaction rates. Overall, these tools are of societal and economic significance, as they will help facilitate the discovery and our understanding of materials.

Contents

Declaration	1
Abstract	2
Impact Statement	3
Acknowledgements	10
1 Introduction	11
1.1 Motivation	12
1.2 Thesis outline	13
2 Quantum Chemistry Background	14
2.1 The electronic structure problem	14
2.2 The Pauli exclusion principle	16
2.3 Orbitals	18
2.4 The many-electron wavefunction	18
2.5 The Hartree-Fock approximation	20
2.5.1 The Schrödinger equation and first Slater-Condon rule	21
2.5.2 Hartree-Fock integro-differential equation	23
2.5.3 Integrals	24
2.5.4 The Fock operator	26
2.5.5 Roothan equations and basis sets	28
2.6 Atomic to molecular orbitals	30
2.7 Second quantization	31
2.8 Post-Hartree-Fock methods	32
2.9 Density functional theory	34
2.9.1 Hohenberg-Kohn theorems	34
2.9.2 Kohn-Sham DFT	40
2.9.3 The Kohn-Sham equations	42
2.9.4 Self-interaction and delocalization error in DFT	43
3 Quantum Chemistry on	
Quantum Computers	45
3.1 Phase estimation	46
3.2 The variational quantum eigensolver	50
3.2.1 Fermion to qubit mapping	53
3.2.2 Ansätze	54
3.2.3 Classical optimization in VQE	55
3.3 VQE and quantum advantage	57
4 The Measurement Roadblock to Implementation of VQE	59
4.1 Measurement in quantum mechanics	59

4.1.1	Expectation variance of individual operators	61
4.1.2	Expectation variance of linear combination of operators	63
4.1.3	Standard variance in VQE	63
4.2	Prior work on measurement strategies	66
4.2.1	Partial tomography approaches	66
4.2.2	Variance reduction method	68
4.2.3	Basis rotation technique	69
4.2.4	Shadow methods	69
4.2.5	Partitioning approaches	72
4.2.6	Variance in grouped measurement strategies	80
4.3	Unitary partitioning study	86
4.3.1	Sequence of rotations approach	87
4.3.2	Linear combination of unitaries approach	88
4.3.3	Numerical study	99
4.3.4	LCU circuit optimization	105
4.4	Conclusion	109
5	The Contextual-Subspace	
	Variational Quantum Eigensolver	112
5.1	Contextuality	113
5.2	Contextual-subspace VQE	115
5.2.1	Partitioning the Hamiltonian	116
5.2.2	Solving the noncontextual Hamiltonian	118
5.2.3	Noncontextual stabilizers	122
5.2.4	Mapping to contextual subspace	124
5.2.5	Linear combination of unitaries construction of R	126
5.2.6	CS-VQE implementation	127
5.3	Numerical results	128
5.3.1	Method	128
5.3.2	Toy example	129
5.3.3	Measurement reduction	133
5.4	Conclusion	136
6	Multiscale Quantum Simulation via	
	Projection-Based Embedding	138
6.1	Density functional embeddings	138
6.2	Theory of projection-based embedding	144
6.2.1	Global calculation	145
6.2.2	Localization of molecular orbitals	146
6.2.3	Projection	148
6.3	Molecular orbital localization	151
6.3.1	Foster-Boys	151
6.3.2	Edmiston-Ruedenberg	152
6.3.3	Pipek-Mezey	152
6.3.4	Intrinsic bonding orbitals	153

CONTENTS

6.3.5	Subsystem projected AO decomposition	154
6.3.6	Partitioning into subsystems	156
6.4	Numerical study	156
6.4.1	Method	156
6.4.2	Results and discussion	158
6.5	Conclusion	163
7	Conclusions and Outlook	165
A	CS-VQE Appendix	190
A.1	Mapping Pauli operators to single-qubit Pauli Z operators	190
A.2	CS-VQE via a sequence of rotations	191
A.3	CS-VQE via a linear combination of unitaries	199
A.4	Numerical details of the toy example	205
A.5	Graphical results for CS-VQE simulation of each molecular Hamiltonian	209
A.6	Tabulated results of simulation	213
B	PBE Appendix	214
B.1	Molecular ground state energy	214
B.2	Strong correlation	214

List of Figures

3.1	Phase kickback circuit primitives	47
3.2	Quantum circuit for the first part of the phase estimation algorithm .	49
4.1	Example graph using standard notation	78
4.2	Example clique covers	78
4.3	Circuit construction to perform rotation generated by a Pauli operator	89
4.4	General circuit to block encode an operator via a linear combination of unitaries	90
4.5	Quantum circuit to disentangle an $n + 1$ qubit state	96
4.6	Quantum multiplexor circuits	97
4.7	Quantum circuit templates to implement a multicontrol single qubit Z gate with a negative or imaginary phase	99
4.8	Illustration of graph approach to finding anticommuting sets of P_i op- erators.	100
4.9	Probability density functions of single-shot VQE estimates of the ground state energy of H_2	101
4.10	Probability density function of single-shot VQE estimates of the ground state energy of LiH	102
4.11	Circuit template for $Z \cdot Z$ operation.	106
4.12	Result of recursively applying the template given in Figure 4.11 to a sequence of single qubit Z gates.	107
4.13	Circuit template of a n_c -control P , where $P \in \{I, Z\}^{\otimes 3}$	108
4.14	Circuit template to implement a multicontrol Pauli operator	108
4.15	Example of a multicontrol $\pm iZ$ gate, implemented via a multicontrol $R_z(\mp\pi)$ multiplexor.	109
4.16	Decomposition of a diagonal unitary Δ as a sequence of multicontrol R_z multiplexors [218].	109
5.1	Compatibility graphs for contextual sets of four Pauli operators . . .	116
5.2	Summary of different contextual subspace approximations for the toy problem	129
5.3	Number of Pauli operators requiring separate measurement to deter- mine the ground-state energy of a particular molecular Hamiltonian .	131
5.4	Number of qubits required to simulate different electronic structure Hamiltonians	133

LIST OF FIGURES

6.1	Planar representations of the molecules used in embedding calculations	157
6.2	Results for WF-in-DFT calculation of small molecules, where the SPADE localisation method has been used	159
6.3	Potential energy curve PBE results for H ₂ O, with changing OH bond length	162
A.1	Summary of different molecular Hamiltonians under different CS-VQE approximations	212
B.1	Results for embedding of small molecules using the IBO localisation method	216
B.2	IBO PBE potential energy curve for water, with changing active bond under a lower active threshold setting	219

List of Tables

2.1	Table of atomic units in the international system of units	15
3.1	Summary of different experimental realizations of VQE.	52
3.2	Summary of different works that encode a second-quantized fermionic Hamiltonian in qubits	53
4.1	Graph based colouring strategies to partition a Hamiltonian into cliques	79
4.2	Statistical summary of ground state calculations for H ₂ and LiH . . .	104
5.1	Example Peres-Mermin square of nine possible observables for a physical system, where each measurement can be assigned a ± 1 value. . .	114
5.2	Example Peres-Mermin square of nine Hermitian operators, all with ± 1 eigenvalues - representing observables.	114
5.3	Gate requirements to implement R as a sequence of rotations for different molecular Hamiltonians	134
6.1	A brief history of different DFT embedding methods.	139
A.1	Different contextual subspace Hamiltonians derived from toy example under updated implementation	207
A.2	Different contextual subspace Hamiltonians derived from toy example, under old implementation	208
A.3	Different resource requirements to study different electronic structure Hamiltonians required to achieve chemical accuracy	213
B.1	Full-system reference values for embedding calculations of small molecules	214
B.2	Numerical values for IBO PBE calculations	215
B.3	Numerical values for SPADE PBE calculations	215
B.4	Cartesian coordinates of atoms in water	217
B.5	Numerical values of SPADE PBE calculation of water with active region bond length changing	217
B.6	Numerical values of SPADE PBE calculation of water with active region bond length fixed	217
B.7	Numerical values of IBO PBE calculation of water with active region bond length changing	218
B.8	Numerical values of IBO PBE calculation of water with active region bond length fixed	218

Acknowledgements

I would like to express my deepest appreciation to my main and secondary supervisors, Peter V. Coveney, Peter J. Love, and Dan E. Browne for their supervision during this project. Each contributed their own insightful perspective on this work. In particular, I appreciated the weekly discussions with both Peters. I'm also extremely grateful to the others who joined these meetings: William Kirby, Andrew Tranter and Tim Weaving. Everyone really pushed my work in interesting directions and helped me learn a great deal along the way. Their constructive criticism and patience were valuable - I don't think I will mistake 'symmetry generators' for a 'symmetry basis' anytime soon.

Over the project, I was lucky to be able to visit Tufts in 2022. I'm thankful to Peter Love for hosting me there. Even though it was during the mask mandate in Cambridge MA, I appreciate all the lunches at Semolina and for Will Kirby meeting me in the office to work in person. I gained valuable experience working alongside Will and had many interesting discussions - in particular on DFT and block encodings techniques.

I also had the great pleasure of working with Tim Weaving on 'Symmer', an open-source Python library we co-produced during the final year and a half of my studies. It has been an extremely interesting and fun project and I hope we will continue to have constructive discussions on where to take our code next.

Similarly, to Michael I. Williams de la Bastida I would like to thank his support on our code 'Nbed'. The project has been very interesting and I learnt many good software development practices from Michael.

In London, I would like to thank all the members of the Centre for Doctoral Training in Delivering Quantum Technologies and the Centre of Computational Science. In particular, James Dborin, Jon McCullough and Oscar Higgott who all put up with my incessant questions. They really helped me out at the beginning of my studies.

Outside of work, my appreciation goes to all my friends. As few are likely to read this, I think I am safe to name a few outright. I would like to thank Daniel Arme, Harry Slade, James Waterfield, Jordanne Armstrong, Sebastian Fomin. From the pints, cycling, boardgames, and holidays - you have all made my time as a doctoral student extremely enjoyable.

To Antara Bate, I'd like to thank you for all your love and support. You really helped me in the final months to get everything done.

The final word, of course, must go to my family. I'd like to thank them for their unwavering support and guidance.

Chapter 1

Introduction

In the early twentieth century, physics was undergoing a series of problems. Old theories were predicting unrealistic results. The “ultraviolet catastrophe” is an archetype of this [1], where the radiation rate of a black body (idealized object which absorbs and emits all frequencies) would go to infinity as the electromagnetic wavelength went to zero. Planck solved the problem by postulating that electromagnetic energy was emitted in quanta. These problems in so-called *classical physics* were explained with a new theory known as quantum mechanics. An extremely important tool, which has had huge success explaining the world around us - from the fundamental particles of nature to reactions occurring in our sun. We recommend an interesting review by Gearhart that gives a historical context of Planck’s achievement [2].

Quantum mechanics at its core is a mathematical framework that allows physical theories to be constructed. Unlike special relativity, which can be derived from physically motivated assumptions that the laws of physics should be consistent in all inertial frames and the speed of light is constant, Quantum mechanics cannot be derived in the same way. Instead, quantum mechanics is derived from a set of mathematical postulates. These postulates are motivated by experimental observations over the last century, rather than from any physical assumptions. It is currently an open question whether a simple set of physical principles would allow the theory to be derived.

However, utilizing this mathematical framework has allowed huge leaps forward in science. For example in the physical theory of quantum electrodynamics (QED), which describes the interaction of atoms and light, has proved to be one of the most accurate theories of our world. For instance, QED predicts the electron magnetic dipole moment g_e that matches closely with experiment, with an accuracy of 7.6×10^{-13} [3–5].

In this thesis, we are interested in the application of quantum mechanics to chemistry - commonly known as quantum chemistry. In 1929 Paul Dirac famously stated [6]:

“The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is

only that the exact application of these laws leads to equations much too complicated to be soluble.”

He is describing how the whole of chemistry can be explained using quantum mechanics (QM), but the mathematical equations encountered end up being too complex to solve. The details for why will be explained in Chapter 2. However, it is worth noting that the details of all chemical systems can in theory be determined from first principles and so solved given a sufficiently powerful computer. The class of first principle methods which aim to solve such problems are referred to as *ab initio* methods.

1.1 Motivation

The focus of this thesis is *ab initio* or ‘from the beginning’ computational methods that attempt to solve the Schrödinger equation. At the core of all conventional *ab initio* techniques is the fundamental problem of the exponential computational cost of simulating quantum systems exactly. In 1982 Richard Feynman addressed this issue by proposing that the best way to simulate quantum systems is by utilizing a quantum system [7]. A quantum simulator is simply a controllable quantum system that can be used to replicate other quantum systems [8]. Many experimental devices are analogue quantum simulators, where the Hamiltonian of the device is engineered to approximate the problem Hamiltonian [9]. Another paradigm used is the digital quantum simulator proposed by Lloyd [10]. This uses Trotter decompositions to map universal unitary evolution to a circuit, which can be made arbitrarily accurate with error correction. This in effect is a universal quantum computer.

Quantum simulators are inherently quantum themselves and thus able to efficiently simulate quantum systems. The underlying reason why digital quantum simulators have an advantage over classical simulators is the difference in the number of bits required to store the wavefunction of an N -particle quantum system. On a classical computer, the number of classical bits required grows exponentially and so they cannot be used to simulate a quantum system efficiently. On the other hand, for a quantum simulator, the number of quantum bits or “qubits” required to represent an N -particle quantum system scales linearly [11]. A quantum state can consequently be represented more efficiently on a quantum computer [12]. However, it is important to note that the amplitudes of the wavefunction stored on such a device cannot be accessed efficiently.

Overall, the motivation for this work comes from the fact that quantum computers can represent and evolve quantum states much more efficiently than conventional computers [13]. This may allow new insights into systems previously not possible to simulate, facilitating the discovery of new materials [14], drugs [13] and catalysts [15]. However, given the current constraints on quantum devices useful application in these

fields remains to be seen [16]. The goal of the work in this thesis was to study and create tools to bridge this gap.

1.2 Thesis outline

In this thesis, Chapter 2 will give the basic theoretical background required to understand the underlying quantum chemistry. Chapter 3 then provides an overview of quantum algorithms related to chemistry - in particular the phase estimation and variational quantum eigensolver algorithms. The author's contribution to the field will be discussed in Chapters 4, 5 and 6. Chapter 7 then summarises the work and proposes future directions.

Chapter 2

Quantum Chemistry Background

2.1 The electronic structure problem

The Schrödinger equation gives the quantum mechanical description of any system. The time-independent non-relativistic Schrödinger equation is written as $H|\psi\rangle = E|\psi\rangle$ and solving it is one of the fundamental goals of quantum chemistry. H is the Hamiltonian operator and $|\psi\rangle$ a set of eigenstates of the Hamiltonian. Each eigenstate $|\psi_i\rangle$ yields an associated eigenvalue E_i . The physical system being studied defines the form of the Hamiltonian. Well-known examples are the particle in a box and harmonic oscillator, where the Schrödinger equation can be solved exactly.

In this thesis, the Hamiltonian of interest is the molecular Hamiltonian, which for η electrons and M nuclei is defined as [17]:

$$H = -\frac{\hbar}{2m_e} \sum_{i=0}^{\eta-1} \nabla_i^2 - \sum_{A=0}^{M-1} \frac{\hbar^2}{2M_A} \nabla_A^2 - \frac{e^2}{4\pi\epsilon_0} \sum_{i,A} \frac{Z_A}{|\vec{r}_i - \vec{R}_A|} + \frac{e^2}{4\pi\epsilon_0} \sum_{i \neq j} \frac{1}{|\vec{r}_i - \vec{r}_j|} + \frac{e^2}{4\pi\epsilon_0} \sum_{A \neq B} \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|}. \quad (2.1)$$

M_A and Z_A define the position, mass and atomic number of the A^{th} nucleus. \vec{r}_i and \vec{R}_A are the spatial coordinates (x, y, z) of the i^{th} electron and A^{th} nucleon respectively.

In atomic units the Hamiltonian becomes [18]:

$$H = -\frac{1}{2} \sum_{i=0}^{\eta-1} \nabla_i^2 - \sum_{A=0}^{M-1} \frac{\nabla_A^2}{2M_A} - \sum_{i=0}^{\eta-1} \sum_{A=0}^{M-1} \frac{Z_A}{|\vec{r}_i - \vec{R}_A|} + \sum_{i=0}^{\eta-1} \sum_{j>i}^{\eta-1} \frac{1}{|\vec{r}_i - \vec{r}_j|} + \sum_{A=0}^{M-1} \sum_{B>A}^{M-1} \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|}. \quad (2.2)$$

Table 2.1 gives values of different quantities in atomic units. The unit of length is the Bohr (a_0), the unit of mass is the electron mass (m_e) and the unit of energy is Hartree (Ha). The Laplacian operators ∇_i^2 and ∇_A^2 are second-order differential operators with respect to the i^{th} electron and A^{th} nucleon. The first and second terms in equation 2.2 are operators for the kinetic energy of electrons and nuclei

2.1. THE ELECTRONIC STRUCTURE PROBLEM

Quantity	Atomic Unit	SI value	Symbol
mass	electron mass	9.1091×10^{-31} Kg	m_e
charge	proton charge	1.6021×10^{19} C	e
action	Planck's constant $/2\pi$	1.0545×10^{-34} Js	\hbar
permittivity	$4\pi \times$ vacuum permittivity	1.1127×10^{-10} C ² N ⁻¹ m ⁻²	$4\pi\epsilon_0$
length	radius of 1st Bohr orbit	0.52917×10^{-11} m	a_0
energy	$2 \times$ Ionisation energy of H	4.3594×10^{-18} J	E_h or Ha

Table 2.1: Table of atomic units in the international system of units (commonly known as SI units) [19].

respectively; the third term gives the Coulombic attraction between electrons and nuclei; the repulsion between electrons and nuclei is given by the final two terms.

In order to reduce the complexity of this problem the Born-Oppenheimer approximation [20] can be employed, where the motion of electrons and nuclei are decoupled. The motivation for this is nucleons being over a thousand times heavier than electrons, hence the nuclei can be treated as stationary while solving for the electronic motion. What this means is that, to a good approximation, one considers the electrons moving in a field of fixed nuclei. The consequence of this is the kinetic energy of the nuclei can be neglected and the repulsion between nuclei is assumed constant. Under these assumptions equation 2.2 can be written as [18]:

$$H_e = -\frac{1}{2} \sum_{i=0}^{\eta-1} \nabla_i^2 - \sum_{i=0}^{\eta-1} \sum_{A=0}^{M-1} \frac{Z_A}{|\vec{r}_i - \vec{R}_A|} + \sum_{i=0}^{\eta-1} \sum_{j>i}^{\eta-1} \frac{1}{|\vec{r}_i - \vec{r}_j|}. \quad (2.3)$$

This is known as the Born–Oppenheimer or clamped-nuclei approximation. The resulting Hamiltonian is known as the “electronic” Hamiltonian and is solved for a given set of nuclear configurations (positions of atoms). To find the total energy for a fixed system the constant nuclear repulsion energy must be added to the solution of the electronic Hamiltonian [18]:

$$E_{total} = E_{elec} + \sum_{A=0}^{M-1} \sum_{B>A}^{M-1} \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|}. \quad (2.4)$$

Equations 2.3 and 2.4 define the electronic structure problem.

Writing out the electronic Schrödinger equation under the Born-Oppenheimer approximation

$$\left[-\frac{1}{2} \sum_i \nabla_i^2 - \sum_{A,i} \frac{Z_A}{|\vec{r}_i - \vec{R}_A|} + \sum_{A>B} \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} + \sum_{i>j} \frac{1}{|\vec{r}_i - \vec{r}_j|} \right] \Psi(\vec{r}; \vec{\mathbf{R}}) = E_{el} \Psi(\vec{r}; \vec{\mathbf{R}}), \quad (2.5)$$

which we can write in short as

$$\left[\hat{T}_e(\vec{\mathbf{r}}) + \hat{V}_{eN}(\vec{\mathbf{r}}; \vec{\mathbf{R}}) + \hat{V}_{NN}(\vec{\mathbf{R}}) + \hat{V}_{ee}(\vec{\mathbf{r}}) \right] \Psi(\vec{\mathbf{r}}; \vec{\mathbf{R}}) = E_{el} \Psi(\vec{\mathbf{r}}; \vec{\mathbf{R}}). \quad (2.6)$$

Here the bold $\vec{\mathbf{r}}$ and $\vec{\mathbf{R}}$ denote objects containing the positions of all electrons and nuclei respectively. Note the semicolon in $\Psi(\vec{\mathbf{r}}; \vec{\mathbf{R}})$ denotes a parametric dependence on $\vec{\mathbf{R}}$. By finding the energy eigenstates $|\psi_i\rangle$ and corresponding energy eigenvalues E_i of the electronic Hamiltonian H_e at different nuclear configurations one solves for nuclear motion - this provides a potential energy surface.

Potential energy surfaces $E_{elec}(\vec{\mathbf{R}})$ are of fundamental importance in chemistry as minima correspond to stable chemical structures and first-order saddle points to transition states. The energy difference between reactant minima and the transition state gives the energy barrier to reaction, hence reaction pathways can be mapped out. These surfaces are also used to find vibrational frequencies of molecular structures at stationary points (minimum or saddle point). To achieve this, the quantum harmonic/anharmonic oscillator approximation is solved yielding vibrational states (eigenvectors) and vibronic energy levels (eigenvalues).

The ground state wavefunction $\Psi(\vec{\mathbf{r}}; \vec{\mathbf{R}})$ also contains a lot of important information about molecular properties such as dipole/multipole moments and polarizability.

Unless stated, from here we only consider the electronic Hamiltonian H_e and electronic wavefunctions and we drop the subscript e .

2.2 The Pauli exclusion principle

In the standard model elementary particles, such as the electron, carry intrinsic angular momentum known as spin. The electronic Hamiltonian defined in equation 2.3 only depends on the spatial coordinates of electrons in a given molecular system and has no mention of spin. To completely describe the electrons in such systems each electron spin must be specified. To do this in the presented nonrelativistic theory it is sufficient to define two functions $\alpha(\omega)$ and $\beta(\omega)$. The first representing spin-up and the latter spin-down. Each function depends on an unspecified spin variable ω [18]. These spin functions are complete:

$$\int d\omega \alpha^*(\omega) \alpha(\omega) = \int d\omega \beta^*(\omega) \beta(\omega) = 1, \quad (2.7a)$$

$$\langle \alpha | \alpha \rangle = \langle \beta | \beta \rangle = 1, \quad (2.7b)$$

and orthonormal:

$$\int d\omega \alpha^*(\omega)\beta(\omega) = \int d\omega \beta^*(\omega)\alpha(\omega) = 0, \quad (2.8a)$$

$$\langle \alpha | \beta \rangle = \langle \beta | \alpha \rangle = 0. \quad (2.8b)$$

In this formalism, each electron i is described by 3 spatial coordinates \vec{r}_i and one spin coordinate ω_i . These are collectively described as

$$\vec{x}_i = \{\vec{r}_i, \omega_i\}. \quad (2.9)$$

The wavefunction for an η -electron system is written as $\Psi(\vec{x}_0, \vec{x}_1, \vec{x}_2, \dots, \vec{x}_{\eta-1})$.

Currently, we have labelled each electron in the wavefunction as \vec{x}_i to describe the coordinates of electron i . Electrons are fermions and thus indistinguishable from one another. Employing the Born interpretation of the wavefunction, the probability density function (distribution) of the electrons shouldn't depend on how we label them. An example of this is

$$|\Psi(\vec{x}_0, \vec{x}_1, \vec{x}_2, \dots, \vec{x}_{\eta-1})|^2 = |\Psi(\underbrace{\vec{x}_1, \vec{x}_0}_{\text{interchanged}}, \vec{x}_2, \dots, \vec{x}_{\eta-1})|^2, \quad (2.10)$$

where the probability density functions are the same even though the electron \vec{x}_0 and \vec{x}_1 have been swapped. In fact, any of the $N!$ permutations of electron coordinates leave the probability density unchanged. If we let \hat{P} be a particular permutation, then

$$|\Psi|^2 = \hat{P}|\Psi|^2. \quad (2.11)$$

However, the goal is to find the solution to the Schrödinger equation: Ψ , rather than $|\Psi|^2$ [19]. As it stands, the procedure of including the property of spin in our model is not satisfactory as the Hamiltonian operator makes no reference to spin. To properly account for spin, the antisymmetry or Pauli exclusion principle is used - an independent postulate of quantum mechanics that states if two fermions (such as electrons) exchange variables then the sign of the wavefunction must change. The explicit result of this is as follows [18, 19]:

$$\Psi(\dots, \vec{x}_i, \dots, \vec{x}_j, \dots) = -\Psi(\dots, \vec{x}_j, \dots, \vec{x}_i, \dots) \quad \forall i \neq j. \quad (2.12)$$

For a general permutation, we find:

$$P\Psi = (-1)^p\Psi, \quad (2.13)$$

where p is the parity of the permutation (the number of simple interchanges to which the permutation can be reduced too) [19]. The exact wavefunction of a molecular system must also obey this property, as well as satisfying the Schrödinger equation. This requirement is achieved through the use of Slater determinants of orbitals.

2.3 Orbitals

An orbital is defined as the wavefunction for a single particle [18]. In this thesis we only consider electron orbitals. A spatial orbital $\psi_i(\vec{r})$ describes the spatial distribution of an electron as a function of the position vector \vec{r} , where $|\psi_i(\vec{r})|^2 d\vec{r}$ is the probability of finding an electron in the small volume element $d\vec{r}$ surrounding \vec{r} . Usually, the spatial molecular orbitals will form an orthonormal set:

$$\int d\vec{r} \psi_i^* \psi_j = \delta_{ij}. \quad (2.14)$$

If the set of spatial molecular orbitals $\{\psi_i\}$ were complete then any arbitrary function $f(\vec{r})$ over space could be expanded as [18]:

$$f(\vec{r}) = \sum_k^{\infty} a_k \psi_k(\vec{r}), \quad (2.15)$$

where a_k are coefficients. In general, this set would need to be infinite to be complete, in practice this can never be done and only a finite set $\{\psi_i | i = 0, 1, \dots, K-1\}$ of K orbitals are used. This finite set will only span a certain region of the complete set, but results can be described as “exact” within the subspace spanned by the finite set.

From each spatial molecular orbital, two spin molecular orbitals can be defined by multiplying the spatial orbital by a $\alpha(\omega)$ or $\beta(\omega)$ spin function (Equation 2.7a). This is written as:

$$\left. \begin{aligned} \chi_{2K+1}(\vec{x}) &= \psi_{2K+1}(\vec{r})\alpha(\omega) = \psi_{2K+1}(\vec{r}) \\ \chi_{2K}(\vec{x}) &= \psi_{2K}(\vec{r})\beta(\omega) = \bar{\psi}_{2K}(\vec{r}) \end{aligned} \right\} i = 0, 1, \dots, K-1. \quad (2.16)$$

Therefore, given a finite set of K spatial orbitals $\{\psi_i | i = 0, 1, \dots, K-1\}$ one can form a set of $2K$ spin orbitals $\{\chi_i | i = 0, 1, \dots, 2K-1\}$.

Now the spin orbitals have been defined, we can define the many-electron wavefunction.

2.4 The many-electron wavefunction

As discussed in Section 2.3 the wavefunction describing a single electron is a spin orbital $\chi_i(\vec{x})$. Next, we want to consider wavefunctions of many electrons. If we

2.4. THE MANY-ELECTRON WAVEFUNCTION

simply take a product of spin orbital wavefunctions for all N electrons, then the N -electron wavefunction would be:

$$\Psi^{HP}(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) = \chi_i(\vec{x}_0)\chi_j(\vec{x}_1) \dots \chi_k(\vec{x}_{\eta-1}), \quad (2.17)$$

and is known as the Hartree product (HP) [18]. Here electron zero is described by spin orbital χ_i and the other electrons by the same process. There are two major problems with this formalism. First this wavefunction distinguishes electrons from each other. For example, electron one is in spin orbital χ_j and electron $\eta - 1$ is in spin orbital χ_k . However, electrons are indistinguishable. The antisymmetry principle does not distinguish electrons and only requires the electronic wavefunction to change sign when the space and spin coordinates of any two electrons change. Second, the current form of the wavefunction (equation 2.17) is uncorrelated. By this, we mean that the probability of finding electron zero in volume element $d\vec{x}_0$, electron one in volume element $d\vec{x}_1$, etc, is just equal to the product of their probabilities. This means that the motion of electrons are independent [18]:

$$|\Psi^{HP}(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1})|^2 d\vec{x}_0 d\vec{x}_1 \dots d\vec{x}_{\eta-1} = |\chi_i(\vec{x}_0)|^2 d\vec{x}_0 \cdot |\chi_j(\vec{x}_1)|^2 d\vec{x}_1 \cdot \dots \cdot |\chi_k(\vec{x}_{\eta-1})|^2 d\vec{x}_{\eta-1}. \quad (2.18)$$

Thus the probability of finding electron zero in volume element $d\vec{x}_0$ is unaffected by the positions of the other electrons and visa versa for all the other electrons. However, it is well known that electrons repel each other via the Coulomb interaction and thus electron zero will “avoid” regions of space occupied by other electrons. This means the motion of electrons is correlated [18].

To correct these deficiencies the following is done. First to enforce the antisymmetry principle a Slater determinant of spin orbitals is used [18, 21]:

$$\Phi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) = \frac{1}{\sqrt{(\eta!)}} \begin{vmatrix} \chi_i(\vec{x}_0) & \chi_j(\vec{x}_0) & \dots & \chi_k(\vec{x}_0) \\ \chi_i(\vec{x}_1) & \chi_j(\vec{x}_1) & \dots & \chi_k(\vec{x}_1) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(\vec{x}_{\eta-1}) & \chi_j(\vec{x}_{\eta-1}) & \dots & \chi_k(\vec{x}_{\eta-1}) \end{vmatrix}, \quad (2.19)$$

where $(\eta!)^{-1/2}$ is a normalization factor. This Slater determinant has η electrons occupying spin orbitals $(\chi_i, \chi_j, \dots, \chi_k)$ without specifying which electron is in which orbital. In the Slater determinant, rows are given by electron labels and columns by spin orbital labels. Interchanging the coordinates of two electrons:

$$\Phi(\dots, \vec{x}_v, \dots, \vec{x}_w, \dots) = -\Phi(\dots, \vec{x}_w, \dots, \vec{x}_v, \dots) \quad (2.20)$$

corresponds to interchanging the two rows (determined by the electron label) of the Slater determinant, which changes the sign of the determinant. Slater determinants of spin orbitals thus result in wavefunctions that obey the antisymmetry principle. Furthermore, if two electrons occupy the same orbital (aka two columns of the Slater determinant or equal) then the determinant is zero. Thus a unique spin orbital can only be occupied by one-electron satisfying the Pauli exclusion principle.

A convenient shorthand notation for a normalized Slater determinant is to only show the diagonal elements of the determinant:

$$\Phi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) = |\chi_i(\vec{x}_0) \chi_j(\vec{x}_1) \dots \chi_k(\vec{x}_{\eta-1})|. \quad (2.21)$$

By ordering electrons labels in order $(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1})$ this can be further shortened to:

$$\Phi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) = |\chi_i \chi_j \dots \chi_k|. \quad (2.22)$$

Another useful way of expressing Slater determinants is via the antisymmetrizer operator A [19]:

$$\Phi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) = \sqrt{\eta!} A \chi_0(\vec{x}_0) \chi_1(\vec{x}_1) \dots \chi_{\eta-1}(\vec{x}_{\eta-1}), \quad (2.23)$$

that gives a shorthand way of writing out the whole determinant. Here $A = \frac{1}{\eta!} \sum_{P \in S_\eta} (-1)^p P$, which is a summation of all permutation operators that exist for an ordered sequence of η objects (here electrons), where odd permutations are negative ($p = 1$) and even permutations are positive ($p = 0$) [19, 22].

To include the effect of correlation, the many-electron wavefunction is approximated by a linear combination of Slater determinants:

$$\Psi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) \approx \sum_{q=0}^Q D_q \Phi_q(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}). \quad (2.24)$$

Note we use a capital phi to denote a Slater determinant and capital psi to denote any normalized linear combination of Slater determinants. As will be discussed later, the Hartree-Fock model attempts to find the optimal expansion for $Q = 0$ - aka a single determinant. Møller-Plesset, Coupled-Cluster and Configuration Interaction methods improve upon Hartree-Fock by extending this expansion using virtual orbitals [19].

2.5 The Hartree-Fock approximation

The Hartree-Fock approximation is a very important technique in quantum chemistry and is equivalent to the molecular orbital approximation. It is the simple picture that

all undergraduate chemists are taught. Taking the electronic Hamiltonian (Equation 2.3), we can re-write it as follows:

$$H = H(\vec{r}) = H(\vec{r}_0, \vec{r}_1, \dots, \vec{r}_{\eta-1}) = \sum_{i=0}^{\eta-1} \hat{h}(\vec{r}_i) + \sum_{i=0}^{\eta-1} \sum_{j>i}^{\eta-1} g(\vec{r}_i, \vec{r}_j), \quad (2.25)$$

where

$$\hat{h}(\vec{r}_i) = \hat{h}(i) = -\frac{1}{2} \nabla_{\vec{r}_i}^2 - \sum_A^M \frac{Z_A}{r_{iA}}, \quad (2.26a)$$

$$g(\vec{r}_i, \vec{r}_j) = g(i, j) = \frac{1}{r_{ij}} = \frac{1}{|\vec{r}_i - \vec{r}_j|}. \quad (2.26b)$$

The one-electron operator $\hat{h}(\vec{r}_i)$ (sometimes called the one-electron Hamiltonian) contains the (Laplacian) kinetic energy operator and the Coulombic attraction between all the electrons and nuclei. The $\frac{1}{r_{ij}}$ term is the Coulomb repulsion between all distinct pairs ($i \neq j$) of electrons [19].

The associated Schrödinger equation is [19]:

$$H(\vec{r}_0, \vec{r}_1, \dots, \vec{r}_{\eta-1}) \Psi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) = E \Psi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}). \quad (2.27)$$

Note the many-electron wavefunction depends on both spatial and spin coordinates ($\{\vec{x}_i | i = 0, 1, \dots, \eta-1\}$), whereas the Hamiltonian only depends on spatial coordinates ($\{\vec{r}_i | i = 0, 1, \dots, \eta-1\}$) [19]. The importance of this will be seen in Section 2.5.3, when certain integrals are calculated.

The Hartree-Fock method, employing the variational principle in quantum mechanics, seeks to minimize the electronic energy defined as [18]:

$$E_0 = \langle \Psi_0 | H | \Psi_0 \rangle, \quad (2.28)$$

where $|\Psi_0\rangle$ is constructed from a single determinant of spin molecular orbitals [18]:

$$|\Psi_0\rangle = |\chi_0, \chi_1, \dots, \chi_{\eta-1}\rangle. \quad (2.29)$$

As will be seen in Section 2.5.5, the molecular spin orbitals ($\{\chi_i\}$) are varied to achieve this with the only constraint being that they must remain orthonormal $\langle \chi_i | \chi_j \rangle = \delta_{ij}$ [18]. This constraint allows us to define the Hartree-Fock procedure.

2.5.1 The Schrödinger equation and first Slater-Condon rule

Thus far we have defined the Molecular Hamiltonian (equation 2.25) and an approximate form for an η -electron wavefunction - a single Determinant (equation 2.29). The

time-independent Schrödinger equation states:

$$H |\Psi_0\rangle = E |\Psi_0\rangle, \quad (2.30)$$

where the wavefunction is a Slater determinant and the Hamiltonian is a set of operators acting on the three-dimensional space of each electron. Left multiplying by $\langle\Psi_0|$ gives:

$$\langle\Psi_0| H |\Psi_0\rangle = E \langle\Psi_0|\Psi_0\rangle. \quad (2.31)$$

Assuming the spin orbitals are orthonormal, then the Slater determinant is normalized ($\langle\Psi_0|\Psi_0\rangle = 1$) yielding:

$$\langle\Psi_0| H |\Psi_0\rangle = E. \quad (2.32)$$

The form of this equation hides a lot of complexity. In detail, this is an integro-differential equation, explicitly:

$$E = \int \sqrt{\eta!} A \chi_0^*(\vec{x}_0) \chi_1^*(\vec{x}_1) \dots \chi_{\eta-1}^*(\vec{x}_{\eta-1}) \left[\sum_{i=0}^{\eta-1} \hat{h}(\vec{r}_i) + \sum_{i=0}^{\eta-1} \sum_{j>i}^{\eta-1} g(\vec{r}_i, \vec{r}_j) \right] \sqrt{\eta!} A \chi_0(\vec{x}_0) \chi_1(\vec{x}_1) \dots \chi_{\eta-1}(\vec{x}_{\eta-1}) d\vec{x}_0 d\vec{x}_1 \dots d\vec{x}_{\eta-1}. \quad (2.33)$$

Here the antisymmetrizer A produces all linear combinations of products of spin orbitals. The sums over different electrons can also be expanded and separated into separate integrals according to the rules:

1. $\int W + V = \int W + \int V$
2. $\int \int f(\vec{r}_1) q(\vec{r}_2) d\vec{r}_1 d\vec{r}_2 = \int f(\vec{r}_1) d\vec{r}_1 \cdot \int q(\vec{r}_2) d\vec{r}_2$

Essentially, the energy expression is a complex linear combination of products of simple integrals over spin orbitals and operators. These simplify to the following:

$$E = \sum_i \langle \chi_i(\vec{x}_i) | \hat{h}(i) | \chi_i(\vec{x}_i) \rangle + \sum_{i<j} \left[\langle \chi_i(\vec{x}_i) \chi_j(\vec{x}_j) | \hat{g}(i, j) | \chi_i(\vec{x}_i) \chi_j(\vec{x}_j) \rangle - \langle \chi_i(\vec{x}_i) \chi_j(\vec{x}_j) | \hat{g}(i, j) | \chi_j(\vec{x}_j) \chi_i(\vec{x}_i) \rangle \right], \quad (2.34)$$

using the Slater-Condon rules [21, 23]. These allow integrals of one and two-body operators over wavefunctions constructed as Slater determinants of orthonormal orbitals to be determined in terms of the individual Slater determinants. We omit the derivation of these rules, which can be found in the original work [21] and many quantum

chemistry textbooks. We recommend an interested reading to look at Appendix M in [22]. These rules allow equation 2.34 to be evaluated. The next goal is to show how to minimize the Hartree-Fock energy with respect to the spin orbitals.

2.5.2 Hartree-Fock integro-differential equation

To determine the best (Hartree-Fock) spin orbitals, the following integro-differential equation must be solved [18]:

$$\begin{aligned} \hat{h}(\vec{x}_1)\chi_i(\vec{x}_1) + \sum_{j \neq i} \left[\int d\vec{x}_2 |\chi_j(\vec{x}_2)|^2 \frac{1}{r_{12}} \right] \chi_i(\vec{x}_1) - \sum_{j \neq i} \left[\int d\vec{x}_2 \chi_j^*(\vec{x}_2) \chi_i(\vec{x}_2) \frac{1}{r_{12}} \right] \chi_j(\vec{x}_1) \\ = \epsilon_i \chi_i(\vec{x}_1), \end{aligned} \quad (2.35)$$

where ϵ_i gives the energy of spin orbital χ_i .

In equation 2.35, the first square bracket term is known as the Coulomb term v_i^{coul} . In an exact theory, this interaction is represented by the two-electron operator r_{ij}^{-1} [18]. In Hartree-Fock theory the one-electron Coulomb potential is given by [18]:

$$v_i^{\text{coul}} = \sum_{j \neq i} \left[\int d\vec{x}_2 |\chi_j(\vec{x}_2)|^2 \frac{1}{r_{12}} \right]. \quad (2.36)$$

To interpret this equation, suppose electron 2 occupies spin orbital χ_j . The $d\vec{x}_2 |\chi_j(\vec{x}_2)|^2$ term gives the probability electron 2 occupies the small volume element $d\vec{x}_2$ at \vec{x}_2 . By integrating over all space and spin coordinates, this equation gives the average interaction r_{12}^{-1} of electron 1 and 2. Note this is a one-electron potential. This replaces the true two-electron potential r_{12}^{-1} , which would be the instantaneous repulsion felt by electron 1 and due to electron 2 at the position of electron 2 (aka no averaging over all space has been done). Note the sum over all electrons $j \neq i$ gives the total average potential acting on the electron in spin orbital χ_i given from all the other $\eta - 1$ electrons in the other spin orbitals. With this interpretation it is standard to define the Coulomb operator:

$$J_{\chi_j}(\vec{x}_1) = \int d\vec{x}_2 |\chi_j(\vec{x}_2)|^2 \frac{1}{r_{12}}, \quad (2.37)$$

which is the average local potential at \vec{x}_1 due to the electron in χ_j [18].

The final square bracket term in equation 2.35 is the exchange term. This is defined by its effect when operating on a particular spin orbital χ_i :

$$K_{\chi_j}(\vec{x}_1)\underline{\chi_i(\vec{x}_1)} = \left[\int d\vec{x}_2 \chi_j^*(\vec{x}_2) \frac{1}{r_{12}} \underline{\chi_i(\vec{x}_2)} \right] \underline{\chi_j(\vec{x}_1)}. \quad (2.38)$$

In contrast, the action of J_{χ_j} on $\chi_i(\vec{x}_1)$ is:

$$J_{\chi_j}(\vec{x}_1)\underline{\chi_i(\vec{x}_1)} = \left[\int d\vec{x}_2 \chi_j^*(\vec{x}_2) \frac{1}{r_{12}} \chi_j(\vec{x}_2) \right] \underline{\chi_i(\vec{x}_1)}. \quad (2.39)$$

We highlight the different actions of these two operators by underlining certain parts of the equations. Note how operating with $K_{\chi_j}(\vec{x}_1)$ on $\chi_i(\vec{x}_1)$ involves an ‘‘exchange’’ of electron 1 and electron 2 to the right of r_{12}^{-1} compared with equation 2.39. This is the reason the Coulomb operator is known as a local operator and the exchange operator is a non-local operator - since K_{χ_j} is not uniquely defined at a local point in space \vec{x}_1 [18]. K_{χ_j} does not depend just on the function and its infinitesimal neighbourhood [19]. The non-local nature of K_{χ_j} makes it difficult to interpret physically. It can be shown to arise entirely due to the anti-symmetry requirement of fermions.

In this section, we have seen how restricting our wavefunction to a single Slater determinant is what causes the averaging of interelectron repulsions, as the Hamiltonian being measured is fixed. One needs to be careful of some chemistry texts which talk about the ‘‘average’’ electron repulsion term in the Fock operator, which can be misleading as to why it arises. The origin lies in approximating the ground state as a single Slater determinant and has nothing to do with the Hamiltonian, which remains unchanged.

Next, we show how the integrals over one and two-body operators are determined with respect to spin orbitals.

2.5.3 Integrals

A major part of quantum chemistry calculations is determining integrals over spin $\chi_i(\vec{x})$ and spatial $\psi_i(\vec{r})$ molecular orbitals. The notation used is as follows [18]:

$$\langle \chi_i | h | \chi_j \rangle = \langle i | h | j \rangle = [i|h|j] = \int d\vec{x}_1 \chi_i^*(\vec{x}_1) h(\vec{x}_1) \chi_j(\vec{x}_1), \quad (2.40a)$$

$$\langle \chi_i \chi_j | \chi_k \chi_l \rangle = \langle ij | kl \rangle = \int d\vec{x}_1 \int d\vec{x}_2 \underbrace{\chi_i^*(\vec{x}_1) \chi_j^*(\vec{x}_2)}_{\text{electron 1+2}} \frac{1}{r_{12}} \underbrace{\chi_k(\vec{x}_1) \chi_l(\vec{x}_2)}_{\text{electron 1+2}}, \quad (2.40b)$$

$$[\chi_i \chi_j | \chi_k \chi_l] = [ij|kl] = \int d\vec{x}_1 \int d\vec{x}_2 \underbrace{\chi_i^*(\vec{x}_1) \chi_j(\vec{x}_1)}_{\text{electron 1}} \frac{1}{r_{12}} \underbrace{\chi_k^*(\vec{x}_2) \chi_l(\vec{x}_2)}_{\text{electron 2}}, \quad (2.40c)$$

$$\langle \chi_i \chi_j | \chi_k \chi_l \rangle = \langle ij | kl \rangle = \langle ij | kl \rangle - \langle ij | lk \rangle, \quad (2.40d)$$

$$\langle ij | kl \rangle = [ik|jl], \quad (2.40e)$$

and

$$h_{ij} = (\psi_i|h|\psi_j) = (i|h|j) = \int d\vec{r}_1 \psi_i^*(\vec{r}_1) h(\vec{r}_1) \psi_j(\vec{r}_1), \quad (2.41a)$$

$$(\psi_i\psi_j|\psi_k\psi_l) = (ij|kl) = \int d\vec{r}_1 \int d\vec{r}_2 \psi_i^*(\vec{r}_1)\psi_j(\vec{r}_1) \frac{1}{r_{12}} \psi_k^*(\vec{r}_2)\chi_l(\vec{r}_2), \quad (2.41b)$$

$$J_{ij} = (\psi_i\psi_i|\psi_j\psi_j) = (ii|jj), \quad (2.41c)$$

$$K_{ij} = (\psi_i\psi_j|\psi_j\psi_i) = (ij|ji). \quad (2.41d)$$

The square brackets denote chemist notation and the standard bra ket notation is known as physicist notation. The rounded brackets represent integrals over spatial orbitals $\{\psi\}$.

The operators in Hartree-Fock theory (\hat{h} and $\frac{1}{r_{12}}$ - equation 2.26) do not depend on spin coordinates ($\alpha(\omega)/\beta(\omega)$ - note we use $\sigma(\omega)$ as a placeholder for these). This means integrals over \vec{x} can be factorised into integrals over the spin coordinate involving no operators and a more complicated integral over spatial coordinates (involving operators). We find that:

$$\begin{aligned} \langle i|h|j \rangle &= \int d\vec{x}_1 \chi_i^*(\vec{x}_1) h(\vec{r}_1) \chi_j(\vec{x}_1) \\ &= \int d\vec{r}_1 \int d\omega_1 \underbrace{\psi_i^*(\vec{r}_1)\sigma_i^*(\omega_1)}_{\text{expanded } \chi_i^*(\vec{x}_1)} \hat{h}(\vec{r}_1) \underbrace{\psi_j(\vec{r}_1)\sigma_j(\omega_1)}_{\text{expanded } \chi_j(\vec{x}_1)} \\ &= \left[\int d\omega_1 \sigma_i^*(\omega_1)\sigma_j(\omega_1) \right] \int d\vec{r}_1 \psi_i^*(\vec{r}_1) \hat{h}(\vec{r}_1) \psi_j(\vec{r}_1) \\ &= \left[\int d\omega_1 \sigma_i^*(\omega_1)\sigma_j(\omega) \right] \times (i|h|j). \end{aligned} \quad (2.42)$$

Likewise, the spin in the two-electron integrals can also be factored out:

$$\begin{aligned} [ij|kl] &= \int d\vec{x}_1 \int d\vec{x}_2 \chi_i^*(\vec{x}_1)\chi_j(\vec{x}_1) \frac{1}{r_{12}} \chi_k^*(\vec{x}_2)\chi_l(\vec{x}_2) \\ &= \int d\vec{r}_1 \int d\omega_1 \int d\vec{r}_2 \int d\omega_2 \underbrace{\psi_i^*(\vec{r}_1)\sigma_i^*(\omega_1)}_{\text{expanded } \chi_i^*(\vec{x}_1)} \underbrace{\psi_x(\vec{r}_1)\sigma_i(\omega_1)}_{\text{expanded } \chi_j(\vec{x}_1)} \frac{1}{r_{12}} \underbrace{\psi_k^*(\vec{r}_2)\sigma_k^*(\omega_2)}_{\text{expanded } \chi_k^*(\vec{x}_2)} \underbrace{\psi_l(\vec{r}_2)\sigma_l(\omega_2)}_{\text{expanded } \chi_l(\vec{x}_2)} \\ &= \left[\int d\omega_1 \sigma_i^*(\omega_1)\sigma_j(\omega_1) \right] \left[\int d\omega_2 \sigma_k^*(\omega_2)\sigma_l(\omega_2) \right] \int d\vec{r}_1 \int d\vec{r}_2 \psi_i^*(\vec{r}_1)\psi_j(\vec{r}_1) \frac{1}{r_{12}} \psi_k^*(\vec{r}_2)\chi_l(\vec{r}_2) \\ &= \left[\int d\omega_1 \sigma_i^*(\omega_1)\sigma_j(\omega_1) \right] \left[\int d\omega_2 \sigma_k^*(\omega_2)\sigma_l(\omega_2) \right] \times (ij|kl). \end{aligned} \quad (2.43)$$

Using the rules of spin integration (Section 2.2):

$$\begin{aligned}
 h_{ij} = \langle i|h|j\rangle &= \left[\int d\omega_1 \sigma_i^*(\omega_1) \sigma_j(\omega) \right] \times (i|h|j) \\
 &= \begin{cases} (i|h|j), & \text{if } \sigma_i = \sigma_j \\ 0, & \text{if } \sigma_i \neq \sigma_j \end{cases},
 \end{aligned} \tag{2.44}$$

and

$$\begin{aligned}
 [ij|kl] = \langle ik|jl\rangle &= \left[\int d\omega_1 \sigma_i^*(\omega_1) \sigma_j(\omega_1) \right] \left[\int d\omega_2 \sigma_k^*(\omega_2) \sigma_l(\omega_2) \right] \times (ij|kl) \\
 &= \begin{cases} (ij|kl), & \text{if } \sigma_i = \sigma_j \text{ and } \sigma_k = \sigma_l \\ 0, & \text{if } \sigma_i \neq \sigma_j \text{ or } \sigma_k \neq \sigma_l \end{cases}.
 \end{aligned} \tag{2.45}$$

Equations 2.42 and 2.45 represent the generic one and two-electron integrals.

Using these results, we can write the Hartree-Fock energy (equation 2.34) in terms of spin orbitals as:

$$\begin{aligned}
 E_{HF} = \langle \Psi_0 | H | \Psi_0 \rangle &= \sum_i^{N_{elec}} \langle i | \hat{h} | i \rangle + \sum_i^{N_{elec}} \sum_{\substack{j \\ \forall j < i}}^{N_{elec}} [ii|jj] - [ij|ji] \\
 &= \sum_i^{N_{elec}} \langle i | \hat{h} | i \rangle + \frac{1}{2} \sum_i^{N_{elec}} \sum_j^{N_{elec}} [ii|jj] - [ij|ji].
 \end{aligned} \tag{2.46}$$

This can be simplified to equations involving spatial orbitals by integrating over the spin variables, as outlined in this section. The restricted Hartree-Fock energy for a closed-shell ground state in terms of spatial orbitals is [18]:

$$\begin{aligned}
 E_{HF} = \langle \Psi_0 | H | \Psi_0 \rangle &= 2 \sum_i^{N_{elec}/2} (\psi_i | \hat{h} | \psi_i) + \sum_i^{N_{elec}/2} \sum_j^{N_{elec}/2} 2(\psi_i \psi_i | \psi_j \psi_j) - (\psi_i \psi_j | \psi_j \psi_i) \\
 &= 2 \sum_i^{N_{elec}/2} h_{ii} + \sum_i^{N_{elec}/2} \sum_j^{N_{elec}/2} 2J_{ij} - K_{ij},
 \end{aligned} \tag{2.47}$$

where the spatial orbitals for α and β spins are ‘restricted’ to be the same. Next we will see how the energy of a single determinant can be minimized.

2.5.4 The Fock operator

Given our Hartree-Fock approximation to the ground state wavefunction being a single Slater determinant, we can now evaluate $E_{HF}[\{\chi_i\}] = \langle \Psi_0 | H | \Psi_0 \rangle$ (equation

2.46). The Hartree-Fock energy E_{HF} is a functional (function of a function) of the spin orbitals $\{\chi_i\}$. The goal of Hartree-Fock is to minimize $E_{HF}[\{\chi_i\}]$, subject to the constraint that the spin orbitals remain orthonormal [18]:

$$\int d\vec{x}_1 \chi_i^*(\vec{x}_1) \chi_j(\vec{x}_1) = \langle \chi_i | \chi_j \rangle = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}. \quad (2.48)$$

The Lagrangian undetermined multipliers method can be used to tackle such an optimization problem that is subject to constraints. For brevity, we omit the derivation, which is provided in many quantum chemistry textbooks [18, 19]. The constrained optimization results in the Hartree-Fock equation:

$$\left[\hat{h}(\vec{x}_1) + \sum_{j \neq i} J_{\chi_j}(\vec{x}_1) - \sum_{j \neq i} K_{\chi_j}(\vec{x}_1) \right] \chi_i(\vec{x}_1) = \epsilon_i \chi_i(\vec{x}_1). \quad (2.49)$$

The restriction over $j \neq i$ turns out to be unnecessary as [18]:

$$[J_{\chi_i}(\vec{x}_1) - K_{\chi_i}(\vec{x}_1)] \chi_i(\vec{x}_1) = 0, \quad (2.50)$$

which is clear from equations 2.38 and 2.39. This cancellation is important, as it stops the Coulombic attraction between an electron and itself, which is not possible. If this term didn't cancel it would lead to self-interaction error, which is a problem in density functional theory [24]. This issue will be discussed later in Section 2.9.4. Equation 2.49 can therefore be written without any summation restriction:

$$\left[\hat{h}(\vec{x}_1) + \sum_j J_{\chi_j}(\vec{x}_1) - \sum_j K_{\chi_j}(\vec{x}_1) \right] \chi_i(\vec{x}_1) = \epsilon_i \chi_i(\vec{x}_1). \quad (2.51)$$

Equation 2.51 allows for the (Hartree-Fock) energy to be minimized with respect to the spin orbitals and is commonly known as the Hartree-Fock equations. The quantity in the square brackets is known as the Fock operator f , which is defined as:

$$f(\vec{x}_1) = \hat{h}(\vec{x}_1) + \sum_j J_{\chi_j}(\vec{x}_1) - \sum_j K_{\chi_j}(\vec{x}_1). \quad (2.52)$$

The Hartree-Fock equations can therefore be written concisely as:

$$\begin{aligned} \left[\hat{h}(\vec{x}_1) + v^{HF}(\vec{x}_1) \right] \chi_i(\vec{x}_1) = \\ f(\vec{x}_1) \chi_i(\vec{x}_1) = \epsilon_i \chi_i(\vec{x}_1), \end{aligned} \quad (2.53)$$

where $v^{HF}(\vec{x}_1)$ is known as the Hartree-Fock potential [18]. This is an eigenvalue problem, where the spin orbitals are the eigenfunctions and spin orbital energies are

the eigenvalues. The exact solutions to these integro-differential equations are the “exact” Hartree-Fock spin orbitals. However, this is a pseudo-eigenvalue problem, as the Fock operator depends on the solutions $\{\chi_i\}$ through the Coulomb and exchange operators. This means these equations are nonlinear and need to be solved iteratively.

2.5.5 Roothan equations and basis sets

In the previous section, we wrote the Hartree-Fock equations (equation 2.53) now they need to be solved. However, as before we need to remove the spin dependency so it becomes a spatial problem. The Hartree-Fock equation (2.51) as it stands depends on spin orbitals and so alpha and beta spin functions. E.g. for an alpha function:

$$f(\vec{x}_1)\psi_k(\vec{r}_1)\alpha(\omega_1) = \epsilon_k\psi_k(\vec{r}_1)\alpha(\omega_1), \quad (2.54)$$

where ϵ_k is the energy of the spatial orbital ψ_k . We can integrate out the spin component by first multiplying on the left by $\alpha^*(\omega_1)$:

$$\left[\int d\omega_1 \alpha^*(\omega_1) f(\vec{x}_1) \alpha(\omega_1) \right] \psi_k(\vec{r}_1) = \epsilon_k \psi_k(\vec{r}_1) \alpha(\omega_1). \quad (2.55)$$

We omit the details of the full derivation, which involves integration over alpha and beta components respectively, which can be found in many textbooks such as [18, 19]. While this step is similar for restricted and unrestricted formalisms, care must be taken in unrestricted calculations due to the spin-up-and-down orbitals having different spatial parts [18]. Overall, the result is that the Hartree-Fock equations can be reduced to a problem dependent on space only. The following spatial integro-differential equation must then be solved:

$$f(\vec{r}_1)\psi_i(\vec{r}_1) = \epsilon_i\psi_i(\vec{r}_1). \quad (2.56)$$

This can be done numerically; however, it is often far too hard except for the simplest systems. From here we consider the problem for restricted calculations (the approach for unrestricted calculations is similar, but requires careful treatment of the alpha and beta parts of the problem - see the Pople-Nesbet equations in [18, 19]). Roothan and Hall independently solved this issue [25, 26] where they showed that by introducing a set of known spatial basis functions the differential could then be converted to a system of linear equations and then solved via standard matrix techniques.

In effect what is done is one introduces a set of K known spatial basis functions $\{\phi_i | i = 0, 1, \dots, K-1\}$. Each unknown spatial molecular orbital $\psi_i(\vec{r})$, each a function of standard three-dimensional space, is expanded as a linear combination of these

known spatial basis functions [18, 19]:

$$\psi_i(\vec{r}) = \sum_{\mu=0}^{K-1} C_{\mu i} \phi_{\mu}(\vec{r}). \quad (2.57)$$

For a given calculation the set of K known basis functions is fixed. Here C is a $K \times K$ matrix, where each column i defines the expansion coefficients for molecular orbital i . A technical note on the basis functions ϕ_j is they are normalized, but in general not orthogonal to each other [18]. If the set of basis functions used is complete then each $\psi_i(\vec{r})$ could be exactly expanded. However, in general this requires infinite spatial basis functions and so for practical reasons a finite number are used in real calculations i.e. within a Galerkin approximation rendering the problem finite dimensional [27, 28].

We can substitute equation 2.57 into equation 2.56 to obtain the Hartree-Fock equation:

$$f(\vec{r}_1) \left(\sum_{\mu=0}^{K-1} C_{\mu i} \phi_{\mu}(\vec{r}_1) \right) = \epsilon_i \left(\sum_{\mu=0}^{K-1} C_{\mu i} \phi_{\mu}(\vec{r}_1) \right). \quad (2.58)$$

Then left multiplying by ϕ_{ν}^* and integrating [18]:

$$\sum_{\mu=0}^{K-1} C_{\mu i} \left(\int d\vec{r}_1 \phi_{\mu}^*(\vec{r}_1) f(\vec{r}_1) \phi_{\mu}(\vec{r}_1) \right) = \epsilon_i \sum_{\mu=0}^{K-1} C_{\mu i} \phi_{\mu}^*(\vec{r}_1) \phi_{\mu}(\vec{r}_1), \quad i = 0, 1, \dots, (K-1). \quad (2.59)$$

This integrated Hartree-Fock equation can be written as [18]:

$$\sum_{\mu=0}^{K-1} F_{\mu i} C_{\mu i} = \epsilon_i \sum_{\mu=0}^{K-1} S_{\mu i} C_{\mu i}, \quad (2.60)$$

where S is the overlap matrix [18]:

$$S_{\mu\nu} = \int d\vec{r}_1 \phi_{\mu}^*(\vec{r}_1) \phi_{\nu}(\vec{r}_1) = \langle \phi_{\mu} | \phi_{\nu} \rangle, \quad (2.61)$$

and F is the Fock matrix [18]:

$$F_{\mu\nu} = \int d\vec{r}_1 \phi_{\mu}^*(\vec{r}_1) f(\vec{r}_1) \phi_{\nu}(\vec{r}_1) = \langle \phi_{\mu} | f | \phi_{\nu} \rangle. \quad (2.62)$$

This is a representation of the Fock operator with the set of basis functions $\{\phi_i\}$. The set of equations defined by each i (see equation 2.60) are known as the Roothan-Hall equations (sometimes just Roothan equations), which can be written as a single

matrix equation [18]:

$$FC = SC\epsilon. \quad (2.63)$$

Here C is the $K \times K$ matrix of expansion coefficients $C_{\mu i}$ and ϵ is a $K \times K$ diagonal matrix of orbital energies ϵ_i . The columns of C define the Hartree-Fock molecular orbitals ψ_i (equation 2.57) and are determined by solving equation 2.63. Importantly, as the Fock matrix F depends on the expansion coefficients C , the Roothan equations are nonlinear. They are therefore solved in an iterative diagonalization process until the matrix of expansion coefficients C converges [18]. A full breakdown of the algorithm is given in many textbooks such as [18, 19, 29, 30].

2.6 Atomic to molecular orbitals

From a Hartree-Fock calculation, we obtain results in terms of integrals and matrices over the basis set $\{\phi_i\}$ rather than over the set of molecular orbitals (solutions) $\{\psi_i\}$. For subsequent treatment of correlation effects, it is very useful to represent the basic integrals in terms of the basis functions to integrals in terms of the molecular orbitals. This transformation is commonly known as the atomic-to-molecular orbital transformation. The one-electron integrals are transformed as:

$$h_{ij} = (\psi_i|h|\psi_j) = \sum_{\mu} \sum_{\nu} C_{\mu i}^* C_{\nu j} H_{\mu\nu}^{\text{core}}, \quad (2.64)$$

where:

$$H_{\mu\nu}^{\text{core}} = T_{\mu\nu} + V_{\mu\nu}, \quad (2.65a)$$

$$T_{\mu\nu} = -\frac{1}{2} \langle \phi_{\mu} | \nabla_{\vec{r}_1}^2 | \phi_{\nu} \rangle = \int d\vec{r}_1 \phi_{\mu}^*(\vec{r}_1) \nabla_{\vec{r}_1}^2 \phi_{\nu}(\vec{r}_1), \quad (2.65b)$$

$$V_{\mu\nu} = \sum_{A=0}^{M-1} \langle \phi_{\mu} | \frac{Z_A}{|\vec{r}_1 - \vec{R}_A|} | \phi_{\nu} \rangle = \int d\vec{r}_1 \phi_{\mu}^*(\vec{r}_1) \frac{Z_A}{|\vec{r}_1 - \vec{R}_A|} \phi_{\nu}(\vec{r}_1). \quad (2.65c)$$

The two-electron integrals are transformed as:

$$(\psi_i \psi_j | \psi_k \psi_l) = \sum_{\mu} \sum_{\nu} \sum_{\lambda} \sum_{\sigma} C_{\mu i}^* C_{\nu j} C_{\lambda k}^* C_{\sigma l} (\phi_{\mu} \phi_{\nu} | \phi_{\lambda} \phi_{\sigma}), \quad (2.66)$$

where the two-electron integrals over basis functions are given by:

$$(\phi_{\mu} \phi_{\nu} | \phi_{\lambda} \phi_{\sigma}) = \int d\vec{r}_1 \int d\vec{r}_2 \phi_{\mu}^*(\vec{r}_1) \phi_{\nu}(\vec{r}_1) \frac{1}{r_{12}} \phi_{\lambda}^*(\vec{r}_2) \phi_{\sigma}(\vec{r}_2). \quad (2.67)$$

Importantly the transformation in equation 2.66 scales as $\mathcal{O}(K^5)$ [31]. Once transformed, the molecular Hamiltonian can be written in this new basis and is known as the second quantized molecular Hamiltonian.

2.7 Second quantization

The antisymmetry principle is not present in the Schrödinger equation. It must therefore be imposed by alternate means. In the first quantisation of the electronic Hamiltonian, it is enforced by the properties of Slater determinants [17, 18, 32, 33]. These are matrix determinants of spin orbitals.

An alternate method to include this property of fermions is to have certain algebraic properties obeyed by operators [18]. This is second quantization and importantly no new physics is present. First, we associate to every spin orbital χ_i a creation operator a_i^\dagger and an annihilation operator a_i . In order for these operators to maintain the antisymmetry principle these operators must obey the following fermionic anti-commutation relations [34, 35]:

$$\{a_i, a_j\} = \{a_i^\dagger, a_j^\dagger\} = 0, \quad (2.68a)$$

$$\{a_i, a_j^\dagger\} = \delta_{ij} \mathcal{I}_i, \quad (2.68b)$$

where anti-commutator is defined as $\{A, B\} \equiv AB + BA$. The action of these fermionic operators on the determinant of M orbitals $|f_0, f_1, \dots, f_{M-1}\rangle = |f\rangle$ is given as:

$$\begin{aligned} \alpha_j^\dagger |f_0, \dots, f_{j-1}, 0, f_{j+1}, \dots, f_{M-1}\rangle &= (-1)^{\sum_{s=0}^{j-1} f_s} |f_0, \dots, f_{j-1}, 1, f_{j+1}, \dots, f_{M-1}\rangle \\ \alpha_j |f_0, \dots, f_{j-1}, 1, f_{j+1}, \dots, f_{M-1}\rangle &= (-1)^{\sum_{s=0}^{j-1} f_s} |f_0, \dots, f_{j-1}, 0, f_{j+1}, \dots, f_{M-1}\rangle \\ \alpha_j^\dagger |f_0, \dots, f_{j-1}, 1, f_{j+1}, \dots, f_{M-1}\rangle &= 0 \\ \alpha_j |f_0, \dots, f_{j-1}, 0, f_{j+1}, \dots, f_{M-1}\rangle &= 0 \end{aligned} \quad (2.69)$$

An example of their use is instructive, consider a system of four orbitals ($M = 4$):

$$\alpha_0^\dagger |vac\rangle = \alpha_0^\dagger |0000\rangle = |1000\rangle, \quad (2.70a)$$

$$\alpha_1^\dagger |vac\rangle = \alpha_1^\dagger |0000\rangle = |0100\rangle, \quad (2.70b)$$

$$\alpha_2^\dagger |vac\rangle = \alpha_2^\dagger |0000\rangle = |0010\rangle, \quad (2.70c)$$

$$\alpha_3^\dagger |vac\rangle = \alpha_3^\dagger |0000\rangle = |0001\rangle. \quad (2.70d)$$

Note that due to eq. 2.68, the occupation number of fermions cannot be greater than

one or less than zero, i.e:

$$\alpha_0^\dagger |1000\rangle = \alpha_0^\dagger \alpha_0^\dagger |vac\rangle = 0, \alpha_0 \alpha_0 |1000\rangle = \alpha_0 |vac\rangle = 0. \quad (2.71)$$

The fermion antisymmetry can also be seen to be built into these operators:

$$|1100\rangle = \alpha_0^\dagger |0100\rangle = \alpha_0^\dagger \alpha_1^\dagger |vac\rangle \xrightarrow{\substack{\text{exchange} \\ \text{electron} \\ \text{labels}}} \alpha_1^\dagger \alpha_0^\dagger |vac\rangle = \alpha_1^\dagger |1000\rangle = -(1)^1 |1100\rangle = -|1100\rangle. \quad (2.72)$$

The molecular Hamiltonian can be written in this formalism. Under the Born-Oppenheimer approximation and a basis ϕ of K orthonormal molecular orbitals, constructed as a linear combination of atomic orbitals (LCAO) often computed from mean field calculations such as Hartree Fock [36], the second quantised form of the electronic Hamiltonian is defined as [35]:

$$H = \sum_{p=0}^{K-1} \sum_{q=0}^{K-1} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{p=0}^{K-1} \sum_{q=0}^{K-1} \sum_{r=0}^{K-1} \sum_{s=0}^{K-1} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \quad (2.73)$$

where:

$$h_{pq} = \langle \chi_p | h | \chi_q \rangle = \int d\vec{x}_i \phi_p^*(\vec{x}_i) \left(-\frac{\nabla_{\vec{r}_i}^2}{2} - \sum_I \frac{Z_I}{|\vec{r}_i - \vec{R}_I|} \right) \phi_q(\vec{x}_i), \quad (2.74)$$

$$h_{pqrs} = \langle \chi_p \chi_q | \chi_r \chi_s \rangle = \int d\vec{x}_i d\vec{x}_j \frac{\phi_p^*(\vec{x}_i) \phi_q^*(\vec{x}_j) \phi_r(\vec{x}_i) \phi_s(\vec{x}_j)}{|\vec{r}_i - \vec{r}_j|}. \quad (2.75)$$

Note the integrals over spatial and spin coordinates can be simplified to integrals over spatial coordinates by integrating over spin - see Section 2.5.3 (specifically equations 2.44 and equation 2.45).

The electronic Hamiltonian H in equation 2.73 is known as the second quantized molecular Hamiltonian. The h_{pq} term represents the kinetic energy of each electron and its Coulombic interaction with the nuclei. The second integral gives the electron-electron Coulomb repulsion contribution. The four-fold sum describing this interaction means the number of terms in H scales as $\mathcal{O}(K^4)$, where K is the number of basis functions [17].

2.8 Post-Hartree-Fock methods

The Hartree-Fock technique is the standard starting point for most wavefunction-based quantum chemistry algorithms. The so-called post-Hartree-Fock methods then

seek to obtain further correlation effects by relaxing the condition on the ground state wavefunction being a single Slater determinant. In the literature what is commonly discussed is the “correlation energy”: E_{corr} . This is given by the difference between the non-relativistic ground state energy of the system E_0 and the Hartree-Fock energy E_{HF} - in full: $E_{corr} = E_0 - E_{HF}$ [18]. As the Hartree-Fock energy is an upper bound on the ground state energy, due to the variational principle, E_{corr} must be negative. Post-Hartree-Fock methods seek ways of finding E_{corr} . Commonly used routines include Møller-Plesset (MP) [37, 38], Configuration Interaction (CI) [39], and Coupled-cluster (CC) [40]. In each technique, the number of allowed Slater determinants is increased allowing more correlation effects to be captured and when all Slater determinants are considered then all correlation effects are included. In that scenario, the full Hilbert space is described and the true ground state energy (full configuration interaction) energy can be obtained. However, this problem scales exponentially due to the number of Slater determinants required scaling as:

$$N_{det} = \binom{K}{\eta} = \frac{K!}{\eta!(K-\eta)!}, \quad (2.76)$$

for K spin orbitals and η electrons. As $\sum_{\eta=0}^K \binom{K}{\eta} = 2^K$, equation 2.76 scales at worst as $\mathcal{O}(2^K)$. This means for large basis sets FCI is not possible, due to the combinatorial overhead. Note, the space of all possible Slater determinants of η particles in K spin orbitals is known as the η -particle Hilbert space.

To solve systems in large basis sets, methods that approximately expand the wavefunction using fewer Slater determinants are used - a polynomial number rather than an exponential. Examples include Møller-Plesset second (MP2), third (MP3), fourth (MP4), etc order, configuration interaction singles (CIS), doubles (CID) and single doubles (CISD) etc, and coupled-cluster singles (CCS), doubles, (CCD) single doubles (CCSD) etc. Details on these are outside the scope of this thesis. The advantage of studying these types of problems on a quantum device is a K orbital space can be described on K -qubits (in the second quantization). The memory overhead of classical methods required to track the coefficient of each Slater determinant has an exponential cost in K rather than a linear qubit cost. This memory advantage of quantum devices is why many believe quantum computers offer a route to quantum advantage via quantum chemistry. However, it is important to note a few caveats. First, it is not possible to know the amplitudes of an exponentially large quantum state as this would require storing an exponential number of wavefunction amplitudes (and would require an exponential amount of time to determine). Secondly, the ground state problem is Quantum Merlin Arthur (QMA) complete [41, 42]. QMA is a complexity class that contains decision problems that are easy to verify on a quantum computer, but not necessarily easy to solve [42]. Often this is said to be the quantum counterpart of NP

problems (here an NP-complete problem).

Next, we introduce an alternate strategy to solve the time-independent Schrödinger equation known as density functional theory (DFT).

2.9 Density functional theory

Density functional theory (DFT) is a method that solves the electronic Schrödinger equation. However, rather than using the many-electron wavefunction in calculations the electron density $\rho(\vec{r})$ of a system is used. For a normalized η -electron wavefunction:

$$\int d\vec{x}_0 \dots d\vec{x}_{\eta-1} |\Psi(\vec{x}_0, \dots, \vec{x}_{\eta-1})|^2 = 1, \quad (2.77)$$

where the integral is taken over all space and sum over both spins. The electron density is defined as:

$$\rho(\vec{r}) = \eta \sum_{\sigma} \int d\vec{r}_1 \dots \int d\vec{r}_{\eta-1} |\Psi(\vec{r}, \sigma, \vec{x}_1, \dots, \vec{x}_{\eta-1})|^2, \quad (2.78)$$

where $\rho(\vec{r})d^3r$ has the interpretation of being the probability density for finding any electron in the volume d^3r around \vec{r} . The density is normalized and so integrating over all space will give the number of electrons:

$$\int \rho(\vec{r})d^3r = \eta. \quad (2.79)$$

To see how these objects allow the electronic Schrödinger equation to be solved, we first need to introduce the Hohenberg-Kohn theorems that underpin the workings of DFT.

2.9.1 Hohenberg-Kohn theorems

In the 1960s, Walter Kohn and Pierre Hohenberg proved two important mathematical theorems that DFT hinges on. First, consider the molecular Hamiltonian under the clamped-nuclei approximation (equation 2.3):

$$H = \underbrace{-\frac{1}{2} \sum_{i=0}^{\eta-1} \nabla_i^2}_T + \underbrace{\sum_{i=0}^{\eta-1} v_{ne}(\vec{r}_i)}_{V_{ne}} + \underbrace{\sum_{i=0}^{\eta-1} \sum_{j>i}^{\eta-1} \frac{1}{|\vec{r}_i - \vec{r}_j|}}_{V_{ee}}, \quad (2.80)$$

where the nuclear-electron interaction has been written as:

$$v_{ne}(\vec{r}_i) = - \sum_{A=0}^{M-1} \frac{Z_A}{|\vec{r}_i - \vec{R}_A|}. \quad (2.81)$$

The Hamiltonian can be written as $H = T + V_{ee} + V_{ne}$. Let us consider the electronic structure obtained by replacing the potential $v_{ne}(\vec{r})$ by an arbitrary potential $v_{ext}(\vec{r})$, so that $H = T + V_{ee} + V_{ext}$. The ground state wavefunction obtained by solving the Schrödinger equation allows for a ground state density to be defined (equation 2.78). This means there is a mapping from the external potential to the ground state density $v_{ext}(\vec{r}) \mapsto \rho(\vec{r})$. Hohenberg and Kohn showed that this mapping can be inverted up to a constant [43], i.e. $\rho(\vec{r}) \mapsto v_{ext}(\vec{r}) + const$. This is the first Hohenberg-Kohn theorem, which proved that the external potential $v_{ext}(\vec{r})$ is uniquely determined by the corresponding ground-state electronic density, to within an additive constant. Quoting the original paper:

“the $v_{ext}(\vec{r})$ is (to within a constant) a unique functional of $\rho(\vec{r})$ since, in turn, $v_{ext}(\vec{r})$ fixes H we see that the full many-particle ground state is a unique functional of $\rho(\vec{r})$ ” [43].

The two-step proof of this is obtained via a contradiction. Suppose there are two local potentials differing by more than a constant $v_{ext,1}(\vec{r}) \neq v_{ext,2}(\vec{r}) + const$. Let these external potentials define two Hamiltonians $H_1 = T + V_{ee} + V_1$ and $H_2 = T + V_{ee} + V_2$. Now assuming the ground state wavefunction of these Hamiltonians $|\Psi\rangle$ is the same - i.e. $H_1 |\Psi\rangle = E_1 |\Psi\rangle$ and $H_2 |\Psi\rangle = E_2 |\Psi\rangle$. Subtracting $H_1 - H_2 = V_1 - V_2$ gives:

$$(H_1 - H_2) |\Psi\rangle = (V_1 - V_2) |\Psi\rangle = (E_1 - E_2) |\Psi\rangle \quad (2.82)$$

in the position representation:

$$\sum_i^\eta (v_{ext,1}(\vec{r}_i) - v_{ext,2}(\vec{r}_i)) \Psi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) = (E_1 - E_2) \Psi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}). \quad (2.83)$$

If we re-arrange these equations, we obtain:

$$\left(\underbrace{\Delta V}_{V_1 - V_2} - \underbrace{\Delta E}_{E_1 - E_2} \right) |\Psi\rangle = 0. \quad (2.84)$$

Assuming $\Psi(\vec{x}_0, \vec{x}_1, \dots, \vec{x}_{\eta-1}) = |\Psi\rangle \neq 0$, this implies that $\Delta V = \Delta E$ and thus $v_{ext,1}(\vec{r}) - v_{ext,2}(\vec{r}) = const$ [44, 45], which contradicts the original assumption [43]. A rigorous mathematical analysis of this is outside the scope of this thesis, for an interested reader we recommend the work of Pino *et al.* in [45]. The result of this is that two local potentials differing by more than an additive constant cannot share the same ground-state wavefunction.

The second part of the proof then considers two Hamiltonians (H_1 , and H_2) that

have two potentials differing by more than a constant that necessarily have different ground state wavefunctions (by the first part of proof) Ψ_1 and Ψ_2 , but the same ground state density $\rho(\vec{r})$. Assuming H_1 and H_2 each have non-degenerate ground states, then by the variational principle:

$$\begin{aligned} E_1 &= \langle \Psi_1 | H_1 | \Psi_1 \rangle < \langle \Psi_2 | H_1 | \Psi_2 \rangle = \langle \Psi_2 | \underbrace{H_2 + V_1 - V_2}_{H_1} | \Psi_2 \rangle \\ &= \langle \Psi_2 | H_2 | \Psi_2 \rangle + \langle \Psi_2 | (V_1 - V_2) | \Psi_2 \rangle = E_2 + \int [v_{ext,1}(\vec{r}) - v_{ext,2}(\vec{r})] \rho(\vec{r}) d\vec{r}, \end{aligned} \quad (2.85)$$

where the inequality comes from Ψ_2 not being the ground state of H_1 . The labels of systems 1 and 2 can be arbitrarily swapped and so equation 2.85 can be written:

$$\begin{aligned} E_2 &= \langle \Psi_2 | H_2 | \Psi_2 \rangle < \langle \Psi_1 | H_1 | \Psi_1 \rangle + \langle \Psi_1 | (V_2 - V_1) | \Psi_1 \rangle \\ &= \langle \Psi_1 | H_1 | \Psi_1 \rangle - \underbrace{\langle \Psi_1 | (V_1 - V_2) | \Psi_1 \rangle}_{\text{note sign and thus order change}} = E_1 - \int [v_{ext,1}(\vec{r}) - v_{ext,2}(\vec{r})] \rho(\vec{r}) d\vec{r}. \end{aligned} \quad (2.86)$$

However, adding equation 2.85 and 2.86 results in the following contradiction [43, 46]:

$$E_1 + E_2 < E_1 + E_2. \quad (2.87)$$

This implies that there cannot be two local potentials differing more than an additive constant, which share the same ground state density. We note that if a constant is added to the potential then the wavefunction remains unchanged and thus the density is unaltered to [47] (as the eigenvectors will remain unchanged and only the eigenvalues will be shifted by the constant). Overall the ground state density $\rho(\vec{r})$ determines the potential $v_{ext}(\vec{r})$ which determines the Hamiltonian and so the many-body problem. This means that the potential v_{ext} is a unique functional¹ (up to an additive constant) of the ground-state density ρ . This will also determine all the ground-state properties of the system, since the Hamiltonian and ground-state wavefunction are determined. Furthermore, as the external potential gives a mapping to the ground state density this defines the ground state wavefunction via equation 2.78 i.e. $v_{ext}(\vec{r}) \mapsto \rho_0(\vec{r}) \mapsto \Psi_0$.

¹The formal definition of a functional usually depends on subfield. Here, it is in the context of functional analysis, where it is a mapping from a space X into the field of real \mathbb{R} or complex \mathbb{C} numbers. Such mappings may or may not be assumed to be linear or to be defined on the whole space X . In colloquial language, a function maps a number to a number. Here a functional maps a function to a number.

Using this, Hohenberg and Kohn define a universal functional $F[\rho]$:

$$F[\rho] = T[\rho] + V_{ee}[\rho] = \min_{\substack{\Psi \mapsto \rho(\vec{r}) \\ \rho \in \mathcal{V}}} \left[\langle \Psi | T + V_{ee} | \Psi \rangle \right] \quad (2.88)$$

that is independent of the external potential [43]. Here \mathcal{V} is the set of all v -representable ρ and $\Psi \mapsto \rho$ constrains the search to any wavefunction that yields the fixed density ρ , determined by equation 2.78. Then for a given potential $v_{ext}(\vec{r})$, the total electronic energy of a system is given by the functional [43]:

$$E[\rho] = \underbrace{F[\rho]}_{\text{system independent}} + \underbrace{\int v_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r}}_{\text{system dependent}}. \quad (2.89)$$

Note the Hohenberg-Kohn universal functional (equation 2.88) is only defined over the set of v -representable densities. A v -representable ρ is associated with an antisymmetric (satisfies the Pauli principle [46]) ground-state wavefunction of some Hamiltonian H with local external potential $v_{ext}(\vec{r})$ [48].

The second Hohenberg and Kohn theorem proved that the density functional $E[\rho]$ satisfies a variational principle, assuming the constraint of η electrons in a given external potential $v_{ext}(\vec{r})$ [43]. We can show this as follows. Take a trial density $\tilde{\rho}(\vec{r})$, that satisfies the boundary conditions $\tilde{\rho}(\vec{r}) \geq 0$ and $\int \tilde{\rho}(\vec{r}) d^3\vec{r} = \eta$, this defines a Hamiltonian \tilde{H} with an external potential $\tilde{v}_{ext}(\vec{r})$ that has its own ground state wavefunction $|\tilde{\Psi}\rangle$. This wavefunction can be taken as the trial wavefunction for the true Hamiltonian, generated by the true external potential $v_{ext}(\vec{r})$, and we obtain [47]:

$$E_0 = \langle \Psi_0 | H | \Psi_0 \rangle \leq \langle \tilde{\Psi} | H | \tilde{\Psi} \rangle = E[\tilde{\rho}] = F[\tilde{\rho}] + \int v_{ext}(\vec{r}) \tilde{\rho}(\vec{r}) d\vec{r}. \quad (2.90)$$

The ground state energy can therefore be found by minimizing the energy functional with respect to the set of all v -representable densities \mathcal{V} . This can be written as:

$$E_0 = \min_{\rho \in \mathcal{V}} \left[F[\rho] + \int v_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r} \right]. \quad (2.91)$$

Stated in words, equation 2.91 gives the ground state energy if and only if the input density is the ground state density $\rho_0(\vec{r})$. Therefore, for any trial density $\tilde{\rho}$, that satisfies the boundary conditions $\tilde{\rho}(\vec{r}) \geq 0$ and $\int \tilde{\rho}(\vec{r}) d^3\vec{r} = \eta$, evaluating $E[\tilde{\rho}]$ through equation 2.89 represents an upper bound of the true ground state energy [47, 49].

To summarise, the first Hohenberg-Kohn theorem shows there exists a mapping from a ground state density to a local potential. The second theorem shows there exists a universal density functional that obeys the variational principle with respect to v -representable densities. However, in this original formulation equation 2.91 involves

a minimization over all v -representable densities. A limitation of this approach is $F[\rho]$ is not defined for ρ that are not v -representable. The difficulty with this assumption is that the set of v -representable densities is unknown [50, 51].

The issue of v -representability was subsequently solved by Levy and Lieb, who extended the Hohenberg-Kohn universal functional to consider N -representable densities, which are known [46, 48]. The universal functional now takes the following form:

$$F[\rho] = \min_{\substack{\Psi \mapsto \rho(\vec{r}) \\ \rho \in \mathcal{D}}} \left[\langle \Psi | T + V_{ee} | \Psi \rangle \right], \quad (2.92)$$

where $\Psi \mapsto \rho$ constrains the search to any wavefunction that yields the fixed density ρ , determined by equation 2.78. Here the set \mathcal{D} denotes all the N -representable densities coming from a particular wavefunction. Equation 2.92 is known as the Levy-Lieb functional and does not require the existence of a local potential associated with the density. The set of all N -representable densities coming from different wavefunctions is a larger set than v -representable densities. Cioslowki in [52–54] provided a formal process to generate, in principle, all antisymmetric wavefunctions that lead to a density. This establishes the domain of the constrained search. We note this formulation shows that the ground state density ρ_0 determines the ground state wavefunction $|\Psi_0\rangle$, which determines H uniquely up to an additive constant. Degeneracies pose no problems in this approach [48, 55, 56].

The variational theorem decomposes the energy minimization over Ψ in two steps. The outer step requires defining a set of ρ that define all N -electron densities \mathcal{N} . For each $\rho \in \mathcal{N}$, the inner loop performs a minimization over all wavefunctions that generate that density. The ground state energy is given by the minimum energy from this search. Overall:

$$\begin{aligned} E_0 &= \min_{\rho \in \mathcal{N}} \left[\langle \Psi | T + V_{ee} + V_{Ne} | \Psi \rangle \right] \\ &= \min_{\rho \in \mathcal{N}} \left[\min_{\substack{\Psi \mapsto \rho(\vec{r}) \\ \rho \in \mathcal{D} \subset \mathcal{N}}} \left(\langle \Psi | T + V_{ee} + V_{Ne} | \Psi \rangle \right) \right] \\ &= \min_{\rho \in \mathcal{N}} \left[\min_{\substack{\Psi \mapsto \rho(\vec{r}) \\ \rho \in \mathcal{D} \subset \mathcal{N}}} \left(\langle \Psi | T + V_{ee} | \Psi \rangle \right) + \int v_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r} \right] \\ &= \min_{\rho \in \mathcal{N}} \left[F[\rho] + \int v_{ne}(\vec{r}) \rho(\vec{r}) d\vec{r} \right]. \end{aligned} \quad (2.93)$$

The minimum of this search is obtained for ground-state density $\rho_0(\vec{r})$ corresponding to the potential $v_{ne}(\vec{r})$ (equation 2.81). The existence of a universal functional $F[n]$ that is independent of the external potential is remarkable, as it simplifies a mini-

mization over the many-electron wavefunction to a minimization that only depends on three variables (the density). This is a major reduction in complexity, where the problem now in principle scales linearly with system size. However, a problem with this formulation is an expression for the universal functional $F[\rho]$ in terms of a density is not known. Given equations 2.92 and 2.93, the exact ground state wavefunction is any wavefunction yielding $\rho(\vec{r})$ that minimizes $T + V_{ee}$. We can decompose the universal functional $F[\rho] = T[\rho] + V_{ee}[\rho]$ as:

$$T[\rho] = \min_{\substack{\Psi \rightarrow \rho(\vec{r}) \\ \rho \in \mathcal{D}}} \left[\langle \Psi | T | \Psi \rangle \right], \quad (2.94a)$$

$$V_{ee}[\rho] = \min_{\substack{\Psi \rightarrow \rho(\vec{r}) \\ \rho \in \mathcal{D}}} \left[\langle \Psi | V_{ee} | \Psi \rangle \right]. \quad (2.94b)$$

Here $T[\rho]$ and $V_{ee}[\rho]$ are the exact functionals for the kinetic and electron-electron repulsion terms. Restating our problem, $E[\rho]$ is minimized for a given $v_{ext}(\vec{r})$, while keeping the particle number constant. To enforce this constraint the method of Lagrange multipliers can be used. The following Lagrangian is used: $\mathcal{L}[\rho] = E[\rho(\vec{r})] - \mu(\int \rho(\vec{r})d^3\vec{r} - \eta)$ for integer η . We find [57, 58]:

$$\frac{\partial \left\{ E[\rho] - \mu(\int \rho(\vec{r})d^3\vec{r} - \eta) \right\}}{\partial \rho(\vec{r})} = 0 \quad (2.95)$$

$$\implies \frac{\partial E[\rho]}{\partial \rho(\vec{r})} = \mu,$$

where μ is the Lagrange multiplier and is known as the chemical potential in DFT. We recommend a recent discussion on the misconceptions of μ in [59], which is outside the scope of this thesis to discuss. Equation 2.95 is known as the Euler-Lagrange equation and can be solved for the exact density. Given that:

$$E[\rho] = F[\rho] + \int d^3\vec{r} v_{ext}(\vec{r})\rho(\vec{r}) \quad (2.96)$$

the Euler-Lagrange equation is sometimes expressed as:

$$\mu = \frac{\partial F[\rho]}{\partial \rho(\vec{r})} + v_{ext}(\vec{r}), \quad (2.97)$$

using equation 2.95. Equation 2.97 is the basic working equation of DFT [49]. The exact density is therefore the functional derivative of F , which is equal to the negative of the external potential (up to a constant μ). It would be great if equation 2.97 could be solved directly, as it would be a single integro-differential equation that could be

solved self-consistently yielding a density that could be normalized and inserted back into equation 2.89 to recover the ground-state energy. However, as the form of $T[\rho]$ and $V_{ee}[\rho]$ is unknown with respect to ρ , we need to approximate $F[\rho]$ somehow. One approach is known as Thomas-Fermi theory, where $F[\rho]$ is approximated by the non-interacting kinetic energy of a uniform gas plus the Hartree energy. We do not consider this theory, which is limited by the error caused by approximations of the kinetic energy. We focus on the Kohn-Sham formulation, which is of primary interest in this thesis. This is introduced next.

2.9.2 Kohn-Sham DFT

In 1965, Kohn and Sham introduce a way of approximating $F[\rho]$ directly (equation 2.88) [60]. They did this by considering a fictitious system of non-interacting particles, aka $V_{ee} = 0$ (equation 2.80). The universal functional for this system can be written as:

$$F[\rho] = T_s[\rho] + E_{Hxc}[\rho], \quad (2.98)$$

where subscript s denotes single-electron and is used to denote the system being non-interacting. $E_{Hxc}[\rho] = J[\rho] + E_{xc}[\rho]$ is the Hartree-exchange-correlation functional, that is made up of a classical $J[\rho]$ and exchange-correlation term $E_{xc}[\rho]$ defined later in this section. The kinetic energy functional in this problem is written as:

$$T_s[\rho(\vec{r})] = \min_{\substack{\Phi \rightarrow \rho \\ \rho \in \mathcal{SCN}}} \left[\langle \Phi | T | \Phi \rangle \right], \quad (2.99)$$

for non-interacting electrons. Here we have assumed the Kohn-Sham wavefunction is a single Slater determinant (as equation 2.21). We represent the set of all single Slater determinants as \mathcal{S} . Equation 2.99 represents a constrained search over all single-determinant wavefunctions of N -representable densities. The restriction to single-determinant wavefunctions Φ does not introduce any approximation, as any N -representable density can be obtained from a single-determinant wavefunction [61, 62]². Equation 2.98 can therefore be recast into the constrained search formalism

²This is ensured in complete basis sets, but has been shown to not hold in small basis sets [63].

(equation 2.93) in terms of a single-Slater determinant, yielding:

$$\begin{aligned}
 E_0 &= \min_{\rho \in \mathcal{N}} \left[F[\rho] + \int v_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r} \right] \\
 &= \min_{\rho \in \mathcal{N}} \left[\left(\min_{\substack{\Phi \mapsto \rho \\ \rho \in \mathcal{SCN}}} \langle \Phi | T | \Phi \rangle \right) + E_{Hxc}[\rho] + \int v_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r} \right] \\
 &= \min_{\rho \in \mathcal{N}} \left[\left(\min_{\substack{\Phi \mapsto \rho \\ \rho \in \mathcal{SCN}}} \langle \Phi | T + V_{ext} | \Phi \rangle + E_{Hxc}[\rho] \right) \right] \\
 &= \min_{\substack{\Phi \mapsto \rho \\ \rho \in \mathcal{SCN}}} \left[\langle \Phi | T + V_{ext} | \Phi \rangle + E_{Hxc}[\rho] \right].
 \end{aligned} \tag{2.100}$$

The single Slater determinant $|\Phi\rangle$ that minimizes equation 2.100 is the Kohn-Sham wavefunction that yields the exact ground state density $\rho_0(\vec{r})$. A major improvement of this formulation over equation 2.93, is the constrained search is over single Slater determinant wavefunctions Φ rather than multi-determinant wavefunctions Ψ . The importance of this work was recognised in 1998 when Walter Kohn shared the Nobel prize in Chemistry “for his development of the density-functional theory” [64].

It is worth noting that $T_s[\rho]$ (equation 2.99) is not equal to $T[\rho]$ (equation 2.94a). However, by the clever formulation of Kohn and Sham [60], the correction to $T_s[\rho]$ is included elsewhere. Explicitly, we rewrite equation 2.98 as [47]:

$$F[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho]. \tag{2.101}$$

$E_{xc}[\rho]$ is known as the exchange-correlation functional, which contains the difference between $T_s[\rho]$ and $T[\rho]$ and also the nonclassical part of $V_{ee}[\rho]$. $J[\rho]$ is the Hartree potential [65], which gives the classical electrostatic repulsion energy between two charge distributions $\rho(\vec{r}_1)$ and $\rho(\vec{r}_2)$:

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2. \tag{2.102}$$

The exchange-correlation functional is exactly defined as [47]:

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho]. \tag{2.103}$$

The Euler-Lagrange equation (equation 2.97) for this problem is [47]:

$$\begin{aligned}
 \mu &= \frac{\partial F[\rho]}{\partial \rho(\vec{r})} + v_{ext}(\vec{r}) \\
 &= \frac{\partial T_s[\rho]}{\partial \rho(\vec{r})} + v_{eff}(\vec{r}),
 \end{aligned} \tag{2.104}$$

where:

$$\begin{aligned}
 v_{eff}(\vec{r}) &= \frac{\partial J[\rho]}{\partial \rho(\vec{r})} + \frac{\partial E_{xc}[\rho]}{\partial \rho(\vec{r})} + v_{ext}(\vec{r}) \\
 &= \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' + v_{xc}(\vec{r}) + v_{ext}(\vec{r}) \\
 &= v_H(\vec{r}) + v_{xc}(\vec{r}) + v_{ext}(\vec{r}).
 \end{aligned} \tag{2.105}$$

Here $v_{xc}(\vec{r}) = \frac{\partial E_{xc}[\rho]}{\partial \rho(\vec{r})}$ and $v_H = \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}'$ are known as the exchange-correlation potential and Hartree potential respectively [47]. Interestingly, equation 2.104 is the same as equation 2.97. What this means is that the density of the non-interacting system with external potential $v_{eff}(\vec{r})$ is the same as the density of the interacting system. Overall, given, $v_{eff}(\vec{r})$ and the constraint $\int \rho(\vec{r}) d^3\vec{r} = \eta$, we obtain a $\rho(\vec{r})$ that satisfies equation 2.104 by solving η one-electron equations:

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}(\vec{r}) \right] \psi_i(\vec{r}) = \epsilon_i \psi_i(\vec{r}), \tag{2.106}$$

where

$$\rho(\vec{r}) = \sum_i^\eta \sum_\sigma |\psi(\vec{r}, \sigma)|^2. \tag{2.107}$$

Equations 2.106, 2.107 and 2.105 are known as the Kohn-Sham equations [47]. These have a very similar form to the Roothan equations in Section 2.5.5 and we can solve them with a similar approach.

2.9.3 The Kohn-Sham equations

To practically solve equation 2.106 a Galerkin approximation of atomic orbitals is often used. Each molecular orbital ψ_i is formed from a linear combination of K known atomic orbital (AO) basis functions $\{\phi_j(\vec{r}) | j = 1, 2, \dots, K\}$. We omit the full details on the full formulation, which follows similarly to Section 2.5.5 and can be found in many texts - such as [47]. We write the resulting conclusions here. First, the Fock matrix of equation 2.106 is written as:

$$f_{KS} = -\frac{1}{2}\nabla^2 + v_{ne}(\vec{r}) + v_{xc}(\vec{r}) + v_H(\vec{r}), \tag{2.108}$$

such that:

$$f_{KS}\psi_i(\vec{r}) = \epsilon_i\psi_i(\vec{r}). \tag{2.109}$$

In the same approach as equation 2.59, inserting the definition of $\psi_i(\vec{r})$ (equation 6.13) and left multiplying by $\phi_\nu^*(\vec{r})$ and integrating over \vec{r} yields:

$$F_{\mu\nu} = \int d\vec{r}_1 \phi_\mu^*(\vec{r}_1) f_{KS} \phi_\nu(\vec{r}_1) = \langle \phi_\mu | f_{KS} | \phi_\nu \rangle. \tag{2.110}$$

Leading to:

$$\sum_{\nu}^K F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_{\nu}^K S_{\mu\nu} C_{\nu i}, \quad (2.111)$$

where S and C have the same definition as equations 2.61 and 6.13. In practice, the Fock matrix is often decomposed as:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + J_{\mu\nu}^{xc} + V_{\mu\nu}^{xc}, \quad (2.112)$$

where:

$$H_{\mu\nu}^{core} = \int d\vec{r}_1 \phi_{\mu}^*(\vec{r}_1) \left(-\frac{1}{2} \nabla^2 + v_{ne}(\vec{r}) \right) \phi_{\nu}(\vec{r}_1), \quad (2.113a)$$

$$J_{\mu\nu} = \int d\vec{r}_1 \phi_{\mu}^*(\vec{r}_1) \left(v_H(\vec{r}) \right) \phi_{\nu}(\vec{r}_1) = \sum_a^K \sum_b^K \gamma_{ab} (\phi_{\mu} \phi_{\nu} | \phi_a \phi_b), \quad (2.113b)$$

$$V_{\mu\nu}^{xc} = \int d\vec{r}_1 \phi_{\mu}^*(\vec{r}_1) \left(v_{xc}(\vec{r}) \right) \phi_{\nu}(\vec{r}_1). \quad (2.113c)$$

Here γ_{ab} is the density matrix defined in the atomic orbital basis. A full definition is provided later in Chapter 6 - see equation 6.21. The only undefined term is the exchange-correlation potential $v_{xc}(\vec{r})$ (equation 2.113c), which is approximated in DFT by what is known in the field as ‘functionals’. There are many options to use. A discussion on these is outside the scope of this thesis. For an interested reader, we recommend a recent extensive assessment of 200 density functionals by Mardirossian and Head-Gordon [66]. We note one technical point when approximating $v_{xc}(\vec{r})$. The integral in equation 2.113c usually cannot be calculated analytically. Instead, a numerical integration is performed over grids that have been optimized for these types of calculations [67, 68]. Once evaluated (equation 2.113c), the same self-consistent field approach as Hartree-Fock can be used to solve the problem. In fact most quantum chemistry codes use the same underlying code for their Hartree-Fock and DFT solvers. The only difference in each method is how the Fock matrix is constructed in each SCF optimization loop.

We note DFT has been extended to include spin and is known as spin density functional theory [69]. This is outside the scope of this thesis. A recent overview can be found in [70].

2.9.4 Self-interaction and delocalization error in DFT

In Hartree-Fock, we saw in Section 2.5.4 that the spurious self-interaction term due to the Coulombic attraction of an electron and itself exactly cancels out with the Exchange term. DFT also contains a Coulombic term (equation 2.94b) split into $J[\rho]$ (equation 2.102) and $E_{xc}[\rho]$ (equation 2.103), which has the same problem of

self-interaction. The correction for this term is exactly treated by the true exchange-correlation functional (equation 2.103). However, in all practical calculations, DFT approximates the exchange and correlation with an approximate exchange-correlation functional. The self-interaction error is now not guaranteed to cancel, which may introduce significant error in some calculations [24, 71]. The definition for this error was later expanded and is now more generally known as delocalization error [72]. An analysis of the history and causes of this error is outside the scope of this thesis, a great review on the topic can be found in [73]. The causes and treatment of delocalization error is still an active area of research.

Chapter 3

Quantum Chemistry on Quantum Computers

Since Feynman’s proposal of using quantum computers to study quantum systems [7], there have been numerous algorithms proposed. Many of these use the quantum Fourier transform (QFT) [74, 75] and quantum phase estimation (QPE) [76, 77] as subroutines. These often require deep circuits implying the need for a fault-tolerant quantum computer. However, current quantum computing platforms have significant constraints, such as short coherence times and low qubit numbers. We are currently in the Noisy Intermediate-Scale Quantum (NISQ) technology era, where computers of roughly 10 – 500 qubits are available [78]. These computers are expected to execute certain tasks that outperform the current capabilities of current classical computers [78]. The term “quantum advantage” is used to denote computations involving a quantum device that cannot be performed classically within reasonable amounts of time and energy resources [79].

In 2019, researchers at Google performed their ‘quantum supremacy’ experiment [80]. They showed that pseudorandom quantum circuits are difficult to simulate classically - and what took 200 seconds on their Sycamore quantum processor would take approximately 10,000 years on the world’s most powerful supercomputers to date. The practical use of random quantum circuits is unknown; however, the experiment was an important milestone in the field as a quantum processor was shown to be so complex that its behaviour could not be predicted by conventional computers [81]. Since their result was released, there has been fierce debate about whether supremacy was actually shown. Shortly after Google published their paper, a blog post¹ by IBM’s competing research group stated “an ideal simulation of the same task can be performed on a classical system in 2.5 days and with far greater fidelity” details of which they give in [82]. Even more doubt has recently been cast on random circuit sampling as a scalable approach to prove quantum supremacy, where a polynomial-time classical algorithm for sampling the distribution of a noisy random quantum circuit in the regime of anti-concentration to within inverse polynomial total variation distance was introduced [83].

¹<https://www.ibm.com/blogs/research/2019/10/on-quantum-supremacy/>

Another quantum advantage experiment was recently performed by Jian-Wei Pan’s group using a Jiuzhang photonic quantum device performing Gaussian boson sampling (GBS) with 50 indistinguishable single-mode squeezed states [84]. Advantage was observed in the time complexity of sampling a Torontonian matrix (determines the probability distribution of measurement outcomes) that scales exponentially with the photon click output [79, 84].

The main takeaway of both these supremacy experiments is small quantum processors cannot be simulated well by the world’s largest supercomputers. A more important question is how do we take advantage of these machines to solve useful problems? This is often coined “practical quantum advantage”. To answer this, many different groups around the world are trying to understand the capabilities and limitations of NISQ computers, given small qubit numbers, lack of error-correction and short coherence times [85]. Algorithms that offload as much of the computation as possible to a conventional computer are therefore natural to investigate for use on such devices. These are often known as ‘hybrid quantum-classical algorithms’.

The variational quantum eigensolver (VQE) is one of these and will be introduced in section 3.2. These hybrid quantum algorithms only need to perform a short sequence of operations - rather than long sequences required by algorithms designed for use with fault-tolerant quantum computers, such as the phase estimation algorithm (Section 3.1). The importance of hybrid algorithms is they circumvent some of the shortfalls of current hardware and allow problems to be approached on present-day devices.

In this chapter, we review the phase estimation and variational quantum eigensolver algorithms. We then comment on the possibility of near-term quantum advantage with a focus on application to chemistry problems.

3.1 Phase estimation

Quantum phases are important in many quantum algorithms and we are going to see how they can be estimated on a quantum device via the phase estimation algorithm. This algorithm was independently proposed by Kitaev [77] and Abrams and Lloyd [76].

The quantum phase estimation algorithm is a method for estimating the eigenphase of any unitary operator. Given any Hermitian operator H , one tries to find the eigenvalue E_j such that:

$$e^{iH} |\lambda_j\rangle = e^{iE_j} |\lambda_j\rangle, \quad (3.1)$$

where the Hermitian operator H has been exponentiated to obtain a unitary operator.

3.1. PHASE ESTIMATION

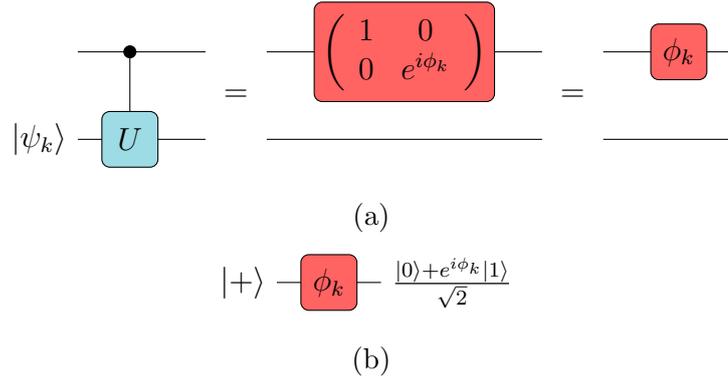


Figure 3.1: (a) Phase kickback when eigenstate of a unitary U , here denoted $|\psi_k\rangle$, is inserted into a controlled version of U , (b) Outcome of phase gate applied to the plus state.

Note that the phase can be rewritten as $e^{iE_j} = e^{i2\pi\phi}$, where ϕ is the phase to be determined via phase estimation.

Before we give an analysis of how this algorithm works, we quickly review how phase kickback works. We can generate a relative phase for a general unitary U by inserting an eigenstate of U denoted $|\psi_k\rangle$ into a controlled version of U . As $U|\psi_k\rangle = e^{i\phi_k}|\psi_k\rangle$, the state on the system register is left unaltered. However, if the state on the control qubit has any $|1\rangle$ component a relative phase is generated on these states. Figure 3.1a illustrates this and shows how the phase is kicked back onto the control qubit. Figure 3.1b summarises the action of a phase gate on the plus state $|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$. These ideas underpin the phase estimation algorithm, next we see how these phases are estimated.

Figure 3.2 illustrates the first part of the phase estimation algorithm. The output quantum state state of the ancilla register in Figure 3.2 is [86]:

$$\begin{aligned} \frac{1}{2^{t/2}} \left(|0\rangle + e^{2\pi i 2^{t-1}\phi} |1\rangle \right) \left(|0\rangle + e^{2\pi i 2^{t-2}\phi} |1\rangle \right) \dots \left(|0\rangle + e^{2\pi i 2^0\phi} |1\rangle \right) \\ = \frac{1}{2^{t/2}} \sum_{k=0}^{2^t-1} e^{2\pi i \phi k} |k\rangle. \end{aligned} \quad (3.2)$$

Suppose the unknown phase ϕ can be expressed in exactly t -bits as: $\phi = 0.\phi_0\phi_1 \dots \phi_{t-1}$. Then the output state can be written in product form as:

$$\frac{1}{2^{t/2}} \left(|0\rangle + e^{2\pi i (0.\phi_{t-1})} |1\rangle \right) \otimes \left(|0\rangle + e^{2\pi i (0.\phi_{t-2}\phi_{t-1})} |1\rangle \right) \otimes \dots \otimes \left(|0\rangle + e^{2\pi i (0.\phi_0\phi_1 \dots \phi_{t-1})} |1\rangle \right). \quad (3.3)$$

3.1. PHASE ESTIMATION

This looks like the output of a quantum Fourier transform [75, 86]:

$$\text{QFT: } |j\rangle \mapsto \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{i\frac{2\pi}{2^n}jk} |k\rangle. \quad (3.4)$$

This is easier to see by re-writing the output of equation 3.4 in product form [86]:

$$\begin{aligned} \text{QFT: } |j\rangle &\mapsto \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{i2\pi j \frac{k}{2^n}} |k\rangle \\ &= \frac{1}{\sqrt{2^n}} \sum_{k_0=0}^1 \dots \sum_{k_{n-1}=0}^1 e^{i2\pi j \left(\sum_{l=0}^{n-1} k_l 2^{-(l+1)} \right)} |k_0, k_1, \dots, k_{n-1}\rangle \\ &= \frac{1}{\sqrt{2^n}} \sum_{k_0=0}^1 \dots \sum_{k_{n-1}=0}^1 \bigotimes_{l=0}^{n-1} e^{i2\pi j k_l 2^{-(l+1)}} |k_l\rangle \\ &= \frac{1}{\sqrt{2^n}} \bigotimes_{l=0}^{n-1} \left[\sum_{k_l=0}^1 e^{i2\pi j k_l 2^{-(l+1)}} |k_l\rangle \right] \\ &= \frac{1}{\sqrt{2^n}} \bigotimes_{l=0}^{n-1} \left[|0\rangle + e^{i2\pi j 2^{-(l+1)}} |1\rangle \right] \\ &= \frac{\left(|0\rangle + e^{2\pi i(0.j_{n-1})} |1\rangle \right) \otimes \left(|0\rangle + e^{2\pi i(0.j_{n-2}j_{n-1})} |1\rangle \right) \otimes \dots \otimes \left(|0\rangle + e^{2\pi i(0.j_0j_1\dots j_{n-1})} |1\rangle \right)}{2^{n/2}}, \end{aligned} \quad (3.5)$$

where we have used the fractional binary notation:

$$\frac{b}{2^n} = 0.b_0b_1b_2\dots b_{n-1} = \sum_{k=0}^{n-1} b_k 2^{-(k+1)}. \quad (3.6)$$

Here b is a binary integer in the range 0 to $2^n - 1$. Using this, it should be clear that the final term in equation 3.5 is the same as the final term of equation 3.3 when $n = t$. In order to estimate the phase, the last part of the algorithm requires performing the inverse quantum Fourier transform on the ancilla register, resulting in:

$$\frac{1}{2^{t/2}} \sum_{k=0}^{2^t-1} e^{2\pi i\phi k} |k\rangle_a |\psi\rangle_s \mapsto |\phi\rangle_a |\psi\rangle_s, \quad (3.7)$$

where $|\phi\rangle_a$ is the ancillary quantum state. Measuring the ancillary register in the computational basis, therefore, gives us the phase ϕ exactly in binary. However, so far we have assumed the phase could exactly be written in t -ancillary bits. In practice,

3.1. PHASE ESTIMATION

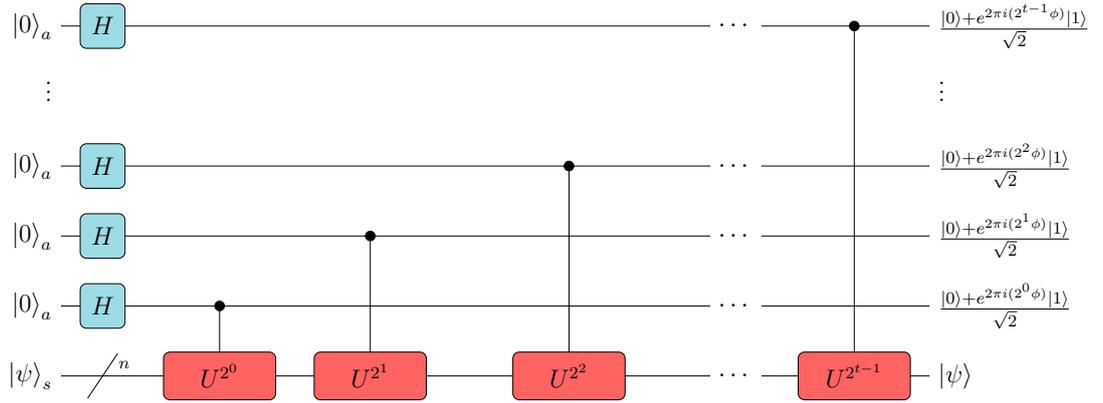


Figure 3.2: First part of the phase estimation algorithm, which is completed with an inverse Fourier transform (QFT^\dagger) applied to the ancillary register. The ancillary and system qubit registers are denoted with a and s respectively.

this assumption cannot be made and so we write:

$$\frac{1}{2^{t/2}} \sum_{k=0}^{2^t-1} e^{2\pi i \phi k} |k\rangle_a |\psi\rangle_s \mapsto |\tilde{\phi}\rangle_a |\psi\rangle_s, \quad (3.8)$$

where $|\tilde{\phi}\rangle_a$ is a good estimator of $|\phi\rangle_a$ when measured in the computational basis. We omit the proof of why $|\tilde{\phi}\rangle_a$ is a good estimator, which is derived in [86]. We summarise the result, which proved the number of ancilla qubits required to obtain the phase to a precision ϵ scales as $\mathcal{O}(\frac{1}{\log_2(\epsilon)})$. Intuitively this makes sense, as each ancillary qubit provides one bit of information. The number of controlled unitaries is doubled for each ancillary qubit and therefore for precision ϵ will scale as $\mathcal{O}(\frac{1}{\epsilon})$ [87].

A component of the phase estimation we have not addressed is how an eigenstate $|\psi\rangle_s$ is required to be input on the system register. What would happen if some other general state, such as $|\Omega\rangle_s$, was prepared? We can write this general state in the eigenbasis of the unitary U as:

$$|\Omega\rangle_s = \sum_i^{2^n} c_i |\psi_i\rangle_s. \quad (3.9)$$

Applying the phase estimation algorithm to this state results in:

$$\sum_i^{2^n} c_i |\bar{0}\rangle_a |\psi_i\rangle_s \mapsto \sum_i^{2^n} c_i |\tilde{\phi}_i\rangle_a |\psi_i\rangle_s. \quad (3.10)$$

This superposition state contains approximated eigenphases of the different eigenstates $|\psi_i\rangle$. The probability of measuring $|\tilde{\phi}_i\rangle_s$ and thus $|\psi_i\rangle_s$ on the system register, will be $|\langle \psi_i | \Omega \rangle|^2$. Therefore, to approximate an eigenphase of a particular state $|\psi_i\rangle_s$ - such as the ground state - with high probability will depend on its overlap with

the input state $|\Omega\rangle$. A central assumption for phase estimation is that there is access to some oracle capable of producing a quantum state with “good overlap” with the desired state to estimate the phase of. Here good means overlap that is polynomial in the size of the system [88].

In 2005, Aspuru-Guzik, Dutoi, Love and Head-Gordon proposed applying phase estimation to quantum chemistry problems [89]. This was the first method introduced to estimate quantum chemical states on a quantum computer. In their numerical results, the Hartree-Fock state was used as the input state for phase estimation [89]. This input was used in subsequent works [42, 90], where in all cases the single Slater determinant reference state had good overlap with the ground state. However, the systems considered in all these works are very small. The Van Vleck catastrophe [91] refers to an expected exponential decline in the quality of trial wavefunctions (measured by overlap with the true wavefunction of a system) as a function of size [88]. The poly overlap of these classical methods with the true ground state must be considered in the context of large basis set sizes K . Recently Chan *et al.* explored this problem and found that exponential quantum advantage for quantum chemistry problems using current quantum algorithms may not be possible, instead it seems like they are more likely to provide polynomial speed-ups [92]. However, no proof was given just numerical evidence. The problem of generating good trial states is therefore still an important open question in the field.

Finally, we note it is well-known that the lowest eigenvalue problem is Quantum Merlin Arthur (QMA) complete [41, 42]. The requirement of an initial state with good overlap with the true ground state is therefore highly non-trivial and this condition has major implications for the practical use of the QPE algorithm for quantum chemistry applications. Ignoring this issue, today’s quantum computers are noisy and QPE requires complicated quantum circuits. An Implementation of QPE on currently available NISQ hardware is extremely limited, as seen in the experimental work of Mohammadbagherpoor *et al.* [93] and O’Malley *et al.* [90]. The variational quantum eigensolver algorithm was introduced to mitigate some of these concerns and is implementable on currently available QPUs [94].

3.2 The variational quantum eigensolver

The variational quantum eigensolver (VQE) algorithm was proposed in an experimental paper by Prezzo *et al.* [94] and later developed by McClean *et al.* [95]. The algorithm was developed to find the lowest eigenvalue of a Hamiltonian on near-term (non-fault tolerant) quantum computers [94]. Since then it has been extended, to allow excited states (higher eigenvalues) to be determined [96, 97]. The crux of the algorithm stems from the Rayleigh-Ritz variational principle [98–100]. This basically

states that for a time-independent Hamiltonian the ground state is bounded by the minimum eigenvalue:

$$E_0 \leq \frac{\langle \Psi(\theta) | H | \Psi(\theta) \rangle}{\langle \Psi(\theta) | \Psi(\theta) \rangle}. \quad (3.11)$$

The algorithm works by first preparing a trial (or ansatz) quantum state, via a parametrised quantum circuit $U(\vec{\theta})$. The energy is then minimized with respect to the parametrization $\vec{\theta}$:

$$E_0 = \underset{\vec{\theta}}{\operatorname{argmin}} \langle \bar{0} | U^\dagger(\vec{\theta}) H U(\vec{\theta}) | \bar{0} \rangle, \quad (3.12)$$

where a classical optimizer is used to update $\vec{\theta}$. Upon convergence, the expectation value is an upper-bound of the ground state. Importantly, this technique can get trapped in local minima, which is a common problem in electronic structure problems.

In a standard VQE experiment, it is generally not possible to measure H at once. Instead, operators that can be measured on a quantum computer must be used, which are tensor products of spin operators. These are the Pauli operators: $\sigma \in \{I, Z, X, Y\}$, where:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (3.13a)$$

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (3.13b)$$

$$Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad (3.13c)$$

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (3.13d)$$

Importantly, all different possible arrangements of n -fold tensor products of Pauli operators (4^n possibilities) form a basis for a Hilbert space $\mathcal{H} \in \mathcal{C}^{2^n \times 2^n}$ and so any Hermitian operator acting on such a Hilbert space can be written as some linear combination of Pauli operators:

$$H = \sum_i^{|H|} c_i P_i = \sum_i^{|H|} c_i \left(\bigotimes_{j=0}^{n-1} \sigma_j^{(i)} \right) = \sum_i^{|H|} c_i \left(\sigma_0^{(i)} \otimes \sigma_1^{(i)} \otimes \dots \otimes \sigma_{n-1}^{(i)} \right). \quad (3.14)$$

For each single qubit Pauli operator σ_j , each subscript j indexes the qubit the operator acts on. The notation $|H|$ means the number of Pauli operators P in H .

3.2. THE VARIATIONAL QUANTUM EIGENSOLVER

Year	Reference	System	Ansatz	Max qubits	Platform	Hardware Vendor
2014	Peruzzo <i>et al.</i> [94]	HeH ⁺	UCC	2	Silicon Photonic	Bristol
2015	Shen <i>et al.</i> [101]	HeH ⁺	UCC	2	Trapped ion	
2015	Google Quantum [90]	H ₂	UCC	2	Superconducting	Google
2016	Santagati <i>et al.</i> [102]	H ₂ , H ₃ , H ₃ ⁺ , H ₄	Parametrized Hamiltonian	2	Silicon Photonic	
2017	Kandala <i>et al.</i> [103]	H ₂ , LiH, BeH ₂ , Heisenberg model	HEA	6	Superconducting	IBM
2017	Colless <i>et al.</i> [15]	H ₂ (excited states)	HEA	2	Superconducting	UCB, LBNL
2018	Hempel <i>et al.</i> [36]	H ₂ , LiH	UCC	3	Trapped ion	
2018	Kandala <i>et al.</i> [104]	H ₂ , LiH (magnetism)	HEA	4	Superconducting	IBM
2018	Ryabinkin <i>et al.</i> [105]	H ₂ , LiH	Qubit CC	4	Superconducting	
2019	Nam <i>et al.</i> [106]	H ₂ O	UCC	4	Trapped ion	IonQ
2019	McCaskey <i>et al.</i> [107]	NaH, RbH, KH	UCC and HEA	4	Superconducting	IBM, Rigetti
2019	Gao <i>et al.</i> [108]	Li superoxide dimer	qUCCCD, R _y	2	Superconducting	IBM
2019	Smart & Mazziotti [109]	H ₃	custom	3	Superconducting	IBM
2020	Google AI Quantum [110]	H ₆ , H ₈ , H ₁₀ , H ₁₂ , HNNH	Hartree-Fock	12	Superconducting	Google
2020	Gao <i>et al.</i> [111]	phenylsulfonyl-carbazole TADF emitters	R _y	2	Superconducting	IBM
2021	Kawashima <i>et al.</i> [112]	H ₁₀	qubit-CC	3	Trapped ion	IonQ
2021	Rice <i>et al.</i> [113]	LiH (dipole moment)	HEA	4	Superconducting	IBM
2022	Eddins <i>et al.</i> [114]	H ₂ O	Entanglement Forging	5	Superconducting	IBM
2022	Motta <i>et al.</i> [115]	H ₃ S ⁺	Entanglement Forging	6	Superconducting	IBM
2022	Yamamoto <i>et al.</i> [116]	crystalline iron model	uCCSD-PBC	2	Superconducting	IBM
2022	Kirsopp <i>et al.</i> [117]	Oxazine derivatives	YXXX	4	Superconducting, Trapped Ion	IBM, Quantinuum
2022	Khan <i>et al.</i> [118]	CH ₃ · (methyl radical)	UCCSD	6	Trapped ion	Quantinuum
2022	O'Brien <i>et al.</i> [119]	Cyclobutene ring	upCCD	10	Superconducting	Google
2022	Zhao <i>et al.</i> [120]	Li ₂ O	oo-upCCD	12	Trapped ion	IonQ
2023	Weaving <i>et al.</i> [121]	HCl	HEA	3	Superconducting	IBM

Table 3.1: Summary of different experimental realizations of VQE.

The linearity of expectation values allows the expected value of H to be determined as:

$$\langle H \rangle = \sum_i c_i \langle P_i \rangle. \quad (3.15)$$

Thus, in a basis of Pauli operators equation 3.12 becomes:

$$E_0 = \operatorname{argmin}_{\vec{\theta}} \sum_i \left[c_i \left(\langle \bar{0} | U^\dagger(\vec{\theta}) P_i U(\vec{\theta}) | \bar{0} \rangle \right) \right]. \quad (3.16)$$

An actual implementation of Equation 3.16 has many components that each impact the design and cost of running a VQE simulation. Table 3.1 summaries many of the experimental realizations of the algorithm to date. An analysis of each is outside the scope of this thesis, but we provide an introduction to some of the important considerations when implementing VQE. For brevity, we only provide a discussion in the context of applying the algorithm to quantum chemistry. The main considerations in VQE are:

1. Which fermion to qubit mapping to use?
2. How to prepare a parametrised quantum state.
3. What optimizer to use to minimize the energy with respect to the Ansatz parameters?
4. How to measure the expectation value of H at each iteration of the algorithm.

We give a brief discussion of points 1-3 in the following subsections and leave the last point for Chapter 4, as it is of primary interest to this thesis.

3.2. THE VARIATIONAL QUANTUM EIGENSOLVER

Citation	Encoding	Qubits	Gates
Jordan-Wigner [123]	Jordan-Wigner	K	$\mathcal{O}(K)$
Bravyi-Kitaev <i>et al.</i> [34, 125]	Bravyi-Kitaev	K	$\mathcal{O}(\log K)$
Bravyi-Kitaev <i>et al.</i> [124]	\mathbb{Z}_2 symmetries	$K - \mathcal{O}(1)$	$\mathcal{O}(K)$
Bravyi-Kitaev <i>et al.</i> [124]	Parity	$K - \frac{K}{\eta}$	$\mathcal{O}(K^3)$
Jiang <i>et al.</i> [126]	Ternary Tree	K	$\mathcal{O}(\log K)$
Steudtner-Wehner [127]	Segment	$K - \frac{K}{2\eta}$	$\mathcal{O}(K^2)$
Babbush <i>et al.</i> [128]	Configuration interaction matrix	$\mathcal{O}(\eta \log K)$	$\mathcal{O}(K)$
Kirby <i>et al.</i> [129]	Optimal-Degree	$\mathcal{O}(\eta^2 \log^4 K)$	$\mathcal{O}(\eta^2 \log^5 K)$
Moll <i>et al.</i> [130]	Subspace Projection	$\lceil \log \binom{K}{\eta} \rceil$	$\mathcal{O}(2^K K)$
Harrison <i>et al.</i> [131]	Clifford permutation	$K - 1$	$\mathcal{O}(K)$
Harrison <i>et al.</i> [131]	Non-Clifford Permutation ($T \leq K$) Toffolis	$\lceil \log \binom{K}{\eta} \rceil$	$\mathcal{O}(4^T K)$
Harrison <i>et al.</i> [131]	General Non-Clifford Permutation	$\lceil \log \binom{K}{\eta} \rceil$	$\mathcal{O}(4^K K)$

Table 3.2: Summary of different works that encode a second-quantized fermionic Hamiltonian in qubits. “Gates” refers to the number of one and two qubit gates to implement an encoding of a conjugate pair of fermionic creation and annihilation operators. The Configuration interaction matrix is an exception, where the gate cost is the cost of the sparse oracle utilized by that particular method. Here K is the number of spin-orbitals (modes) and η the number of fermions.

3.2.1 Fermion to qubit mapping

To simulate systems described in the second quantization, a mapping from indistinguishable fermions to distinguishable qubit operators is required. This requires a mapping from the fermionic Fock space to the Hilbert space of qubits, such that each fermionic state can be represented by a qubit state. We do not discuss first quantized approaches as it is outside the scope of this thesis. Further information can be found in [122].

The most common mapping methods are: Jordan-Wigner [123], parity [124] and Bravyi-Kitaev [34, 125]. Each mapping results in Pauli operators with certain weights (maximum number of non-identity single qubit Pauli matrices in the operator). Table 3.2 summarises the different methods.

It is currently unknown which encoding method is the most noise-resilient - i.e. best for NISQ experiments. Tranter *et al.* did a numerical study comparing the Jordan-Wigner and Bravyi-Kitaev mappings and found that the BK transform was at least as efficient as the JW in finding the ground states of molecular systems and often more efficient [132].

We note the qubit cost for molecular problems can be reduced by tapering off \mathbb{Z}_2 symmetries, including those based on the point group of the molecule under consideration [124, 133].

3.2.2 Ansätze

The choice of Ansatz, to model a given electronic structure wavefunction on a qubit register, often requires a trade-off between expressibility and experimental feasibility. Expressibility in this context means the ability of parameterized quantum circuits to produce quantum states that represent the η -particle Hilbert space well [134]. The goal is to have parameterized circuits with minimum depth that maximize accuracy. There are two flavours of Ansätze known as “hardware efficient” and “chemically inspired”.

Hardware efficient Ansätze (HEA) are composed of repeated blocks of parameterized gates that are most suited (easy to implement) to the available quantum hardware [12]. Such approaches have been used in different VQE experiments to compute the ground state energies of small molecular systems [103, 104]. However, these approaches are agnostic to the chemical problem being simulated [135]. This results in a few drawbacks. First, is the issue of so-called “Barren plateaus” of the variational parameter landscape, where the derivative of the cost function (in this case the derivative of the energy with respect to the Ansatz parameters) is close to zero [136]. Another problem with hardware efficient Ansätze, are physical symmetries, such as electron number and parity, are not accounted for, which increases the size of the search space. Several approaches have been proposed to mitigate these problems [135, 137–139]. We note a recent approach proposed by the Mayhall group, called ctrl-VQE, which is a very unique way of thinking about hardware efficient Ansätze [140]. Rather than working with a parametrised quantum circuit, a parametrised laboratory-frame pulse representation is optimized. They studied the bond dissociation of H_2 and HeH^+ , on a quantum computer, to within chemical accuracy and studied LiH at fixed bond length [140].

Chemically inspired Ansätze are physically motivated by the problem at hand. When Peruzzo *et al.* first introduced the VQE algorithm, they used a unitary coupled cluster singles doubles (UCCSD) Ansatz [94]. This Ansatz was inspired by the classical non-unitary coupled cluster method [40]. The trial wavefunction is prepared from a reference state, usually the Hartree-Fock state $|\Phi_{HF}\rangle$, by applying the exponentiated excitation operator [141, 142]:

$$U(\vec{\theta}) = e^{T-T^\dagger} \quad (3.17)$$

where:

$$|\psi(\vec{\theta})\rangle = U(\vec{\theta}) |\Phi_{HF}\rangle. \quad (3.18)$$

The operator $T = \sum_{k=1}^d T_k$ is the excitation operator, which is usually truncated at

some excitation level d . The UCCSD is realized at $d = 2$, and T will be:

$$T_{UCCSD} = T_1 + T_2 = \sum_{\substack{i \in occ \\ a \in virt}} t_i^a \alpha_a^\dagger \alpha_i + \sum_{\substack{ij \in occ \\ a, b \in virt}} t_{ij}^{ab} \alpha_a^\dagger \alpha_b^\dagger \alpha_i \alpha_j. \quad (3.19)$$

Here t_i^a and t_{ij}^{ab} are the cluster amplitudes (real coefficients), *occ* and *virt* denote orbital indices that are occupied and unoccupied with respect to the Hartree-Fock state [141, 142]. A Trotter expansion is often used to decompose this operator as:

$$T_{UCC} = \sum_{k=1}^d T_k = \left(\prod_{k=1}^d e^{T_k - T_k^\dagger / a} \right)^a, \quad (3.20)$$

where a trotter number of $a = 1$ is usually sufficient for VQE [141]. Note the fermionic operators here can be converted to qubit operators using any of the transformations discussed in Section 3.2.1. The number of parameters for UCCSD scales as $\mathcal{O}(K^2\eta^2)$ for K spin orbitals and η electrons [142]. The number of gates under the Jordan-Wigner transform scales as $\mathcal{O}(K^2\eta^2)$ and Bravyi-Kitaev as $\mathcal{O}(K^3\eta^2)$ (assuming a single Trotter step) [142].

Even though the scaling of UCCSD is tractable, the circuit depth is still a challenge for NISQ devices and optimization may be hard. As many of the excitation terms can have near-zero contribution to the correlation energy, many variants of the UCC ansatz have been proposed that improve this scaling by choosing excitation operators that constitute the Ansatz more efficiently [143]. Examples include: Unitary Pair CC with Generalized Singles and Doubles (k -UpCCGSD) [144], Orbital Optimized UCC (OO-UCC) [145], Adaptive Derivative-Assembled Pseudo-Trotter Ansatz Variational Quantum Eigensolver (ADAPT-VQE) [146], qubit-ADAPT-VQE [147], Qubit Coupled Cluster (QCC) [105]. Other chemically inspired Ansatz impose properties of the system such as the particle number, spin, and time-reversal symmetries [148, 149]. A full analysis of each of these is outside the scope of this thesis.

3.2.3 Classical optimization in VQE

For variational quantum algorithms to be successful it is not enough for the Ansatz to be expressible enough to contain the solution. The cost landscape must also have large enough cost gradients to allow this solution to be found [150]. Take the Ansatz for random parameterized quantum circuits:

$$U(\vec{\theta}) = U(\theta_0, \theta_1, \dots, \theta_{L-1}) = \prod_{j=0}^{L-1} U_j(\theta_j) W_j, \quad (3.21)$$

where $\vec{\theta}$ gives a particular parametrization. Here $U_j(\theta_j) = e^{-i\theta_j V_j}$, V_j is a Hermitian operator and W_j is a general unitary operator that does not depend on the Ansatz parametrization [136]. The component of the gradient corresponding to the parameter θ_j is determined by the partial derivative: $\partial_j E = \frac{\partial E}{\partial \theta_j}$ [136, 150]. This can be determined by the parameter shift rule [151–153]. It was shown in [136] that:

$$\langle \partial_j E \rangle = 0 \quad \forall j. \quad (3.22)$$

In words, the gradient of the cost function averages out to zero rather than being biased in a particular direction [150]. An unbiased cost landscape can be trainable, but this depends on the extent to which the gradient fluctuates away from zero [150]. For a cost (in our case energy) to exhibit a barren plateau, the gradient must vanish exponentially with the number of qubits n . This is related to the variance of the partial derivative vanishing exponentially - $Var[\partial_j E] \in \mathcal{O}(\frac{1}{2^{np}})$ for any integer $p > 0$ [150, 154]. To determine the minimization direction requires exponentially many measurements and makes the optimization problem effectively intractable for large problem sizes [150, 154]. This issue is also present for gradient-free optimizers [155] and can be understood by these methods usually relying on sampling the cost landscape. If the variance across this landscape is too small then it is impossible to accurately progress through the optimization problem [87]. One approach to avoid these landscapes, where the gradient becomes vanishingly small, is to use a structured initial guess [136].

The barren plateau problem has also been linked to the presence of high levels of entanglement in quantum circuits [156] and hardware noise in a given computation [157]. The unitary coupled cluster Ansatz for instance shows these noise-induced barren plateaus [157]. Therefore, physically motivated Ansätze may not be beneficial over hardware efficient Ansätze. Grimsley *et al.* have shown that ADAPT-VQE can mitigate some of the problems described here [158]. The quantum natural-gradient-based strategy is also considered resilient to barren plateaus [159]. However, much more work is still required on the problem of barren plateaus.

Ignoring the issue of Barren plateaus, for a given VQE problem what classical optimizer should be used? Pellow-Jarman *et al.* recently performed a numerical study investigating different classical optimizers and found under realistic noise levels there was no clear winner between gradient-based and gradient-free approaches [160]. In very low noise regimes, they found that gradient-based approaches outperformed the gradient-free optimizers.

3.3 VQE and quantum advantage

Theoretically, when studying the second quantized molecular Hamiltonian for different molecular systems using VQE, the problem will scale polynomially with the number of spin orbitals K (equation 2.73) [94, 95]. However, this hides a lot of extra complexity contained within the algorithm, which limits the ability to achieve quantum advantage. We note this has been discussed in numerous reviews [16, 79, 87, 161, 162]. We summarise some of the main points here.

First, there is the issue of a large pre-factor caused by circuit sampling [87]. At a high level, the number of samples in VQE that is required to achieve a precision ϵ scales as $\mathcal{O}(K^8/\epsilon^2)$ [163]. The consequence of this is the sheer number of measurements required to estimate a single energy can take many years. How to handle this problem will be discussed at length in the next chapter, but the main takeaway here is a naive measurement strategy in VQE makes any practical advantage redundant. Compared with phase estimation, VQE trades circuit depth and number of qubits for a higher number of measurements and circuit repetitions. Wang *et al.* show that QPE requires $\mathcal{O}(1)$ repetitions with quantum circuits scaling inversely with precision ϵ as $\mathcal{O}(\frac{1}{\epsilon})$ [164]. VQE on the other hand has measurements scaling as $\mathcal{O}(\frac{1}{\epsilon^2})$, with circuit depth scaling as $\mathcal{O}(1)$ in precision [164]. We note many other factors affect the scaling of both methods, the metrics presented here just illustrate some of the asymptotic considerations.

VQE also requires solving a classical optimization problem, which has been shown to be NP-hard [165]. This means at worst the optimal solution for a problem is intractable. However, this is expected as all optimization problems can suffer from this issue [166]. As commented on by Tilly *et al.*, the key open question is whether VQE can be optimized in a polynomial number of iterations, via a heuristic, and converge to an approximate but accurate enough solution [87]. However, the presence of barren plateaus means that even if convergence in VQE was theoretically possible, the number of measurements required to obtain the gradient or sample the optimization landscape will scale exponentially under certain Ansätze. Much more work needs to be done on this problem.

One of the major benefits of VQE is it being a noise-resilient quantum algorithm [90, 94], as optimization can effectively adapt to noise [95]. This can be seen with different experimental realizations of VQE on NISQ devices - see Table 3.1. However, it is unknown how this resilience scales when treating larger problems and the quantum circuits become much deeper and more complicated. Saib *et al.* investigated this issue for hardware efficient Ansätze and found the expressibility of the Ansatz did not correlate with performance [167]. This indicates expressibility may not be the best choice for selecting a given Ansatz for VQE applied to chemistry problems. This

was further shown by Dalton *et al.* in a study that investigated how noise affects the Ansatz part of VQE [168]. They conclude that VQE is unlikely to be scalable on near-term hardware without error correction. Moreover, in their concluding remarks they state VQE algorithms are unlikely to provide useful quantum advantage for chemistry applications in the NISQ era of computing. Further research is required on this topic. Error mitigation strategies may offer a route to improve these issues, which we numerically found in a three-qubit study of HCl [121]. However, how these methods scale with problem size can be a problem [169].

Finally, we reiterate a comment made in [16]. This discussion has been presented for VQE applied to quantum chemistry problems. However, the algorithm can be used to find the ground state of any Hamiltonian and so can be applied in many alternate fields. The possibility of quantum advantage may be easier to achieve in these contexts. This remains an open question.

Chapter 4

The Measurement Roadblock to Implementation of VQE

In this chapter, we explore the topic of measurement in the variational quantum eigensolver algorithm, which we will see causes the runtime of the algorithm to become a major obstacle for practical implementation. We first derive where this issue stems from and then provide a brief review of methods that seek to reduce the measurement overhead and thus runtime. We then present our numerical investigation for the unitary partitioning measurement reduction strategy. This was to study the real performance of the method, as opposed to relying on theoretical asymptotic scaling arguments.

4.1 Measurement in quantum mechanics

The third postulate of quantum mechanics states that associated with every observable quantity is a Hermitian operator M [86]. A Hermitian operator is an operator that is equal to its Hermitian conjugate M^\dagger or transpose complex conjugate: $M = M^\dagger = (M^T)^*$ [86]. Two of the most fundamental properties of Hermitian operators are the eigenvalues are real, and that they are diagonalizable. Taking the spectral decomposition of a general Hermitian operator:

$$M = \sum_{i=0}^{d-1} \lambda_i |\phi_i\rangle \langle \phi_i| = \sum_{i=0}^{d-1} \lambda_i M_i, \quad (4.1)$$

where M_i are projection operators onto eigenstates $|\phi_i\rangle$ and must obey the completeness equation $\sum_i M_i^\dagger M_i = \mathcal{I}$ [86]. λ_i are the (real) eigenvalues of M_i that are possible measurement outcomes.

For a general quantum state $|\psi\rangle$, before measurement the probability of measuring $|\phi_k\rangle$ is $p_k = \langle \psi | M_k^\dagger M_k | \psi \rangle$, where $M_k = |\phi_k\rangle \langle \phi_k|$. The state of the system after measurement is [86]:

$$|\psi'\rangle = \frac{M_k |\psi\rangle}{\sqrt{\langle \psi | M_k^\dagger M_k | \psi \rangle}} \quad (4.2)$$

and the eigenvalue λ_k is also returned upon measurement. The projector M_k determines the state after measurement $|\psi'\rangle$ and the denominator in equation 4.2 is the renormalization factor.

In quantum mechanics, the expected (or expectation) value of an operator is given by the weighted average of all possible measurement outcomes. It is not the most probable value of a measurement. The experimental action of measuring a property on a normalized state $|\psi\rangle = \sum_j c_j |\phi_j\rangle$ is formally written as:

$$\begin{aligned} \langle M \rangle_{|\psi\rangle} &= \langle \psi | M | \psi \rangle = \langle \psi | \left(\sum_{i=0}^{d-1} \lambda_i |\phi_i\rangle \langle \phi_i| \right) | \psi \rangle \\ &= \left(\sum_{i=0}^{d-1} \lambda_i |\langle \psi | \phi_i \rangle|^2 \right). \end{aligned} \quad (4.3)$$

This is known as the expectation value of A .

In the context of this work, we are interested in the expectation value of an operator made up of a linear combination of Pauli operators. The different possible arrangements of n -fold tensor products of Pauli operators (4^n) form a basis for a Hilbert space $\mathcal{H} \in \mathbb{C}^{2^n \times 2^n}$ and so any Hermitian operator acting on such a Hilbert space can be written as some linear combination of Pauli operators:

$$H = \sum_i c_i P_i. \quad (4.4)$$

Often, the expectation value of such a given operator cannot be measured directly. However, the linearity of expectation values allows the expected value to be determined as:

$$\langle H \rangle = \sum_i c_i \langle P_i \rangle. \quad (4.5)$$

The expectation value of each Pauli operator $\langle P_i \rangle$ is estimated by preparing M independent copies of a given state ψ and measuring :

$$\begin{aligned} \langle P_i \rangle &= \langle \psi(\vec{\theta}) | P_i | \psi(\vec{\theta}) \rangle \\ &\approx \left(\frac{1}{M} \sum_{j=0}^{M-1} s_j^{(i)} \right) = \langle \tilde{P}_i \rangle, \end{aligned} \quad (4.6)$$

where $s_j^{(i)} \in \{-1, +1\}$. The above expression is exact in the limit that the number of samples $M \rightarrow \infty$. For a finite number of samples, we denote the estimated expectation value as $\langle \tilde{P}_i \rangle$. The only (special) case the expectation value of an operator can be evaluated exactly, with a single sample, is when the wavefunction is an eigenstate of the operator being measured. Note that each $s_j^{(i)}$ is obtained from separate state preparations and all values obtained are thus independent and identically distributed

random variables. In such a scenario it is important to distinguish sample variation and intrinsic variation of an expectation value. We discuss this in detail next.

4.1.1 Expectation variance of individual operators

A quantum state can be thought of as a mathematical object that provides the probability distribution for the outcomes of each possible measurement on a system. In the computational basis, this will be a distribution of different binary states that can be obtained (measured in an experiment) with a certain probability. As we are talking about probability distributions we can analyse certain properties using probability theory.

In quantum mechanics, we are often interested in the expected value of an operator for a given state. Looking at equation 4.3, this can be thought of as the average of all possible outcomes weighted by their likelihood (probability) of occurring. We consider the setting for a quantum state defined on n -qubits, meaning the discrete probability distribution is defined on a finite number of bitstrings¹. This means there will always be an inherent population mean and variance, where the variance describes how spread the distribution is. For a given state (distribution), the variance of the expectation value of an operator is calculated from the expected value of the squared deviation from the mean of an operator. For a general operator O this is written as [170]:

$$\begin{aligned} \text{Var}[O] &= \text{Cov}[O, O] \\ &= \langle (O - \langle O \rangle)^2 \rangle \\ &= \langle O^2 - 2O\langle O \rangle + \langle O \rangle^2 \rangle \\ &= \langle O^2 \rangle - 2\langle O \rangle \langle O \rangle + \langle O \rangle^2 \\ &= \langle O^2 \rangle - 2\langle O \rangle^2 + \langle O \rangle^2 \\ &= \langle O^2 \rangle - \langle O \rangle^2. \end{aligned} \tag{4.7}$$

Note multiplying by a constant c results in $\text{Var}[cO] = c^2\text{Var}[O]$. Applying this

¹In the continuum limit (here a discrete probability distribution over an infinite number of bitstrings) care must be taken to ensure the mean and variance are defined for a given quantum state. Consider equation 4.3, in the continuum limit ($d = \infty$). The expectation value will only be defined if the sum converges. If the sum diverges then the expectation value is undefined. Interestingly, for problems studied on a quantum computer there is always a finite number of qubits and so this divergence issue cannot occur.

relation to a general Pauli operator:

$$\begin{aligned}
 \text{Var}[c_i P_i] &= c_i^2 (\langle P_i^2 \rangle - \langle P_i \rangle^2) \\
 &= c_i^2 (\langle I \rangle - \langle P_i \rangle^2) \\
 &= c_i^2 (1 - \langle P_i \rangle^2) \\
 &\leq c_i^2,
 \end{aligned} \tag{4.8}$$

where the fact that Pauli operators are self-inverse has been utilized.

In real experiments we cannot evaluate the expectation value of an operator O exactly², instead, we use statistical sampling to estimate the true value which we denote $\langle \tilde{O} \rangle$ (equation 4.6). $\langle \tilde{O} \rangle$ approximates the true expectation value $\langle O \rangle$ of the operator on the full distribution (a parameter of the wavefunction or true distribution). The Central Limit Theorem means that when the number of samples is large the sample expectation value will be close to the true value. A useful metric to consider is the standard error of a statistic (here the estimated expectation value due to sampling), which is the standard deviation of its sampling distribution. The sampling distribution is a distribution of $\{+1, -1\}$ values generated by sampling the true distribution. This sampled distribution has its own properties (such as the mean and variance). The variance of a sample expectation value (equation 4.6) is:

$$\begin{aligned}
 \text{Var}[\langle \tilde{O} \rangle] &= \text{Var}\left[\frac{1}{M} \sum_{j=0}^{M-1} s_j\right] \\
 &= \frac{1}{M^2} \text{Var}\left[\sum_{j=0}^{M-1} s_j\right] \\
 &= \frac{1}{M^2} \sum_{j=0}^{M-1} \text{Var}[s_j] \\
 &= \frac{1}{M^2} \sum_{j=0}^{M-1} \sigma^2 \\
 &= \frac{1}{M^2} M \sigma^2 = \frac{M \sigma^2}{M^2} \\
 &= \frac{\sigma^2}{M}.
 \end{aligned} \tag{4.9}$$

Note we have used the fact that each s_j sample is independent and identically distributed and thus all the s_j samples will have the same variance σ^2 . Equation 4.9 indicates that as more samples are taken, the variance of the sample mean $\langle \tilde{O} \rangle$ decreases. The square root of equation 4.9 is sometimes known as the standard error on the mean (SEM). Overall, the standard error tells you how accurate any statis-

²Apart from the special case that the state being measured is an eigenstate of that operator

tic obtained from sampling a given population is likely to be compared to the true population parameter. Intuitively this makes sense as if you increase the number of samples M the uncertainty in a statistic determined from sampling should decrease - aka more data reduces your uncertainty.

4.1.2 Expectation variance of linear combination of operators

For many quantum algorithms, the expectation value of an operator is obtained from a linear combination of expectation values of other operators (equation 4.5). In probability theory, the variance of a linear combination of variables is given by Bienaymé's identity:

$$\begin{aligned}
 \text{Var}[H] &= \text{Var}\left[\sum_i c_i P_i\right] \\
 &= \sum_i \sum_j \text{Cov}[c_i P_i, c_j P_j] \\
 &= \sum_i c_i^2 \text{Var}[P_i] + \sum_i \sum_{\substack{j \\ \forall j \neq i}} c_i c_j \text{Cov}[P_i, P_j].
 \end{aligned}
 \tag{4.10}$$

This is simply a sum of all the covariances, where for Pauli operators [171]:

$$\text{Cov}[P_i, P_j] = \langle P_i P_j \rangle - \langle P_i \rangle \langle P_j \rangle.
 \tag{4.11}$$

As measurements are taken from independent state preparations (equation 4.6), the covariance between different estimators must be zero - $\text{Cov}[P_i, P_j] = 0 \forall i \neq j$ [95] - the measurement outcomes are uncorrelated. The overall variance under these conditions is therefore:

$$\text{Var}[H] = \text{Var}\left[\sum_i c_i P_i\right] = \sum_i c_i^2 \text{Var}[P_i].
 \tag{4.12}$$

Interestingly, we will see later that covariances are not zero if groups of Pauli operators are measured at once.

4.1.3 Standard variance in VQE

When performing VQE, one of the goals is to determine the ground state energy of an operator to a certain precision using the minimum number of samples required. The variance of the expectation value is therefore important to account for. The following section considers this problem, where $\langle H \rangle$ is measured as a linear combination of Pauli operators.

As well as the intrinsic variance of the expectation value of an operator on a given state, there will also be sample variance due to statistical fluctuations that are

uncorrelated from sample to sample. The total variance of the sample expectation value $\langle H \rangle$ scales as [172]

$$\text{Var}[H] = \sum_{i=0}^{|H|-1} \frac{c_i^2 \text{Var}[P_i]}{M_i} = \sum_{i=0}^{|H|-1} \frac{c_i^2 (1 - \langle P_i \rangle^2)}{M_i}, \quad (4.13)$$

and derives from the results of equations 4.9 and 4.12. The total number of samples taken is $M = \sum_{i=0}^{|H|-1} M_i$, where M_i is the number of samples used to determine $\langle P_i \rangle$. We want to take the minimum number of samples M_i for each term in order to minimize the overall measurement cost, such that $\text{Var}[H]$ is minimized. This was discussed in [173] and proven in [172] using Lagrange multipliers. We reproduce their derivation here. The following Lagrangian is used:

$$\mathcal{L} = \sum_{i=0}^{|H|-1} M_i + \lambda \left(\sum_{i=0}^{|H|-1} \frac{c_i^2 \text{Var}[P_i]}{M_i} - \epsilon^2 \right), \quad (4.14)$$

where ϵ^2 denotes the target variance for $\text{Var}[H]$. The following expression for M_i must then be solved:

$$\min_{M_i} \max_{\lambda} \mathcal{L} = \min_{M_i} M. \quad (4.15)$$

This is done by taking the derivative of \mathcal{L} with respect to M_i :

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial M_i} &= \frac{\partial}{\partial M_i} \left[\sum_{i=0}^{|H|-1} M_i + \lambda \left(\sum_{i=0}^{|H|-1} \frac{c_i^2 \text{Var}[P_i]}{M_i} - \underbrace{\epsilon^2}_{\text{target Var}[H]} \right) \right] \\ &= 1 - \lambda \left(\frac{c_i^2 \text{Var}[P_i]}{M_i^2} \right) = 0 \\ &\implies 1 = \lambda \frac{c_i^2 \text{Var}[P_i]}{M_i^2}. \end{aligned} \quad (4.16)$$

Rearranging equation 4.16, we see:

$$\begin{aligned} M_i^2 &= \lambda c_i^2 \text{Var}[P_i] \\ M_i &= \sqrt{\lambda} |c_i| \sqrt{\text{Var}[P_i]}. \end{aligned} \quad (4.17)$$

Plugging this back into equation 4.13:

$$\begin{aligned}
 \underbrace{\epsilon^2}_{\text{Var}[H]} &= \sum_{i=0}^{|H|-1} \frac{c_i^2 \text{Var}[P_i]}{\sqrt{\lambda} |c_i| \sqrt{\text{Var}[P_i]}} \\
 &= \frac{1}{\sqrt{\lambda}} \sum_{i=0}^{|H|-1} |c_i| \sqrt{\text{Var}[P_i]} \\
 &\implies \\
 \sqrt{\lambda} &= \frac{1}{\epsilon^2} \sum_{i=0}^{|H|-1} |c_i| \sqrt{\text{Var}[P_i]}.
 \end{aligned} \tag{4.18}$$

Using this result in combination with equation 4.17, we find:

$$\begin{aligned}
 M &= \sum_{i=0}^{|H|-1} M_i = \sum_{i=0}^{|H|-1} (\sqrt{\lambda} |c_i| \sqrt{\text{Var}[P_i]}) = \sqrt{\lambda} \sum_{i=0}^{|H|-1} (|c_i| \sqrt{\text{Var}[P_i]}) \\
 &= \underbrace{\left(\frac{1}{\epsilon^2} \sum_{i=0}^{|H|-1} |c_i| \sqrt{\text{Var}[P_i]} \right)}_{\sqrt{\lambda}} \left(\sum_{i=0}^{|H|-1} (|c_i| \sqrt{\text{Var}[P_i]}) \right) \\
 &= \frac{1}{\epsilon^2} \left(\sum_{i=0}^{|H|-1} |c_i| \sqrt{\text{Var}[P_i]} \right)^2 \\
 &= \frac{1}{\epsilon^2} \left(\sum_{i=0}^{|H|-1} |c_i| \sqrt{1 - \langle P_i \rangle^2} \right)^2 \leq \frac{(\sum_{i=0}^{|H|-1} |c_i|)^2}{\epsilon^2}.
 \end{aligned} \tag{4.19}$$

The final line in equation 4.19 uses the upper bound for the variance of each Pauli operator, $\text{Var}[P_i] \leq 1$, to remove the state-dependence. This gives a state-independent upper bound on the number of measurements needed to achieve an expectation value to precision ϵ .

It is clear from the penultimate line of equation 4.19 that for the special case that $\sqrt{\text{Var}[P_i]} \propto \frac{1}{|c_i|}$ each M_i is optimal. However, this scenario is highly unlikely. Instead the measurement budget M_i for estimating each $\langle P_i \rangle$ should be set proportional to $M_i \propto |c_i| \sqrt{\text{Var}[P_i]}$. However, $\text{Var}[P_i]$ is state-dependent and not normally known before and so what is used instead is: $M_i \propto |c_i|$, where the upper bound of $\text{Var}[P_i] = 1$ (equation 4.8) is used. This strategy works well when $\text{Var}[P_i]$ is roughly the same $\forall i$. A numerical investigation by Arrasmith *et al.* found that variations in $|c_i|$ were generally higher than $\text{Var}[P_i]$ for random states and thus choosing $M_i \propto |c_i|$ outperformed sampling each term uniformly [174]. An alternate approach could use a small number of experimental runs to estimate each $\text{Var}[P_i]$, this could then be used to define each $M_i \propto |c_i| \sqrt{\text{Var}[P_i]}$.

Since the number of terms in the second quantized molecular Hamiltonian (equation 2.73) scales as $\mathcal{O}(K^4)$, the total number of measurements required (equation 4.19) will scale as $\mathcal{O}(K^8/\epsilon^2)$. For chemistry problems, the desired precision is $\epsilon = 1$ kcal/mol (1.6 mHa) and is colloquially known as chemical accuracy. This is the precision required to match typical thermochemical experiments. Wecker *et al.* showed that to obtain energy estimates for HeH^+ , BeH_2 and H_2O requires $10^8 - 10^9$ samples to achieve an error of 1 mHa [173]. This implies that the number of measurements required is an obstacle for experimental implementations of VQE to the number of qubits currently available on NISQ devices.

For example, take the experimental implementation of VQE by Hempel *et al.*, which took 20 ms to perform each VQE repetition on a trapped ion quantum computer [36]. To obtain the ground state energy of H_2 , in a minimal basis to within chemical precision, took ≈ 14000 repetitions and required 4.6 minutes of averaging. Since then many groups have estimated the runtime of VQE. Gonthier *et al.* provide different estimates for a set of molecular systems and found that for ethanol (260 qubit problem) it would take 71 days to obtain a single energy estimate [163]. Johnson *et al.* also recently estimated the runtime of different VQE problems and for CO_2 (208 qubits), and found that roughly three million years would be required to estimate the energy of the prepared ground state using standard sampling [175]. This is clearly a major obstacle for VQE being of any actual utility.

Various approaches have been proposed for reducing the total number of samples required by VQE. We give a brief review of commonly used strategies next.

4.2 Prior work on measurement strategies

We summarise some important work in the field on reducing the measurement cost for different quantum algorithms that require measurement of Pauli operators. Numerous reviews on the topic have been done and we refer the reader to [163, 176] for further details and approaches.

4.2.1 Partial tomography approaches

Quantum state tomography is the experimental procedure to determine an unknown quantum state ρ , which is a density operator often also called the density matrix [86]. The density matrix is defined as

$$\rho \equiv \sum_i p_i |\psi_i\rangle \langle \psi_i|. \quad (4.20)$$

This represents a state with probability p_i the system is in state $|\psi_i\rangle$. In this formalism, expectation values are written as $\langle A \rangle = \langle \phi | A | \phi \rangle = \text{Tr}(\rho A)$, where Tr denotes the trace operation.

In general a single copy of ρ is not sufficient to determine the state of ρ , due to no quantum measurements being able to determine non-orthogonal states. For example, in the one qubit case, no single qubit measurement can distinguish $|0\rangle$ and $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$. However, we can estimate the state of ρ if multiple copies of the unknown state are available. An arbitrary density matrix over n qubits can be written as [86]:

$$\rho = \sum_{i_0=0}^3 \sum_{i_1=0}^3 \dots \sum_{i_{n-1}=0}^3 \frac{\text{Tr}(\sigma_0^{i_0} \otimes \sigma_1^{i_1} \otimes \dots \otimes \sigma_{n-1}^{i_{n-1}} \rho)}{2^n} \sigma_0^{i_0} \otimes \sigma_1^{i_1} \otimes \dots \otimes \sigma_{n-1}^{i_{n-1}}, \quad (4.21)$$

where

$$\sigma_l^j = \begin{cases} \mathcal{I} & \text{for } j = 0 \\ X & \text{for } j = 1 \\ Z & \text{for } j = 2 \\ Y & \text{for } j = 3 \end{cases}, \quad \forall l = 0, 1, \dots, (n-1). \quad (4.22)$$

Here l denotes the qubit index the Pauli operator acts on and j the type of Pauli operator.

The total number of operators requiring measurement scales exponentially as $4^n - 1$. The -1 occurs as we do not need to measure the n -fold identity term - $\text{Tr}(I^{\otimes n} \rho)$ as it always equals $+1$ [177], and occurs when $\{i_l = 0 \forall l = 0, 1, \dots, n-1\}$. This method forms a set of orthonormal matrices, with respect to the Hilbert space and thus ρ can be expanded as a linear combination of these [86].

If we consider a two-qubit example, then the full density matrix can be estimated by measuring $4^2 - 1 = 15$ expectation values. If we determine each using M measurements, then this has a cost of $15M$. However, Cotler and Wilczek realized there are redundancies when measuring certain operators and recently proposed a method that can improve upon this implementation. When i_0 and i_1 are non-zero, they can be used to find the 1-site expectation values. For example $\text{Tr}(\sigma_0^{i_0=1} \otimes \sigma_1^{i_1=2} \rho) \equiv \text{tr}(X_0 \otimes Z_1 \rho)$ can be used to determine $\text{Tr}(\mathcal{I}_0 \otimes Z_1 \rho)$ and $\text{tr}(X_0 \otimes \mathcal{I}_1 \rho)$ [177]. Therefore the terms $(i_0 = 0, i_1 = 1, 2, 3)$ and $(i_1 = 1, 2, 3, i_0 = 0)$ do not need to be measured, as they can be inferred. Overall only $9M$ measurements are required. In general, for full state tomography only 3^N terms require measurement, which is equivalent to measuring all combinations of X, Z, Y on each independent qubit. A natural question that arises from this is whether we can evaluate k -qubit reduced density matrices, rather than

the full density matrix, more efficiently? We define a k -qubit reduced density matrix as:

$$\rho_k = \frac{1}{2^k} \sum_{i_0=0}^3 \sum_{i_1=0}^3 \dots \sum_{i_{k-1}=0}^3 \text{tr} \left(\sigma_0^{i_0} \otimes \sigma_1^{i_1} \otimes \dots \otimes \sigma_{k-1}^{i_{k-1}} \rho_k \right) \sigma_0^{i_0} \otimes \sigma_1^{i_1} \otimes \dots \otimes \sigma_{k-1}^{i_{k-1}}. \quad (4.23)$$

For a general n qubit state, there will be $\binom{n}{k}$ k -body reduced density operators that require measuring. However, due to overlapping terms Cotler and Wilczek showed that all k -body reduced density matrices can be determined in $e^{\mathcal{O}(k)} \log_2(n)$ rounds of measuring. Therefore, to achieve a precision ϵ requires the number of measurements scaling as:

$$M \approx e^{\mathcal{O}(k)} \left(\frac{\log_2(n)}{\epsilon} \right)^2. \quad (4.24)$$

Their method is known as quantum overlapping tomography (QOT). At a high level, the method provides a measurement protocol - using hash functions - that defines which operators to measure in order to determine all k -qubit reduced density matrices of an n -qubit state efficiently.

Subsequent work, by Bonet-Monroig et al. published just after, devised a binary partitioning scheme allowing the fermionic 2-RDM ($k = 2$ and $k = 4$) to be estimated to a constant error using $\mathcal{O}(n^2)$ measurements. Each measurement requires an additional $\mathcal{O}(n)$ linear depth quantum circuit. The underlying principle is the same, except a binary strategy is employed rather than hash functions. In general, their approach requires $\mathcal{O}(3^k \log^{k-1}(n))$ unique measurements. For $k = 2$, the scheme is identical to QOT; however, for higher k the scaling of QOT is better by polylogarithmic factors [177, 178].

4.2.2 Variance reduction method

In [172], Rubin *et al.* use n -representability constraints on the expectation values of few-fermion operators to construct estimators for the expectation value of the Hamiltonian with lower variance. The underlying principle is adding operators to the Hamiltonian that sum to zero and optimizing their coefficients to reduce the total variance. By using algebraic equalities from the fermionic n -representability constraints, they can cast this as an optimization problem, where linear programming techniques can be used to find the new coefficients. The new Hamiltonian constructed from the original will have the same expectation value, but lower maximum variance.

In a piece of follow-up work, Gonthier *et al.* utilized this technique and also implemented it in the qubit picture comparing the two approaches [163]. They numerically found that performance in the qubit picture resulted in lower variances compared to

the fermionic approach.

4.2.3 Basis rotation technique

The “Basis Rotation Grouping ” strategy [179] aims to reduce the number of separate terms requiring measurement by using tensor factorization techniques applied to the second quantized molecular Hamiltonian [179–182]. This allows the Hamiltonian to be written as:

$$H_e = U_0 \left(\sum_{p=0}^{K-1} g_p a_p^\dagger a_p \right) U_0^\dagger + \sum_{l=0}^{L-1} U_l \left(\sum_{p=0}^{K-1} \sum_{q=0}^{K-1} g_{pq}^{(l)} a_p^\dagger a_p a_q^\dagger a_q \right) U_l^\dagger. \quad (4.25)$$

U_l are unitary operators that implement a single particle change of orbital basis, which yields a diagonal operator in the new basis. The key result is $\langle a_p^\dagger a_p \rangle$ and $\langle a_p^\dagger a_p a_q^\dagger a_q \rangle$ can be sampled simultaneously because under the Jordan-Wigner transformation these are diagonal qubit operators. In an experiment, if measurement in the computational basis is performed then measurement outcomes for all the diagonal qubit operators can be achieved in $L + 1 = \mathcal{O}(K)$ distinct measurements [179]. The overall reduction is $\mathcal{O}(K^4)$ to $\mathcal{O}(K)$, at the cost of implementing U_l (after the Ansatz circuit) prior to measurement. The additional circuit cost to perform these unitary operations scales linearly as $\mathcal{O}(K)$ [179].

Gonthier *et al.* in [183] compared this measurement reduction strategy against qubit-wise commutation (introduced in Section 4.2.5.2) and found significant improvements in the number of measurements required.

4.2.4 Shadow methods

Aaronson noted in [184] that a full classical description of a quantum state is unnecessary for certain tasks. Instead, it is often enough to predict properties of the quantum system - i.e find the expectation values of a set of operators. Shadow tomography is a technique to predict properties of a system without fully characterizing the state. He showed that with a polynomial number of states, one could predict an exponential number of target functions. However, the algorithm requires a quantum memory and exponentially long quantum circuits making it unnameable for NISQ hardware.

Huang *et al.* developed the classical shadow technique, based on the ideas of shadow tomography, to allow application on current hardware [185]. The goal is to efficiently learn a classical representation of a quantum state from N snapshots (measurements) of the quantum state. Consider an n -qubit quantum state ρ prepared in a quantum circuit followed by applying a unitary U , the state has been mapped as $\rho \mapsto U\rho U^\dagger$. Measuring this state in the computational basis gives a binary bitstring $|b\rangle \in \{0, 1\}^{\otimes n}$. If each U is selected randomly from a fixed ensemble, then it is possible

to efficiently store the reverse operation $U^\dagger |b\rangle \langle b| U$ in classical memory. Huang and co-workers note that by taking the expectation value over both the choice of U and the outcome distribution it acts as a quantum channel mapping ρ to its classical snapshot $U^\dagger |b\rangle \langle b| U$:

$$\mathbb{E}[U^\dagger |b\rangle \langle b| U] = \mathcal{M}(\rho). \quad (4.26)$$

This quantum channel \mathcal{M} depends on the fixed ensemble of unitaries U . However, given \mathcal{M} it can be classically inverted \mathcal{M}^{-1} to reconstruct a snapshot of the original quantum state:

$$\tilde{\rho} = \mathbb{E}[\mathcal{M}^{-1} U^\dagger |b\rangle \langle b| U]. \quad (4.27)$$

By repeating this N times, there will be a collection of inverted snapshots known as the classical shadow:

$$S(\rho, N) = \left\{ \begin{aligned} \tilde{\rho}_0 &= \mathcal{M}^{-1}(U_0^\dagger |b_0\rangle \langle b_0| U_0), \\ \tilde{\rho}_1 &= \mathcal{M}^{-1}(U_1^\dagger |b_1\rangle \langle b_1| U_1), \\ &\dots, \\ \tilde{\rho}_{N-1} &= \mathcal{M}^{-1}(U_{N-1}^\dagger |b_{N-1}\rangle \langle b_{N-1}| U_{N-1}) \end{aligned} \right\}. \quad (4.28)$$

The inverted channel is not completely positive and trace-preserving and thus not physical, but as it is calculated in a classical post-processing step this is not a problem. Given such a set of classical shadows, the expectation value of an operator \hat{O}_l is given by the empirical averages of independent shadow predictions:

$$\text{Tr}(\hat{O}_l \rho) = \frac{1}{N} \sum_{i=0}^{N-1} \text{Tr}(\hat{O}_l \cdot \tilde{\rho}_i). \quad (4.29)$$

Due to the fact that classical shadows are random, this produces a random variable that yields the correct prediction in expectation. This estimator can be affected by outliers and can be improved by taking the median of means. This simply requires separating the shadow $S(\rho, N)$ into K equally sized chunks: $S(\rho, N/K)$. The mean of each $S(\rho, K)$ is then taken (equation 4.29) and the median value over the set is used as the estimator:

$$\text{Tr}(\hat{O}_l \rho) = \text{median} \left\{ \langle \hat{O}_l \rangle^{k=0}, \langle \hat{O}_l \rangle^{k=1}, \dots, \langle \hat{O}_l \rangle^{k=N/K} \right\} \quad (4.30)$$

and is more robust to outliers. Huang *et al.* show that the number of samples required

to achieve precision ϵ scales as:

$$M \geq \mathcal{O}\left(\frac{\log N \max_i |O_i|_{shadow}^2}{\epsilon^2}\right), \quad (4.31)$$

where $|O_i|_{shadow}^2$ is the shadow norm that depends on the unitary ensemble chosen. Two important ensembles to consider are:

1. random n -qubit Clifford circuits
2. tensor products of random single qubit Clifford circuits

as the stabilizer formalism can be used to efficiently store the classical shadow $S(\rho, N)$ for both of these ensembles classically. However, the first ensemble requires $n^2/\log(n)$ entangling gates to sample the n -qubit Clifford unitaries making it less NISQ friendly [186]. We focus our subsequent analysis on the second case, which is equivalent to measuring each qubit in random Pauli bases. This measurement approach is standard in many NISQ algorithms performed on hardware to date. A drawback to using this ensemble, is the shadow norm becomes dependent on the locality k of the observables that are to be estimated:

$$|O_i|_{shadow}^2 \leq 4^k |O_i|_{\infty}^2, \quad (4.32)$$

where the spectral norm is defined as $|O_i|_{\infty}^2 = \max_{0 \leq i \leq n-1} \sum_{j=0}^{n-1} |a_{ij}|$, given the absolute magnitude sum of each column, $|O_i|_{\infty}^2$ is the largest value of these. We illustrate the implication of this with an example. Consider you want to measure the expectation value of $\langle P \rangle = \langle \sigma_0 \otimes \sigma_1 \otimes \dots \otimes \sigma_{n-1} \rangle$, where σ are single qubit Pauli matrices (not including the identity) on an n qubit state. Then one would need an exponentially large classical shadow (4^n) and thus $M \approx \mathcal{O}\left(\frac{4^n}{\epsilon^2}\right)$ samples to determine $\langle P \rangle$ to precision ϵ . In the current form, classical shadows based on Pauli measurements only offers an advantage for a large number of observables with modest locality.

Since the classical shadow approach was proposed there have been numerous works to improve the method. A particular focus has been made in the Pauli basis measurement picture; where if one has knowledge of the operator to be measured can one do better than randomly measuring? Hadfield *et al.* provide a classically efficient algorithm for the scenario where knowledge of the operator to measure and a classical approximation to the underlying quantum state is known *a priori* [187]. Their proposal is known as Locally-Biased Classical Shadows. Soon after, Huang and co-workers proposed an alternate strategy using derandomization, which is also classically tractable but formulated as a greedy process [188]. Numerically derandomization was observed to outperform Locally-Biased Classical Shadows; however, the optimality of derandomization is unknown due to it being a greedy heuristic.

4.2.5 Partitioning approaches

For a qubit Hamiltonian written as some linear combination of Pauli operators, it is always possible to decompose it into groups (or cliques) of simultaneously measurable operators. We write this as:

$$H = \sum_i c_i P_i = \sum_j^{N_C} C_j = \sum_j^{N_C} \underbrace{\left(\sum_k^{|C_j|} c_k^{(j)} P_k^{(j)} \right)}_{C_j}, \quad (4.33)$$

where C_j are cliques given as linear combinations of Pauli operators. There are different ways to form each clique according to properties within a clique. We review the following cases:

1. cliques composed of pairwise commuting operators
2. cliques composed of pairwise qubit-wise commuting operators
3. cliques composed of pairwise anticommuting operators

in the subsequent subsections and then discuss how the variance of H is affected by grouping strategies.

4.2.5.1 Commuting cliques

In quantum mechanics, it is known that a set of mutually commuting operators can be measured simultaneously as they share a common basis. Such a set forms an Abelian group. From stabilizer theory [189], it is known that an Abelian group S has a set of independent generators $\{t_0, t_1, \dots, t_w\}$ that can be efficiently determined [124, 189]. Each independent generator can be mapped to a single Pauli matrix on a distinct qubit j :

$$U t_i U^\dagger = \sigma_j \in \{X, Y, Z\}. \quad (4.34)$$

For commuting measurements, it is standard to map each generator to a Pauli operator comprised of only a single non-trivial single qubit Pauli Z matrix via a Clifford operator Q . The same mapping is performed on the clique of commuting Pauli operators: $C_{comm} \mapsto C'_{comm} = Q^\dagger C_{comm} Q$. The expectation value of C'_{comm} can then be inferred by some combination of the expectation values in the generating set. These are a set of single Pauli Z operators.

For example, consider the commuting clique:

$$C_{comm} = ZXZ + YYI + XZI + YYZ, \quad (4.35)$$

4.2. PRIOR WORK ON MEASUREMENT STRATEGIES

where each Pauli operator pairwise commutes. C_{comm} has the following generators:

$$G \equiv \{ZXI, XZI, IIZ\}. \quad (4.36)$$

These can be mapped to the following single qubit Z operators:

$$G' \equiv \{IZI, ZII, IIZ\}, \quad (4.37)$$

via the following Clifford operators $U_1 = e^{i\frac{\pi}{4}ZYI} = \frac{1}{\sqrt{2}}III + \frac{1}{\sqrt{2}}i ZYI$ and $U_2 = e^{i\frac{\pi}{4}YZI} = \frac{1}{\sqrt{2}}III + \frac{1}{\sqrt{2}}i YZI$:

$$\begin{aligned} U_2U_1(ZXI)U_1^\dagger U_2^\dagger &= IZI \\ U_2U_1(XZI)U_1^\dagger U_2^\dagger &= ZII \\ U_2U_1(IIZ)U_1^\dagger U_2^\dagger &= IIZ. \end{aligned} \quad (4.38)$$

In a quantum circuit, U_1 is performed followed by U_2 and then each qubit is measured in the computational basis (giving: ZII , IZI and IIZ). The transformed C_{comm} will be:

$$\begin{aligned} C_{comm} \mapsto C'_{comm} &= U_2U_1C_{comm}U_1^\dagger U_2^\dagger \\ &= IZZ + ZZI + ZII + ZZZ \end{aligned} \quad (4.39)$$

The expectation value of C'_{comm} can then be determined from the expectation values of the operators in the generating set G' . Explicitly:

$$\begin{aligned} \langle IZZ \rangle &= \langle IZI \rangle \times \langle IIZ \rangle \\ \langle ZZI \rangle &= \langle IZI \rangle \times \langle ZII \rangle \\ \langle ZII \rangle &= \langle ZII \rangle \\ \langle ZZZ \rangle &= \langle IZI \rangle \times \langle ZII \rangle \times \langle IIZ \rangle. \end{aligned} \quad (4.40)$$

In this example, only a single ZZZ measurement was required rather than four, at the increased circuit cost of implementing each Clifford rotation prior to measurement.

Gokhale *et al.* in 2019 looked at this strategy in the context of VQE applied to quantum chemistry [190]. They showed the Clifford operation could be constructed using $\mathcal{O}(n^2)$ CNOT gates. This was further improved by Crawford *et al.* who showed that for k commuting n -qubit Pauli operators at most would require $kn - k(k+1)/2$ and $\mathcal{O}(kn/\log k)$ two-qubit gates.

In general, for n qubits, the maximum number of Pauli operators is $4^n - 1$ excluding $I^{\otimes n}$. The optimal commuting cover will have $2^n + 1$ commuting cliques [191, 192]. This indicates that simultaneous measurement can lead to an exponential reduction in quantum cost compared to naively measuring each term separately. However, for quantum chemistry the molecular Hamiltonian (equation 2.73) contains $\mathcal{O}(K^4)$ terms

rather than $4^n - 1$ Pauli operators. The improvement in the number measurements is therefore more complicated.

The number of Pauli operators in a second quantized molecular Hamiltonian $\mathcal{O}(K^4)$ was observed to be decreased to $\mathcal{O}(K^3)$ terms by grouping commuting Pauli operators [190, 193, 194]. This improvement was based on extrapolation across a set of molecular examples. In [195], Gokhale and Chong provide an algorithm that proves this linear-reduction in the number of spin orbitals K is always possible for Hamiltonians encoded under the Jordan-Wigner transformation. They demonstrate that molecular Hamiltonians can always be partitioned into pairwise-commuting families where each family contains $\mathcal{O}(K)$ terms. Next, we summarise the qubit-wise commuting approach, which is in a sense just a more strict commuting clique condition.

4.2.5.2 Qubit-wise commuting cliques

In [196], Verteletskyi *et al.* formalize how to find different terms in the Hamiltonian that share a tensor product basis (TPB) They introduce the idea of qubit-wise commutation (QWC), where two operators qubit-wise commute if the individual Pauli matrices on each qubit in each operator pairwise commute. The QWC commutator is defined as [196]:

$$[P_i, P_j]_{qwc} = \left[\bigotimes_{k=0}^{n-1} \sigma_k^{(i)}, \bigotimes_{k=0}^{n-1} \sigma_k^{(j)} \right]_{qwc} = \begin{cases} 0, & \text{if } [\sigma_k^{(i)}, \sigma_k^{(j)}] = 0 \forall k \text{ where } 0 \leq k \leq n-1 \\ 1, & \text{otherwise.} \end{cases} \quad (4.41)$$

For example $[X_0Z_1, \mathcal{I}_0Z_1]_{qwc} = 0$ and $[X_0Z_1, Y_0Y_1]_{qwc} = 1$. Note the important difference between this and the standard commutator. Any QWC pair of operators must also commute in the standard sense, but if a pair of Pauli operators commute they do not necessarily QWC, e.g. $[X_0Z_1, Y_0Y_1] = 0$.

For a Hamiltonian partitioned into qubit-wise commuting cliques, a single set of $\mathcal{O}(n)$ one qubit measurements can be used to determine the expectation value of all the operators in that clique. The true number is simply any qubit index that is acted on non-trivially. The advantage of this approach over general commuting cliques, is no extra circuit operations are required prior to measurement. For chemistry problems under the Bravyi-Kitaev and Jordan-Wigner transformations, numerical simulations found that the number of qubit-wise commuting groups scaled as $\mathcal{O}(K^4)$ [196]. This is the same scaling as the original Hamiltonian (equation 2.73) and means there is unlikely to be any major measurement reduction obtained from this approach. While not directly related to the $\mathcal{O}(K^4)$ Pauli operators in a chemistry Hamiltonian, Yen *et al.* in [194] show that there are exponentially more commuting Pauli operators in the full set of 4^n operators than qubit-wise commuting Pauli operators. While we offer no

proof, we conjecture this relates to why the qubit-wise commuting strategy does not result in a linear measurement reduction but the general commuting scenario does.

Again, we address how the cliques are determined in Section 4.2.5.4. Next, we review the unitary partitioning strategy which partitions a Hamiltonian into normalized anticommuting cliques.

4.2.5.3 Unitary partitioning (anticommuting cliques)

For a Hamiltonian composed as a linear combination of Pauli operators, the expectation value is usually determined by the linear combination of the expectation value of each Pauli operator P_i , which are each Hermitian ($P_i = P_i^\dagger$), unitary ($P_i P_i^\dagger = I$) operators. The goal of unitary partitioning, a technique independently proposed by Verteletskyi *et al.* [197] and Zhao *et al.* [198], is to group Pauli operators into different linear combinations where each group is unitary and represent measurable sets. Note any real linear combination of Pauli operators will be Hermitian and thus a valid observable. However, general linear combinations of such operators will not be unitary. Both unitary and Hermitian conditions allow the whole group to be determined at once. We see this if we take the spectral decomposition of a general Hermitian operator A :

$$A = \sum_{a=0}^{d-1} \lambda_a |\psi_a\rangle \langle \psi_a|, \quad (4.42)$$

where d is the dimension of the space and A acts on orthonormal states $|\psi_a\rangle$. Each $|\psi_a\rangle$ is an eigenstate of the operator with corresponding real eigenvalue λ_a . Note for a unitary operator each $|\lambda_a| = 1$, but can be complex. As the set of eigenvectors $\{|\psi_a\rangle\}$ form an orthonormal basis there always exists a unitary R that maps this basis to another: $R|\psi_a\rangle = |e_a\rangle$ or $|\psi_a\rangle = R^\dagger|e_a\rangle$. The operator A can be written in this basis:

$$\begin{aligned} A &= \sum_{a=0}^{d-1} \lambda_a |\psi_a\rangle \langle \psi_a| \\ &= \sum_{a=1}^d \lambda_a R^\dagger |e_a\rangle \langle e_a| R \\ &= R^\dagger \left(\sum_{a=1}^d \lambda_a |e_a\rangle \langle e_a| \right) R \\ &= R^\dagger Q R. \end{aligned} \quad (4.43)$$

The expectation value of A can be found by $\langle A \rangle = \langle \psi | A | \psi \rangle = \langle \psi | R^\dagger Q R | \psi \rangle$. As the unitary constraint means the expectation value must be $+1$ or -1 , it can be evaluated at once on a quantum computer. This idea underpins the unitary partitioning

strategy for measurement reduction.

In order to make a general sum of Pauli operators, $L = \sum_a c_a P_a$, unitary three conditions must be met [197, 198]:

1. $\{P_i, P_j\} = 0 \forall P_i, P_j \in L$ and $i \neq j$
2. $\sum_j |c_j|^2 = 1$
3. $\text{Im}(c_j^* c_i) = 0 \forall c_i, c_j \in L$

The first condition requires each Pauli operator in the linear combination to pairwise anticommute with all the other operators, where $\{, \}$ is the anticommutator defined as $\{A, B\} \equiv AB + BA$. The second condition ensures the eigenvalues of the resulting linear combination will have a norm of one. The final condition ensures sum will be Hermitian, as no imaginary coefficient can be generated, and is a constraint that ensures the group is a valid observable. If we take a general linear combination of anticommuting Pauli operators $\chi = \sum_{i=0}^{k-1} c_i P_i$, meeting these conditions then:

$$\begin{aligned}
 \chi^2 &= \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} c_i c_j P_i P_j \\
 &= \sum_{\substack{i=0 \\ \forall j=i}}^{k-1} c_i c_j P_i P_j + \sum_{\substack{i=0 \\ \forall i \neq j}}^{k-1} c_i c_j P_i P_j \\
 &= \sum_{i=0}^{k-1} c_i^2 P_i^2 + \sum_{i>j}^{k-1} \left(c_i c_j P_i P_j + c_j c_i P_j P_i \right) \\
 &= \sum_{i=0}^{k-1} c_i^2 I + \sum_{i>j}^{k-1} c_i c_j \underbrace{\{P_i, P_j\}}_{=0} \\
 &= \sum_{i=0}^{k-1} c_i^2 I = I \underbrace{\left(\sum_{i=0}^{k-1} c_i^2 \right)}_{=1} = I.
 \end{aligned} \tag{4.44}$$

Thus proving a normalized linear combination of anticommuting Pauli operators is a unitary operator. As each coefficient is also real, this operator is also Hermitian.

To implement unitary partitioning, the Hamiltonian is decomposed into normal-

ized anticommuting cliques. This is concisely written:

$$\begin{aligned}
 H &= \sum_i^{\mathcal{O}(N^4)} c_i P_i \\
 &= \sum_j \gamma_j C_j^U \\
 &= \sum_j \gamma_j \left(\sum_{\substack{k \\ \{P_a, P_b\}=0 \\ \forall P_a, P_b \in C_j^U \\ a \neq b}} \frac{c_k}{\gamma_j} P_k \right),
 \end{aligned} \tag{4.45}$$

where $\gamma_j = (\sum_k^{C_j^U} c_k^2)^{0.5}$ and is the normalization factor for each anticommuting clique C_j^U . The superscript U is used to denote the clique is a unitary operator. Using equation 4.43, we can rewrite each C_j^U term yielding:

$$\begin{aligned}
 H &= \sum_j \gamma_j C_j^U \\
 &= \sum_j \gamma_j (R_j^\dagger A_j R_j).
 \end{aligned} \tag{4.46}$$

In an experiment to evaluate $\langle \psi | R_j^\dagger A_j R_j | \psi \rangle$, one performs R followed by measuring A_j . The details on how R_j is constructed and what A_j is will be discussed in detail in Section 4.3.

Numerical evidence of unitary partitioning applied to different quantum chemistry Hamiltonians showed a K -fold reduction in the number of measurable groups as $\mathcal{O}(K^4) \mapsto \mathcal{O}(K^3)$ [197]. An algorithm was given by Zhao *et al.* in [199] that proved a linear reduction is always possible. Their approach is based on writing the second quantized Hamiltonian as a linear combination of majorana operators. They show a linear reduction is always possible by counting the number of anticommuting terms by orbital index. They note their partitioning scheme is not an optimal solution, but can always be applied to obtain a linear term reduction in K . A numerical study is then presented, for different molecular systems, and they show graph based heuristics can improve the partitioning of the molecular Hamiltonian into anticommuting sets. Such graph strategies are discussed next.

4.2.5.4 The problem of finding cliques

So far in this Section, we have reviewed different partitioning approaches that split a Hamiltonian (equation 4.33) into different groups according to some property. Either all terms within a group will pairwise commute, pairwise qubit-wise commute, or all pairwise anticommute. This structure allows the problem to be mapped to a graph

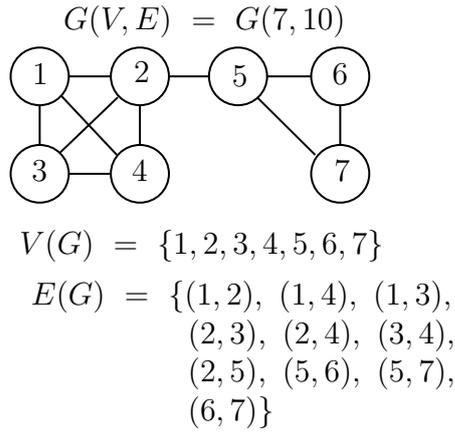


Figure 4.1: Example graph using standard notation.

problem. Before we discuss how this can be done, we quickly review what a graph is along with some basic definitions.

A graph G is a collection of vertices V (sometimes called nodes) and edges E . The following notation is commonly used: $G = (V, E)$, which means a graph that has V vertices and E edges. It is common notation to use $V(G)$ to represent the set of nodes and $E(G)$ to represent the set of edges. Figure 4.1 illustrates a simple example.

A clique is defined as follows: A set of vertices C is a clique of the graph G if and only if $C \subseteq V(G)$ such that $a, b \in E(G) \forall a, b \in C \wedge a \neq b$. Unpacking this definition, a clique is a subset of the vertex set $V(G)$ such that all the distinct vertices in this set are joined by an edge. Another way to think of this is a clique is a complete subgraph of G . A graph where all nodes are connected to each other is known as a complete graph. Figure 4.2 gives two examples.

The degree of a vertex a , is denoted $deg(a)$, which represents the number of vertices adjacent to a , or equivalently, the number of edges incident to a . The maximum degree of all vertices in a graph is denoted $\Delta(G)$ and the minimum by $\delta(G)$.

Next, we define what maximum and maximal cliques are, which can often be

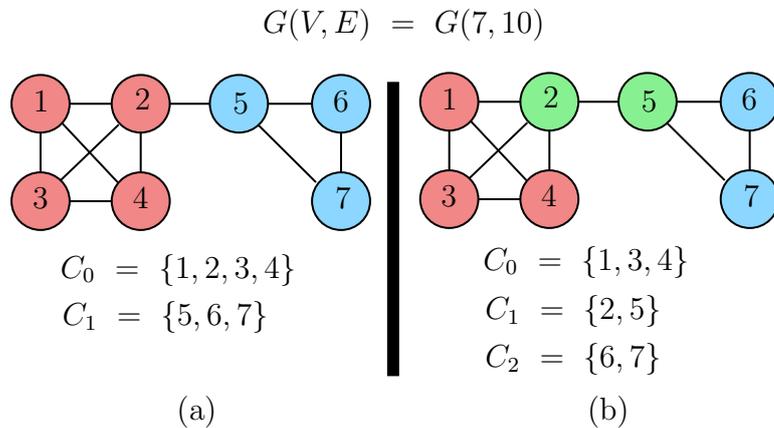


Figure 4.2: Example clique covers of the graph in Figure 4.1.

4.2. PRIOR WORK ON MEASUREMENT STRATEGIES

Clique relation	Strategy
Commuting	<ol style="list-style-type: none"> 1. Construct a graph with edges between anticommuteing Pauli operators 2. Perform a graph colouring
Qubit-wise commuting	<ol style="list-style-type: none"> 1. Construct a graph with edges between qubit-wise commuting Pauli operators 2. Find the complement graph from the previous step 3. Perform a graph colouring of the graph obtained in step 2
Anticommuteing	<ol style="list-style-type: none"> 1. Construct a graph with edges between commuting Pauli operators 2. Perform a graph colouring

Table 4.1: Graph based colouring strategies to partition a Hamiltonian into cliques. For the qubit-wise commuting clique relation, we note the complement of a graph is a different graph with the same vertices that have edges between vertices if and only if they are not adjacent in the original graph.

confused due to their similar names. A clique is maximal if and only if it is not a proper subgraph of another clique. In other words, a clique is maximal if it cannot be made larger by adding another vertex/node (which will need to have an edge with all the nodes within the clique so it still remains a clique). A maximum clique of a graph is a clique that has the most possible vertices. We note there can be many maximum cliques within a graph decomposed into cliques. From these definitions, it should be clear that a maximum clique will be maximal, but not every maximal clique will be a maximum clique.

The minimum degree vertex of a clique bounds the size of the clique as $|C| \leq \delta(G_C) + 1$, where G_C is the graph of the clique. This comes from the definition of a clique, which requires a complete graph.

The goal of measurement reduction is to reduce the number of measurements required to obtain an expectation value within precision ϵ . When partitioning the Hamiltonian into groups, it seems logical to group them into the fewest simultaneously measurable parts as possible. This is an instance of the well-known Minimum Clique Cover problem. This is an NP-hard problem [200]; however, heuristic algorithms can provide sufficiently good approximate solutions to this problem.

To solve the minimum clique cover problem, first our Hamiltonian must be mapped onto a graph. To do this each Pauli operator is represented by a vertex (node) and each edge by a certain commutation property, depending on the type of partitioning desired. Table 4.1 summarises each scenario.

Finding the minimum clique cover can then be done via different graph algorithms. Verteletskyi *et al.* explored different heuristics to solve this problem [196]. Their study found that the largest first graph colouring algorithm found the best qubit-wise commuting cliques for a set of different molecular Hamiltonians. Note a graph colouring searches for the minimum number of colours required to colour the graph, where no neighbours of a node can have the same colour as the node itself.

4.2.6 Variance in grouped measurement strategies

As discussed in the previous section, it is possible to partition a Hamiltonian into N_C cliques:

$$H = \sum_i c_i P_i = \sum_j^{N_C} C_j = \sum_j^{N_C} \underbrace{\left(\sum_k^{|C_j|} c_k^{(j)} P_k^{(j)} \right)}_{C_j}, \quad (4.47)$$

based on certain conditions each clique C_j or linear combination of Pauli operators must obey. We presented commuting, qubit-wise commuting or anti-commuting cliques. However, partitioning into the minimum number of measurable sets does not necessarily minimize the total number of measurements. What is actually needed is a partitioning that requires the fewest number of measurements to obtain $\langle H \rangle$ to a certain precision. We follow the analysis of Crawford *et al.* that gives a way to evaluate the measurement reduction given by a certain partitioning [201]. Using the same process as Section 4.1.3, we find the measurement requirement for measuring a grouped set of terms being:

$$M_g = \frac{1}{\text{Var}[H]} \left(\sum_j^{N_C} \sqrt{\text{Var}[C_j]} \right)^2, \quad (4.48)$$

where the derivation follows in the same manner as before - up to the penultimate line of equation 4.19. The previous result of measuring single Pauli operators (equation 4.19) turns out to be a particular instance of this, where each clique has a size of one and can be written:

$$M_u = \frac{1}{\text{Var}[H]} \left(\sum_j^{N_C} \left[\sum_{k=0}^{|C_j|-1} |c_k^{(j)}| \sqrt{\text{Var}[P_k^{(j)}]} \right] \right)^2. \quad (4.49)$$

The subscripts u and g are used to denote no grouping and grouping respectively. A natural metric to evaluate the measurement cost of a particular grouping of Pauli operators is therefore given by the ratio R of these two terms:

$$R = \frac{M_u}{M_g} = \left(\frac{\sum_j^{N_C} \left[\sum_{k=0}^{|C_j|-1} |c_k^{(j)}| \sqrt{\text{Var}[P_k^{(j)}]} \right]}{\sum_j^{N_C} \sqrt{\text{Var}[C_j]}} \right)^2, \quad (4.50)$$

where the greater the value of R , the better the measurement saving is by assembling these operators into a particular group. The problem with this metric is it assumes the variances of all the operators $P_k^{(j)}$ and C_j must be known, but these are state-dependent properties. This issue was addressed in [201], where all the variance and covariance terms are replaced with their expectation value over uniform spherical

distribution. This gives a new ratio metric \tilde{R} :

$$\tilde{R} = \left(\frac{\left(\sum_j^{N_A} \left[\sum_k^{|A_j|} |c_i^j| \right] \right)}{\left(\sum_j^{N_A} \left[\sqrt{\sum_k^{|A_j|} |c_i^j|^2} \right] \right)} \right)^2. \quad (4.51)$$

Interestingly, this new measure is now state-independent. This is useful as it can approximate R , but can be calculated analytically without prior knowledge of the quantum state being measured. Their proof was restricted to commuting Paul operators and required utilizing Theorem 3 of [190], which proved $\mathbb{E}(\text{Cov}[P_i, P_j]) = 0$ for pairwise commuting Pauli operators if the expectation is taken over a uniform distribution over all possible state vectors (the Haar distribution). However, in general the covariance of commuting operators will not be 0 and at worst will be ± 1 [163]. We therefore find:

$$\begin{aligned} \text{Var}[C_j] &= \text{Var}\left[\sum_i c_i P_i\right] = \sum_i \sum_j \text{Cov}[c_i P_i, c_j P_j] \\ &= \sum_i c_i^2 \text{Var}[P_i] + \sum_i \sum_{\substack{j \\ \forall j \neq i}} c_i c_j \text{Cov}[P_i, P_j]. \end{aligned} \quad (4.52)$$

The maximum variance for a commuting clique is therefore:

$$\begin{aligned} \text{Var}_{max}[\langle C_j \rangle] &= \sum_i c_i^2 \text{Var}[P_i] + \sum_i \sum_{\substack{j \\ \forall j \neq i}} |c_i c_j \sqrt{\text{Var}[P_i] \text{Var}[P_j]}| \\ &\leq \sum_i c_i^2 + \sum_i \sum_{\substack{j \\ \forall j \neq i}} |c_i c_j| \end{aligned} \quad (4.53)$$

and the lowest variance possible will be:

$$\text{Var}_{min}[\langle C_j \rangle] = \sum_i c_i^2 \text{Var}[P_i] - \sum_i \sum_{\substack{j \\ \forall j \neq i}} |c_i c_j \sqrt{\text{Var}[P_i] \text{Var}[P_j]}|, \quad (4.54)$$

where $|\sqrt{\text{Var}[P_i] \text{Var}[P_j]}| \geq \text{Cov}[P_i, P_j]$ has been utilized to bound the covariance [163]. Therefore, for a clique of commuting operators we can introduce an upper and lower bound on R :

$$R = \frac{M_u}{M_g} \geq R_{min} = \left(\frac{\sum_j^{N_C} \left[\sum_{k=0}^{|C_j|-1} |c_k^{(j)}| \sqrt{\text{Var}[P_k^{(j)}]} \right]}{\sum_j^{N_C} \sqrt{\text{Var}_{max}[C_j]}} \right)^2, \quad (4.55a)$$

$$R = \frac{M_u}{M_g} \leq R_{max} = \left(\frac{\sum_j^{N_C} \left[\sum_{k=0}^{|C_j|-1} |c_k^{(j)}| \sqrt{\text{Var}[P_k^{(j)}]} \right]}{\sum_j^{N_C} \sqrt{\text{Var}_{min}[C_j]}} \right)^2. \quad (4.55b)$$

The benefit of both R_{min} and R_{max} is they show the best and worst scenarios $R_{min} \leq R \leq R_{max}$, whereas the approach in [201] turns out to only model the case when $\text{Cov}[P_i, P_j] = 0 \forall i \neq j$. A potentially better way to evaluate equation 4.52 is to use a quantum state obtained from a classical chemistry method. This can be used to estimate the variance and covariance terms. and could help find better cliques. We leave this potential improvement as an avenue for future work.

Next, we provide a new analysis for the anticommuting case. Here we examine the covariance of two anticommuting Pauli operators: $\{P_i, P_j\} = 0$. The amount two random variables vary together (co-vary) is measured by their covariance. Consider the results of random variables x and y , one can obtain a set of M paired measurements:

$$\begin{aligned} & \{(x_0, y_0), \\ & \quad (x_1, y_1), \\ & \quad \dots, \\ & \quad (x_{M-1}, y_{M-1})\}. \end{aligned} \quad (4.56)$$

A positive covariance indicates that higher than average values of one variable tend to be paired with higher than average values of the other variable. A negative covariance indicates that a higher than average value of one variable tends to be paired with lower than average values of the other. If two random variables are independent, then their covariance will be zero. However, a covariance of zero does not mean two random variables are independent, as nonlinear relationships can result in a covariance of zero.

In the context of measuring a quantum state in the Pauli basis on a quantum computer, this would be a set of paired single shot samples $\{s_i^a, s_i^b | i = 0, 1, \dots, N-1\}$, where $s_i^a, s_i^b \in \{-1, +1\}$. Experimentally, each pair is the (single shot) measurement outcome for P_a followed by the (single shot) measurement outcome for P_b . Taking simultaneous projective measurements, without re-preparing the quantum state is a meaningful operation for commuting operators as they share a common eigenbasis. This is due to the order of measurement not affecting the measurement outcomes, but the paired samples will be statistically correlated and have a certain covariance. However, for anticommuting operators this is not the case, as these operators do not share a common eigenbasis. Projective measurement means the expectation value of these operators cannot be known simultaneously. We consider the covariance in this scenario.

Consider the spectral decomposition of two anticommuting Pauli operators $\{P_a, P_b\}$:

$$P_a = +1 |\kappa_0\rangle \langle \kappa_0| - 1 |\kappa_1\rangle \langle \kappa_1|, \quad (4.57)$$

and

$$P_b = +1 |\Omega_0\rangle \langle \Omega_0| - 1 |\Omega_1\rangle \langle \Omega_1|, \quad (4.58)$$

where for Pauli operators:

$$\langle \kappa_0 | \kappa_0 \rangle = \langle \kappa_1 | \kappa_1 \rangle = \langle \Omega_0 | \Omega_0 \rangle = \langle \Omega_1 | \Omega_1 \rangle = 1, \quad (4.59a)$$

$$\langle \kappa_0 | \kappa_1 \rangle = \langle \Omega_0 | \Omega_1 \rangle = 0, \quad (4.59b)$$

$$|\langle \kappa_0 | \Omega_0 \rangle|^2 = |\langle \kappa_0 | \Omega_1 \rangle|^2 = |\langle \kappa_1 | \Omega_0 \rangle|^2 = |\langle \kappa_1 | \Omega_1 \rangle|^2 = 0.5. \quad (4.59c)$$

Without loss of generality, assume P_a is measured first on a general normalized quantum state $|\psi\rangle = \gamma |\kappa_0\rangle + \delta |\kappa_1\rangle$. The only possible post-measurement outcomes are $|\kappa_0\rangle$ or $|\kappa_1\rangle$, with probabilities $|\gamma|^2$ or $|\delta|^2$ respectively. Consider the result of subsequently measuring P_b . The expectation value in each scenario will be:

$$\begin{aligned} \langle \kappa_0 | P_b | \kappa_0 \rangle &= \langle \kappa_0 | (|\Omega_0\rangle \langle \Omega_0| - |\Omega_1\rangle \langle \Omega_1|) | \kappa_0 \rangle \\ &= \langle \kappa_0 | \Omega_0 \rangle \langle \Omega_0 | \kappa_0 \rangle - \langle \kappa_0 | \Omega_1 \rangle \langle \Omega_1 | \kappa_0 \rangle \\ &= |\langle \kappa_0 | \Omega_0 \rangle|^2 - |\langle \kappa_0 | \Omega_1 \rangle|^2 \\ &= \underbrace{0.5}_{\mathbb{P}(\Omega_0|\kappa_0)} - \underbrace{0.5}_{\mathbb{P}(\Omega_1|\kappa_0)} = 0, \end{aligned} \quad (4.60a)$$

$$\begin{aligned} \langle \kappa_1 | P_b | \kappa_1 \rangle &= \langle \kappa_1 | (|\Omega_0\rangle \langle \Omega_0| - |\Omega_1\rangle \langle \Omega_1|) | \kappa_1 \rangle \\ &= \langle \kappa_1 | \Omega_0 \rangle \langle \Omega_0 | \kappa_1 \rangle - \langle \kappa_1 | \Omega_1 \rangle \langle \Omega_1 | \kappa_1 \rangle \\ &= |\langle \kappa_1 | \Omega_0 \rangle|^2 - |\langle \kappa_1 | \Omega_1 \rangle|^2 \\ &= \underbrace{0.5}_{\mathbb{P}(\Omega_0|\kappa_1)} - \underbrace{0.5}_{\mathbb{P}(\Omega_1|\kappa_1)} = 0. \end{aligned} \quad (4.60b)$$

Overall, we find the probabilities of all possible combinations of measurement outcomes to be:

$$\begin{aligned} \mathbb{P}(P_b = |\Omega_0\rangle | P_a = |\kappa_0\rangle) &= \mathbb{P}(P_b = |\Omega_1\rangle | P_a = |\kappa_0\rangle) = 0.5, \\ \mathbb{P}(P_b = |\Omega_0\rangle | P_a = |\kappa_1\rangle) &= \mathbb{P}(P_b = |\Omega_1\rangle | P_a = |\kappa_1\rangle) = 0.5. \end{aligned} \quad (4.61)$$

This result shows that the probability of obtaining $|\Omega_0\rangle$ or $|\Omega_1\rangle$ is not affected by the probability of obtaining $|\kappa_0\rangle$ or $|\kappa_1\rangle$ in the first measurement. The variables are

therefore statistically independent³. We find the covariance of P_a and P_b , where $\{P_a, P_b\} = 0$, to be:

$$\begin{aligned}
 Cov[P_a, P_b] &= \mathbb{E} \left[\left(p_a - \langle P_a \rangle \right) \left(p_b - \langle P_b \rangle \right) \right] \\
 &= \mathbb{E} \left[\left(p_a p_b - p_a \langle P_b \rangle - \langle P_a \rangle p_b + \langle P_a \rangle \langle P_b \rangle \right) \right] \\
 &= \left(\mathbb{E}[p_a p_b] - \mathbb{E}[p_a \langle P_b \rangle] - \mathbb{E}[\langle P_a \rangle p_b] + \mathbb{E}[\langle P_a \rangle \langle P_b \rangle] \right) \quad (4.62) \\
 &= \langle P_a P_b \rangle - \langle P_a \rangle \langle P_b \rangle - \langle P_a \rangle \langle P_b \rangle + \langle P_a \rangle \langle P_b \rangle \\
 &= \langle P_a P_b \rangle - \langle P_a \rangle \langle P_b \rangle \\
 &= \langle P_a \rangle \langle P_b \rangle - \langle P_a \rangle \langle P_b \rangle = 0,
 \end{aligned}$$

where under independence: $\langle P_a P_b \rangle = \langle P_a \rangle \langle P_b \rangle$. Intuitively, this result makes sense. The projective measurement of the first Pauli operator maximally randomizes the expectation value of the other Pauli operator and thus the covariance will be zero. Interestingly, the projective measurement causes the underlying distribution of the quantum state to change and so subsequent measurements generating paired samples are not well defined in this setting (for anticommuting operators). This phenomenon is not present in classical experiments. However, the same statistical analysis can be done if we just take pairs of subsequent measurements and only do a statistical analysis on these random variables. We note that our analysis did not have to account for $\langle P_a P_b \rangle$ not being a valid observable for anticommuting Pauli operators due to the resulting operator not being Hermitian and so not a valid observable.

Given the covariance of two anticommuting Pauli operators is zero, we find the variance of a normalized anticommuting clique $\gamma_j C_j$ to be:

$$\begin{aligned}
 Var[\gamma_j C_j^U] &= \gamma_j^2 Var[C_j^U] = \gamma_j^2 Var \left[\sum_i \frac{c_i}{\gamma_j} P_i \right] = \gamma_j^2 \sum_i \sum_k \frac{|C_j^U|}{\gamma_j} \frac{|C_j^U|}{\gamma_j} Cov \left[\frac{c_i}{\gamma_j} P_i, \frac{c_k}{\gamma_j} P_k \right] \\
 &= \gamma_j^2 \sum_i \frac{|C_j^U|}{\gamma_j^2} c_i^2 Var[P_i] + \gamma_j^2 \sum_i \sum_{\substack{k \\ \forall k \neq i}} \frac{|C_j^U|}{\gamma_j} \frac{|C_j^U|}{\gamma_j} \frac{c_i}{\gamma_j} \frac{c_k}{\gamma_j} \underbrace{Cov[P_i, P_k]}_{=0} \quad (4.63) \\
 &= \sum_i \frac{|C_j^U|}{\gamma_j} c_i^2 Var[P_i].
 \end{aligned}$$

³This analysis is strictly for the case of subsequent measurement of anticommuting Pauli operators.

We use this to obtain the following R ratio (equation 4.50):

$$\begin{aligned}
 R = \frac{M_u}{M_g} &= \left(\frac{\sum_j^{N_C} \left[\sum_{k=0}^{|C_j^U|-1} \sqrt{\text{Var}[c_k^{(j)} P_k^{(j)}]} \right]}{\sum_j^{N_C} \sqrt{\text{Var}[\gamma_j C_j^U]}} \right)^2 \\
 &= \left(\frac{\sum_j^{N_C} \left[\sum_{k=0}^{|C_j^U|-1} |c_k^{(j)}| \sqrt{\text{Var}[P_k^{(j)}]} \right]}{\sum_j^{N_C} \sqrt{\sum_{k=0}^{|C_j^U|-1} |c_k^{(j)}|^2 \text{Var}[P_k^{(j)}]}} \right)^2 \\
 &= \left(\frac{\sum_j^{N_C} \left[\sum_{k=0}^{|C_j^U|-1} |x_k^{(j)}| \right]}{\sum_j^{N_C} \sqrt{\sum_{k=0}^{|C_j^U|-1} |x_k^{(j)}|^2}} \right)^2 \\
 &= \left(\frac{\sum_j^{N_C} \|\vec{x}_j\|_1}{\sum_j^{N_C} \|\vec{x}_j\|_2} \right)^2,
 \end{aligned} \tag{4.64}$$

where $x_k^{(j)} = |c_k^{(j)}| \sqrt{\text{Var}[P_k^{(j)}]}$ and $\vec{x}_j = (x_0^{(j)}, x_1^{(j)}, \dots, x_{|C_j^U|-1}^{(j)})$. Minkowski inequality ensures $\|\vec{x}_j\|_2 \leq \|\vec{x}_j\|_1$. At worst unitary partitioning will achieve the same number of measurements as no grouping and will more often achieve an improvement. However, we can actually bound the improvement in general as:

$$\|u\|_1 = \sum_i^\nu |u_i| = \sum_i^\nu |u_i| \cdot 1 \leq \left(\sum_i^\nu |u_i|^2 \right)^{0.5} \cdot \left(\sum_i^\nu 1^2 \right)^{0.5} = \sqrt{\nu} \|u\|_2, \tag{4.65}$$

where the Cauchy-Schwarz inequality has been utilized. Overall, we find $\|u\|_2 \leq \|u\|_1 \leq \sqrt{\nu} \|u\|_2$ and thus:

$$1 \leq R = \frac{M_u}{M_g} = \left(\frac{\sum_j^{N_C} \|\vec{x}_j\|_1}{\sum_j^{N_C} \|\vec{x}_j\|_2} \right)^2 \leq \left(\frac{\sum_j^{N_C} \sqrt{|C_j^U|} \cdot \|\vec{x}_j\|_2}{\sum_j^{N_C} \|\vec{x}_j\|_2} \right)^2. \tag{4.66}$$

This result gives a recipe for how to maximise the possible measurement reduction obtained by unitary partitioning. The numerator of equation 4.66 should be maximised and the denominator minimized. However, evaluating each term requires knowledge of the variance of each Pauli operator. This can either be found precisely on the true ground state or estimated by a classical approximation to the ground state. Instead, we can also use the upper bound of each Pauli: $\text{Var}[P_i] \leq 1$ to estimate equation 4.66. In this scenario, R can be maximised by finding anticommuting cliques that have their 1-norm maximised. This strategy turns out to be the same as discussed by Crawford *et al.* for commuting cliques [201]. Using this idea as motivation, they introduce a heuristic algorithm called ‘‘sorted insertion’’ to maximise R . This begins by ordering the Pauli operators in a Hamiltonian by the absolute magnitude of their coefficients.

In this order, a Pauli operator is added to a collection if it commutes with all terms within that collection. If not, a new collection is created and that operator is inserted there. During enumeration, each collection is checked in order of their creation. The algorithm is terminated once the final Pauli operator is inserted into a clique. This algorithm can be trivially extended to treat anticommuting cliques, where collections of pairwise anticommuting operators are built rather than commuting operators.

The advantage of “sorted insertion” over graph based approaches is it requires a sorting algorithm, such as merge sort that scales as $\mathcal{O}(|H| \log(|H|))$ [202], followed by a two-fold loop over Pauli operators in the Hamiltonian scaling as $\mathcal{O}(|H|^2)$. Whereas, just constructing the graph of a Hamiltonian has a cost of $\mathcal{O}(|H|^2)$ without even considering the cost of subsequent graph colouring algorithms used to find different cliques.

Yen and Izmaylov in [203] found that the sorted insertion method gave the best measurement reduction for a given precision on a set of molecular problems defined on up to 20 qubits.

4.3 Unitary partitioning study

The primary research question during the first half of the Ph.D. was to investigate the performance of the unitary partitioning measurement reduction strategy applied to VQE simulating chemistry problems on an actual quantum device. Two different circuit implementations of this technique are given in [198]. One is achieved by a sequence of rotations and the other by a linear combination of unitaries (LCU). Determining which construction is best and whether unitary partitioning actually reduces the number of measurements needed by VQE was the goal of the numerical study. To study the circuit costs we restrict our analysis to quantum circuits composed of Clifford gates and arbitrary single qubit gates, which is a universal gate set [204].

Continuing from Section 4.2.5.3, once the Hamiltonian has been partitioned into cliques according to equation 4.46, R_j must be coherently performed in circuit before measuring A_j . In the following subsections, we show how R_j is can be built by two different circuit constructions and what the operator A_j being measured is. Without loss of generality, we show this for a single normalized anticommuting clique C_j^U . For ease of reading, we rewrite the definition of the j -th normalized anticommuting clique:

$$C_j^U = \sum_{\substack{k \\ \{P_a, P_b\}=0 \\ \forall P_a, P_b \in C_j^U \\ a \neq b}}^{|C_j^U|} \beta_k P_k, \quad (4.67)$$

where β_k are real coefficients and $\sum_k^{|C_j^U|} \beta_k^2 = 1$.

4.3.1 Sequence of rotations approach

In this subsection, R_j is constructed as a sequence of rotations and follows the work in [198]. First, we define a set of Hermitian self-inverse operators [198]:

$$\{\chi_{\omega k}^{(j)} = iP_{\omega}^{(j)}P_k^{(j)} \mid \forall P_k^{(j)} \in C_j^U \text{ where } k \neq \omega \text{ and } P_{\omega} \in C_j^U\}. \quad (4.68)$$

As every P_l operator in C_j^U anticommutes with all other operators in the set by definition, it is clear from equation 4.68 that $\chi_{\omega k}^{(j)}$ will commute with all $\{P_l^{(j)} \in C_j^U \mid l \neq \omega, k\}$ and anticommute with $P_{\omega}^{(j)}$ and $P_k^{(j)}$. This property is the crux of the sequence of rotations method.

The adjoint rotation of $\chi_{\omega k}^{(j)}$, for a particular k , can be written [198]:

$$R_{\omega k}^{(j)} = e\left(-i\frac{\theta_{\omega k}^{(j)}}{2}\chi_{\omega k}^{(j)}\right), \quad (4.69)$$

where each $\chi_{\omega k}^{(j)}$ is just a real Pauli operator. The action on a single C_j^U is [198]:

$$\begin{aligned} R_{\omega k}^{(j)}C_j^UR_{\omega k}^{\dagger(j)} &= (\beta_k \cos \theta_{\omega k} - \beta_n \sin \theta_{\omega k})P_k^{(j)} + \\ &(\beta_k \sin \theta_{\omega k} + \beta_n \cos \theta_{\omega k})P_n^{(j)} + \sum_{\substack{P_l \in C_j^U \\ \forall l \neq n, k}} \beta_l P_l. \end{aligned} \quad (4.70)$$

The coefficient of $P_k^{(j)}$ can be made to go to 0, by setting $\beta_k^{(j)} \cos \theta_{\omega k}^{(j)} = \beta_n^{(j)} \sin \theta_{\omega k}^{(j)}$. This approach removes the term with index k and increases the coefficient of $P_{\omega}^{(j)}$ from $\beta_{\omega}^{(j)} \mapsto \sqrt{(\beta_{\omega}^{(j)})^2 + (\beta_k^{(j)})^2}$. This process is repeated over all indices excluding $k = \omega$ until only the $P_{\omega}^{(j)}$ term remains.

The whole procedure can be concisely written using the following operator:

$$R_{S_j} = \prod_{\substack{k=0 \\ \forall k \neq \omega}}^{|C_j^U|-1} R_{\omega k}^{(j)}(\theta_{\omega k}), \quad (4.71)$$

which is simply a sequence of rotations. The angle $\theta_{\omega k}$ is defined iteratively at each step of the removal process, as the coefficient of $P_{\omega}^{(l)}$ increases at each step and thus must be taken into account. Importantly the correct solution for $\theta_{\omega k}$ must be chosen given the signs of β_{ω} and β_k [198]. The overall action of this sequence of rotations is:

$$R_{S_j}C_j^UR_{S_j}^{\dagger} = R_{S_j}\left(\sum_{\substack{l \\ \{P_a, P_b\}=0 \\ \forall P_a, P_b \in C_j^U \\ a \neq b}}^{|C_j^U|} \delta_l P_l^{(j)}\right)R_{S_j}^{\dagger} = P_{\omega}^{(j)}. \quad (4.72)$$

In summary, for a Hamiltonian of m terms partitioned into m_c cliques we obtain:

$$\begin{aligned}
 H &= \sum_i^m c_i P_i = \sum_j^{m_c} \gamma_j C_j^U \\
 &= \sum_j^{m_c} \gamma_j (R_{S_j}^\dagger P_\omega^{(j)} R_{S_j}),
 \end{aligned} \tag{4.73}$$

where each C_j^U has been rewritten by applying the result of equation 4.72. The number of terms requiring separate measurement has been reduced from $m \mapsto m_c$. In an experiment, measuring $\langle C_j \rangle$ requires first performing R_{S_j} coherently in circuit followed by measuring $P_\omega^{(j)}$. We cover the circuits required for this next.

4.3.1.1 Sequence of rotations circuit

To implement the sequence of rotations, defined in equation 4.71, each individual Pauli rotation (equation 4.69) must be done in circuit. Whitfield *et al.* in [205] show how such unitary operators can be composed via quantum circuits using standard gates. An example is provided in Figure 4.3.

The additional circuit requirement to perform a single rotation in the sequence of rotations method is $\mathcal{O}(2(n-1))$ CNOT gates and $\mathcal{O}(2n)$ single qubit change of basis gates. The number of terms in an anticommuting clique is upper bounded as $|C_j^U| \leq 2n + 1$ [206] and so the maximum number of single rotations required will be $2n$. Overall, we find that unitary partitioning via a sequence of rotations has an increased circuit cost scaling as $\mathcal{O}(n^2)$ single qubit and CNOT gates.

It is possible to optimize these circuits, as many of the change of basis gates and CNOT gates between each rotation $R_{\omega k}^{(j)}$ (equation 4.69) can be cancelled. As the order of rotations in R_{S_j} (equation 4.71) is free to choose, the possible gate cancellations can be maximized by choosing sequential Pauli operators that share as many single qubit Pauli operators on each individual qubit as possible. Different circuit optimizations are then possible, for example the Phase Gadget Synthesis method [207], which uses tools from ZX-calculus to optimize circuits built as sequential rotations generated by Pauli operators. Furthermore, if the gates are ordered lexicographically then the Pauli matrices in common between Pauli operators will be maximised [208]. We leave such improvements to future work.

4.3.2 Linear combination of unitaries approach

In this section, we describe how R_j is constructed via a linear combination of unitaries (LCU). To motivate the construction, we first summarise the LCU technique proposed by Childs and Wiebe in 2012 [209].

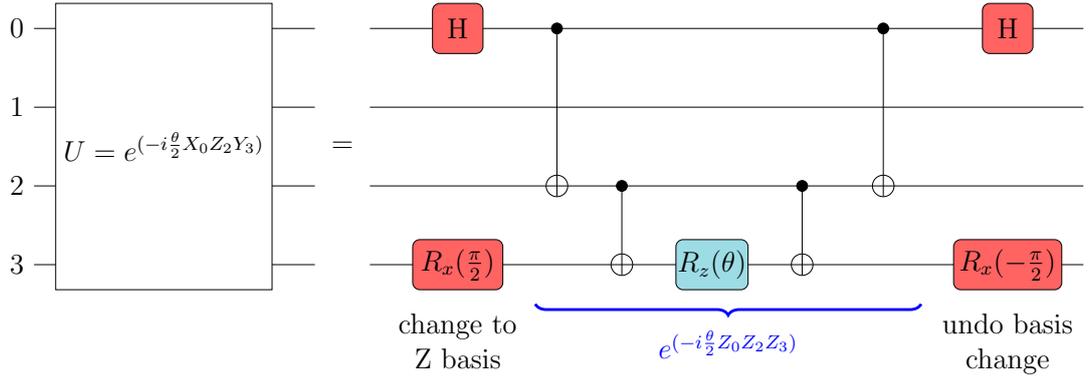


Figure 4.3: Circuit construction to perform rotation generated by a Pauli operator P , in this case $P = XIZY$.

4.3.2.1 The LCU technique

Consider any complex operator A given as some linear combination of d unitary operators [210]:

$$A = \sum_{i=0}^{d-1} \alpha_i U_i, \quad \|A\| \leq \sum_{i=0}^{d-1} \left(|\alpha_i| \cdot \underbrace{\|U_i\|}_{=1} \right) = \sum_{i=0}^{d-1} |\alpha_i| = \|A\|_1, \quad (4.74)$$

where the upper bound on the spectral norm of A is the l_1 norm. Without loss of generality, we can make all $|\alpha_i| > 0$ by absorbing the complex phases into the unitaries U_j . Given a list of α_i and each U_i , which are assumed to be easy to implement as controlled operations on a quantum device, the block encoding can be constructed using the oracles [210]:

$$G = \sum_{i=0}^{d-1} \sqrt{\frac{\alpha_i}{\|A\|_1}} |i\rangle \langle 0|_a + \dots = \begin{bmatrix} \left(\frac{\alpha_0}{\|A\|_1}\right)^{1/2} & \cdot & \dots \\ \left(\frac{\alpha_1}{\|A\|_1}\right)^{1/2} & \cdot & \dots \\ \vdots & \ddots & \dots \\ \left(\frac{\alpha_{d-1}}{\|A\|_1}\right)^{1/2} & \cdot & \dots \end{bmatrix} \quad (4.75)$$

and

$$U_{\text{select}} = \sum_{i=0}^{d-1} \left(|i\rangle \langle i|_a \otimes U_i \right). \quad (4.76)$$

Here subscript s denotes the system register and a the ancilla register. The number of ancilla qubits required will be $n_a = \lceil \log_2(d) \rceil$. In the block encoding literature G and U_{select} are often called the ‘‘Prepare’’ and ‘‘Select’’ oracles [210].

The ‘‘Prepare’’ oracle is a unitary that prepares the state $|G\rangle = \sum_{i=0}^{d-1} \sqrt{\frac{\alpha_i}{\|A\|_1}} |i\rangle$ from the all-zero state on the ancilla register - i.e. $|\bar{0}\rangle \mapsto |G\rangle_a$. This is why in equation 4.75 only the first column is defined. The other columns can be anything so long as they are orthogonal and G remains unitary. This means there is a lot of choice in

4.3. UNITARY PARTITIONING STUDY

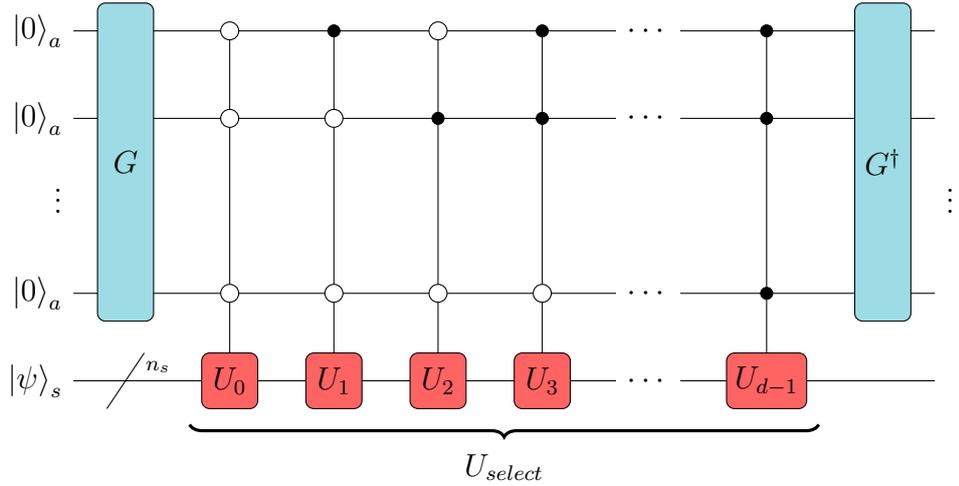


Figure 4.4: General circuit to block encode an operator via a linear combination of unitaries (equation 4.80). For unitary partitioning, this circuit gives a $(\|\vec{\alpha}_q^{(j)}\|_1, \lceil \log_2(|C_j^U| - 1) \rceil, 0)$ - R_j block encoding. Here U_{select} is defined in equation 4.93, where each $P_i^{(j)} = U_i$. Equation 4.92 defines G for a particular anticommuting clique j .

how to construct this operator. A practical note is that if one finds the quantum circuit to do $|\bar{0}\rangle_a \mapsto |G\rangle_a$, then the action of the circuit on the other basis states are automatically accounted for and the whole of G will be defined. This requires arbitrary state preparation, which will be discussed later in Section 4.3.2.3.

We can check these oracles give the correct block encoding, by the following proof [210]:

$$\begin{aligned}
 \left[\langle G|_a \otimes I_s \right] U_{select} \left[|G\rangle_a \otimes I_s \right] &= \left[\langle G|_a \otimes I_s \right] \left(\sum_{i=0}^{d-1} \sqrt{\frac{\alpha_i}{\|A\|_1}} |i\rangle \otimes U_i \right) \\
 &= \left(\sum_{k=0}^{d-1} \sqrt{\frac{\alpha_k}{\|A\|_1}} \langle k| \otimes I_s \right) \left(\sum_{i=0}^{d-1} \sqrt{\frac{\alpha_i}{\|A\|_1}} |i\rangle \otimes U_i \right) \\
 &= \sum_{i=0}^{d-1} \sum_{k=0}^{d-1} \frac{\sqrt{\alpha_i \alpha_k}}{\|A\|_1} \langle k|i\rangle \otimes U_i \\
 &= \frac{1}{\|A\|_1} \sum_{i=0}^{d-1} \alpha_i U_i.
 \end{aligned} \tag{4.77}$$

Figure 4.4 shows how this block encoding is achieved via a quantum circuit. Step-

4.3. UNITARY PARTITIONING STUDY

ping through each operation, we observe the following action:

$$\begin{aligned}
G_a^\dagger U_{\text{select}} G_a |\bar{0}\rangle_a |\psi\rangle_s &= G_a^\dagger U_{\text{select}} |G\rangle_a |\psi\rangle_s \\
&= G_a^\dagger \left[|G\rangle_a \frac{A}{\|A\|_1} |\psi\rangle_s + |G^\perp\rangle_a \sqrt{1 - \left\| \frac{A}{\|A\|_1} |\psi\rangle_s \right\|^2} |\psi\rangle_s \right] \\
&= |\bar{0}\rangle_a \frac{A}{\|A\|_1} |\psi\rangle_s + |\perp\rangle_a \sqrt{1 - \left\| \frac{A}{\|A\|_1} |\psi\rangle_s \right\|^2} |\psi\rangle_s.
\end{aligned} \tag{4.78}$$

Therefore, by post-selecting on the all-zero state in the ancilla register $|\bar{0}\rangle_a$, the system register will be projected onto $\frac{A}{\|A\|_1} |\psi\rangle_s$ and A was successfully applied to the quantum state $|\psi\rangle_s$. If any other state is measured in the ancilla register, then the quantum state will be projected into the wrong part of the Hilbert space and A will not have been applied. This method therefore gives a probabilistic implementation of A , where the probability of success is given by:

$$\begin{aligned}
\mathbb{P}_{\text{success}} &= \langle \psi |_s \langle \bar{0} |_a \frac{A^\dagger}{\|A\|_1} \frac{A}{\|A\|_1} |\psi\rangle_s |\bar{0}\rangle_a \\
&= \left(\frac{1}{\|A\|_1} \right)^2 \langle \psi |_s A^\dagger A |\psi\rangle_s.
\end{aligned} \tag{4.79}$$

As A is usually not unitary, $A^\dagger A$ does not necessarily result in the identity matrix. The probability of success therefore depends on both $\|A\|_1$ and $|\psi\rangle_s$. For example, consider the case of $A = \alpha_0 U_0 + \alpha_1 U_1 = \frac{1}{2} \mathcal{I} + \frac{1}{2} Z = |0\rangle\langle 0|$. This matrix defines the projector onto the all-zero state and is a non-unitary operation. The probability of success when block encoding this matrix will depend heavily on whether $|\psi\rangle_s$ has overlap with $|0\rangle_s$. If $|\psi\rangle_s = |1\rangle_s$, then the probability of success would be zero. Finally, we note that if A is unitary then the probability of success will just be $\|A\|_1^{-2}$.

An alternate way to describe LCU is to see that the oracle $G^\dagger U G$ encodes the matrix A (does not have to be unitary) in the top left block of a larger matrix B that is unitary:

$$B = G^\dagger U_{\text{select}} G = \begin{bmatrix} \frac{A}{\|A\|_1} & * \\ * & * \end{bmatrix}, \tag{4.80}$$

where $*$ denotes different matrices, each determined by the block encoding method. In this diagram, $\frac{A}{\|A\|_1}$ is selected by the projector $|\bar{0}\rangle\langle \bar{0}|_a$. The reason why the block encoding B has A divided by $\|A\|_1$ is due to unitary matrices requiring eigenvalues in the form of $e^{i\theta}$. The spectral norm of a matrix is the maximum singular value of a matrix. As the spectral norm must be less than the 1-norm, normalizing by $\|A\|_1$ ensures that B can be unitary (now dependent on $*$ parts).

A final note on notation. When a matrix, e.g. A , is block encoded using U we usually say “the unitary U gives an (α, k, ϵ) -block encoding of A ”. Here α is a normalization constant applied to the matrix to be block encoded. k is the extra ancilla qubits required to perform the block encoding. For LCU this depends on the number of operators in the linear combination of unitaries (equation 4.74) and scales logarithmically as $k = \lceil \log_2(|A|) \rceil$, where $|A|$ is the number of operators in the linear combination. Finally, ϵ is the error of the block encoding and is determined by [211]:

$$\epsilon = \|A - \alpha \left(\langle \bar{0} |_a B | \bar{0} \rangle_a \right)\|. \quad (4.81)$$

In this work, we focus on exact block encodings - i.e $\epsilon = 0$. Approximate block encoding strategies are also possible [212, 213], but a discussion is outside the scope of this thesis.

4.3.2.2 Unitary partitioning via LCU

To construct R_j via LCU, we first need to manipulate each of the m_c anticommuting sets (C_j^U) that the qubit Hamiltonian was partitioned into (equation 4.45). As before, a particular Pauli operator $P_\omega \in S_l$ in each set is selected to be reduced to. Again this will be denoted by the index ω and written as $P_\omega^{(j)}$, where j indexes the particular clique. To begin the construction as a linear combination of Pauli operators, we first define the operator $C_{j \setminus \omega}^U$:

$$C_{j \setminus \omega}^U = \sum_{\forall k \neq \omega}^{|C_j^U| - 1} \delta_k P_k, \quad (4.82a)$$

$$\text{where } \sum_{\forall k \neq \omega}^{|C_j^U| - 1} \delta_k^2 = 1. \quad (4.82b)$$

$C_{j \setminus \omega}^U$ is simply a normalized anticommuting clique, where the $\beta_\omega^{(j)} P_\omega^{(j)}$ operator has been removed from C_j^U followed by a normalization. Taking each normalised clique C_j^U , we perform the following steps. First, we rewrite C_j^U (equation 4.67) with the term we are reducing to ($\beta_\omega^{(j)} P_\omega^{(j)}$) outside the sum:

$$C_j^U = \beta_\omega^{(j)} P_\omega^{(j)} + \sum_{\substack{l=0 \\ \forall l \neq \omega}}^{|C_{j \setminus \omega}^U| - 2} \beta_l^{(j)} P_l^{(j)}. \quad (4.83)$$

4.3. UNITARY PARTITIONING STUDY

Then re-normalising:

$$C_j^U = \beta_\omega^{(j)} P_\omega^{(j)} + \Omega_j \sum_{\substack{l=0 \\ \forall l \neq \omega}}^{|C_j^U|-2} \delta_l^{(j)} P_l^{(j)}, \quad (4.84a)$$

$$\text{where } \sum_{\substack{l=0 \\ \forall l \neq \omega}}^{|C_j^U|-2} \delta_l^{2(j)} = 1, \quad (4.84b)$$

$$\text{and } \beta_l^{(j)} = \Omega_j \delta_l^{(j)}, \quad (4.84c)$$

we can substitute equation 4.82a into equation 4.84a:

$$C_j^U = \beta_\omega^{(j)} P_\omega^{(j)} + \Omega_j C_{j \setminus \omega}^U, \quad (4.85)$$

where $(\beta_\omega^{(j)})^2 + \Omega_j^2 = 1$. In this form, we can use the trigonometric identity $\cos^2(\theta) + \sin^2(\theta) = 1$ to define the following operator:

$$H_\omega^{(j)} = \cos(\phi_\omega^{(j)}) P_\omega^{(j)} + \sin(\phi_\omega^{(j)}) C_{j \setminus \omega}^U. \quad (4.86)$$

Comparing equations 4.85 and 4.86 it is clear that $\cos(\phi_\omega^{(j)}) = \beta_\omega^{(j)}$ or $\sin(\phi_\omega^{(j)}) = \Omega_j$. Next using the definition of $H_\omega^{(j)}$ in equation 4.86 it was shown in [198] that one can consider rotations of $H_\omega^{(j)}$ around an axis that is Hilbert-Schmidt orthogonal to both $C_{j \setminus \omega}^U$ (equation 4.82a) and $P_\omega^{(j)}$:

$$\chi^{(j)} = \frac{i}{2} [C_{j \setminus \omega}^U, P_\omega^{(j)}] = i \sum_{\substack{k=0 \\ \forall k \neq \omega}}^{|C_j^U|-2} \delta_k^{(j)} P_k^{(j)} P_\omega^{(j)}. \quad (4.87)$$

$\chi^{(j)}$ anticommutes with $H_\omega^{(j)}$, is self-inverse and has the following action [198]:

$$\chi^{(j)} H_\omega^{(j)} = i(-\sin \phi_\omega^{(j)} P_\omega^{(j)} + \cos \phi_\omega^{(j)} C_{j \setminus \omega}^U). \quad (4.88)$$

This defines the rotation:

$$R_j^{LCU} = e\left(-i\frac{\alpha^{(j)}}{2}\chi^{(j)}\right) = \cos\left(\frac{\alpha^{(j)}}{2}\right)\mathcal{I} - i\sin\left(\frac{\alpha^{(j)}}{2}\right)\chi^{(j)} \quad (4.89a)$$

$$= \cos\left(\frac{\alpha^{(j)}}{2}\right)\mathcal{I} - i\sin\left(\frac{\alpha^{(j)}}{2}\right)\left(i\sum_{\substack{k=0 \\ \forall k \neq \omega}}^{|\mathcal{C}_j^U|-2} \delta_k^{(j)} P_k^{(j)} P_\omega^{(j)}\right) \quad (4.89b)$$

$$= \cos\left(\frac{\alpha^{(j)}}{2}\right)\mathcal{I} + \sin\left(\frac{\alpha^{(j)}}{2}\right)\sum_{\substack{k=0 \\ \forall k \neq \omega}}^{|\mathcal{C}_j^U|-2} \delta_k^{(j)} P_{k\omega}^{(j)}, \quad (4.89c)$$

where $P_k^{(j)} P_\omega^{(j)} = P_{k\omega}^{(j)}$. Note the first line of equation 4.89 stems from $\chi^{(j)}$ being a self-inverse Hermitian operator (equation 4.44). Importantly $P_{k\omega}^{(j)}$ will be another tensor product of Pauli operators, as products of n -fold Pauli operators will yield another operator in the Pauli group. Note the multiplication of anticommuting Pauli operators results in each $P_{k\omega}^{(j)}$ having a $\pm i$ phase. The adjoint action of this rotation on $H_\omega^{(j)}$ is:

$$R_j^{LCU} H_\omega^{(j)} R_j^{\dagger LCU} = \sin(\phi_\omega^{(j)} - \theta^{(j)}) C_{j \setminus \omega}^U + \cos(\phi_\omega^{(j)} - \theta^{(j)}) P_\omega^{(j)}. \quad (4.90)$$

By setting $\theta^{(j)} = \phi_\omega^{(j)}$, the coefficient of $C_{j \setminus \omega}^U$ will go to zero and we achieve the intended result of $R_j C_j^U R_j^\dagger = P_\omega^{(j)}$. To build R_j by the LCU method, we use its definition in equation 4.89c. In practice, it is easier to re-write equation 4.89 using the fact that all $P_{k\omega}$ and \mathcal{I} are in the Pauli group. The terms can thus be combined into a single sum:

$$R_j^{LCU} = \alpha^{(j)} \mathcal{I} + \sum_{\substack{k=0 \\ \forall k \neq \omega}}^{|\mathcal{C}_j^U|-2} \alpha_k^{(j)} P_{k\omega}^{(j)} \quad (4.91a)$$

$$= \sum_{\substack{q=0 \\ \forall q \neq \omega}}^{|\mathcal{C}_j^U|-1} \alpha_q^{(j)} P_q^{(j)}. \quad (4.91b)$$

Note all $\alpha_{q^{(j)}}$ must be real and positive for the LCU technique. This is achieved by absorbing negative signs and imaginary phases into each $P_{k\omega}^{(j)}$, hence these operators are n -fold tensor Pauli operators up to a complex phase. When written in this form, it is easy to define the operators G (equation 4.75) and U_{LCU} (equation 4.76):

$$G^{(j)} = \sum_{q=0}^{|C_j^U|-1} \sqrt{\frac{\alpha_q^{(j)}}{\|\vec{\alpha}_q^{(j)}\|_1}} |q\rangle_a \langle 0|_a + \dots = \begin{bmatrix} \left(\frac{\alpha_0^{(j)}}{\|\vec{\alpha}_q^{(j)}\|_1}\right)^{1/2} & \cdot & \dots \\ \left(\frac{\alpha_1^{(j)}}{\|\vec{\alpha}_q^{(j)}\|_1}\right)^{1/2} & \cdot & \dots \\ \vdots & \ddots & \dots \\ \left(\frac{\alpha_{|C_j^U|-1}^{(j)}}{\|\vec{\alpha}_q^{(j)}\|_1}\right)^{1/2} & \cdot & \dots \end{bmatrix}, \quad (4.92)$$

$$U_{LCU}^{(j)} = \sum_{q=0}^{|C_j^U|-1} |q\rangle_a \langle q|_a \otimes P_q^{(j)}, \quad (4.93)$$

that are required to perform R_j as a LCU. Overall the operator is encoded as:

$$\langle \bar{0}|_a G^{(j)\dagger} U_{LCU}^{(j)} G^{(j)} |\bar{0}\rangle_a = \langle G^{(j)}|_a U_{LCU}^{(j)} |G^{(j)}\rangle_a = \frac{R_j^{LCU}}{\|\vec{\alpha}_q\|_1}. \quad (4.94)$$

As R_j^{LCU} is unitary, the probability of success is given by the square of the 1-norm of R_j^{LCU} (equation 4.79). Note that the 1-norm is defined as $\|\vec{\alpha}_q^{(j)}\|_1 = \sum_{q=0}^{|C_j^U|-1} |\alpha_q^{(j)}|$. Figure 4.4 shows the circuit construction to encode R_j via a linear combination of unitaries. Next, we need to consider how G and U_{select} are constructed.

4.3.2.3 Preparation circuit

The prepare operator (equation 4.92) requires being able to generate any real quantum state, on an ancilla register from the all-zero state $|\bar{0}\rangle_a$. There are many different proposals on how to prepare arbitrary quantum states [214–217]. We discuss the approach given in [215–217], which leverages quantum multiplexors.

Quantum multiplexors were first used to describe circuit blocks implementing quantum conditional (*if-then-else*) circuits [216, 218]. *If* true, perform action in the *then* clause and *if* false perform the action in the *else* clause. At the operational level this may be performed by first processing the two clauses in parallel and then multiplexing the output. This is a common primitive in classical computing. The quantum analogue acts in a similar way, but generates linear combinations of *if-else-then* outcomes [216].

Shende *et al.* in [216] show how any normalized quantum state can be efficiently prepared on a quantum computer. The intuition of their approaches comes from recursively disentangling the least significant bit. To understand this, first we need to consider an arbitrary two-dimensional complex vector:

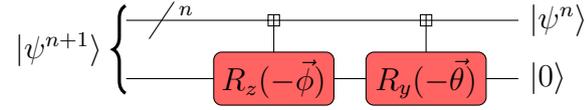


Figure 4.5: Quantum circuit to disentangle an $n + 1$ qubit state (equation 4.100). Figure 4.6b defines the controls of a multiplexor gate and an efficient realisation of the gate is given in Figure 4.6c [216, 219].

$$\begin{aligned}
 |\psi\rangle &= \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \alpha |0\rangle + \beta |1\rangle \\
 &= r \left[\frac{\alpha}{r} |0\rangle + \frac{\beta}{r} |1\rangle \right] \\
 &= r \left[\cos(\theta/2) e^{it} |0\rangle + \sin(\theta/2) e^{ik} |0\rangle \right] \\
 &= r e^{\frac{i}{2}t} e^{\frac{i}{2}k} \left[\cos(\theta/2) e^{\frac{i}{2}(t-k)} |0\rangle + \sin(\theta/2) e^{-\frac{i}{2}(t-k)} |0\rangle \right] \\
 &= r e^{\frac{i}{2}(t+k)} \left[\cos(\theta/2) e^{-\frac{i}{2}(k-t)} |0\rangle + \sin(\theta/2) e^{+\frac{i}{2}(k-t)} |0\rangle \right] \\
 &= r e^{\frac{i}{2}\omega} \left[\cos(\theta/2) e^{-\frac{i}{2}\phi} |0\rangle + \sin(\theta/2) e^{+\frac{i}{2}\phi} |0\rangle \right],
 \end{aligned} \tag{4.95}$$

where:

$$r = \sqrt{|\alpha|^2 + |\beta|^2}, \tag{4.96a}$$

$$t = \arctan 2 \left(\text{Im}(\alpha), \text{Re}(\alpha) \right), \tag{4.96b}$$

$$k = \arctan 2 \left(\text{Im}(\beta), \text{Re}(\beta) \right), \tag{4.96c}$$

$$\theta = 2 \times \arccos \left(\frac{|\alpha|}{r} \right), \tag{4.96d}$$

$$\phi = k - t, \tag{4.96e}$$

$$\omega = t + k. \tag{4.96f}$$

Here r is a normalization constant and $e^{\frac{i}{2}\omega}$ a global phase. The constant factor $r e^{\frac{i}{2}\omega}$ is undetectable and thus we only need to consider the variables θ and ϕ [216]. We can rotate such a 2D complex vector $|\psi\rangle$ as follows [216]:

$$R_y(-\theta) R_z(-\phi) |\psi\rangle = r e^{i\frac{t}{2}} |0\rangle. \tag{4.97}$$

Theorem 9 in [216] proves that equation 4.97 can be used to disentangle the least significant of an arbitrary quantum state. Given any $(n + 1)$ -qubit state, this can

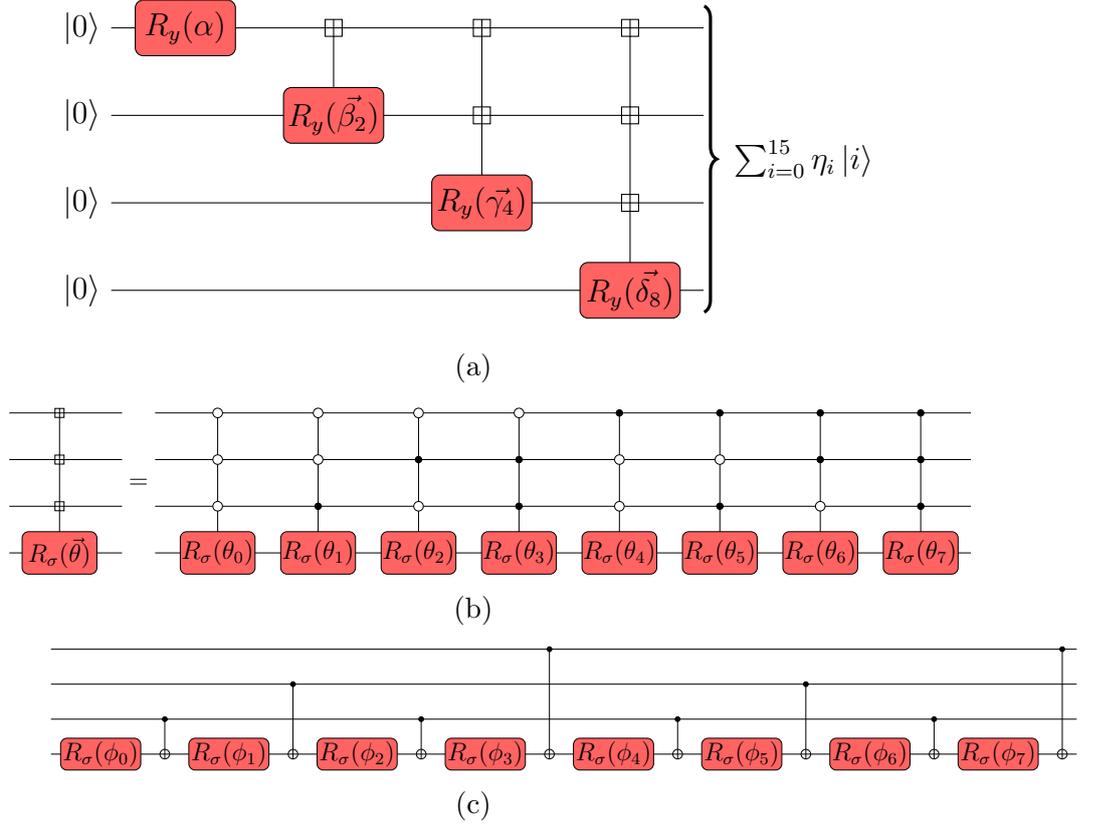


Figure 4.6: (a) Quantum circuit to prepare any real amplitude four-qubit state, (b) uniformly controlled Pauli rotation ($\sigma \in \{X, Y, Z\}$) defined as a multiplexor, (c) Efficient circuit implementation of uniformly controlled multiplexor rotation [215]. The subscript numbers in Figure 4.6a indicate the size of the vector of angles.

be disentangled by the quantum circuit illustrated in Figure 4.5. A 2^{n+1} element vector describing $|\psi^{n+1}\rangle$ is divided into two 2^n sized blocks. Each is interpreted as two-dimensional complex vector, formally:

$$\begin{aligned}
 |\psi^{n+1}\rangle &= \sum_{i=0}^{2^{n+1}-1} \alpha_i |i\rangle = \sum_{i=0}^{2^n-1} \left[|i\rangle \otimes (\alpha_{2i} |0\rangle + \alpha_{2(i+1)} |1\rangle) \right] \\
 &= \sum_{i=0}^{2^n-1} \left[\underbrace{|i\rangle}_{n \text{ qubits}} \otimes \underbrace{|\rho_i\rangle}_{1 \text{ qubit}} \right].
 \end{aligned} \tag{4.98}$$

The angles to disentangle the least significant bit (furthest right) are then found. Each single qubit state $\{|\rho_i\rangle \mid i = 0, 1, \dots, 2^n - 1\}$ is rotated onto $|0\rangle$ by an appropriate θ_i and ϕ_i (equation 4.97). This can be done simultaneously by multiplexing the following unitary:

$$U = \begin{bmatrix} R_y(-\theta_0)R_z(-\phi_0) & 0 & \dots & 0 \\ 0 & R_y(-\theta_1)R_z(-\phi_1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_y(-\theta_{2^n-1})R_z(-\phi_{2^n-1}) \end{bmatrix}, \quad (4.99)$$

that has the following action:

$$U |\psi^{n+1}\rangle = \begin{bmatrix} r_0 e^{i\omega_0} \\ r_1 e^{i\omega_1} \\ \vdots \\ r_{2^n-1} e^{i\omega_{2^n-1}} \end{bmatrix} \otimes |0\rangle = |\psi^n\rangle \otimes |0\rangle. \quad (4.100)$$

Each R_z and R_y operation in U can be performed in parallel via a quantum multiplexor. Figure 4.5 illustrates the circuit template which achieves this. A general rotation multiplexor can be efficiently implemented in parallel via the circuit illustrated in Figure 4.6c. The angles required for each multiplexor are obtained by solving a single matrix-vector problem (see equation 4 in [219]).

For a general state $|\psi\rangle$, these steps can be called recursively on the least significant bit until only a global phase remains. Combining all these steps into a single unitary V and taking the conjugate transpose of the operator will therefore implement $V^\dagger |\bar{0}\rangle = |\psi\rangle$. To obtain the circuit for V^\dagger this involves reversing the circuit for V and negating all the angles in each rotation gate.

The linear combination of unitaries technique only requires a real quantum state to be prepared and so the R_z multiplexor is not required. In [216, 217], it is shown that any real quantum state can be loaded on a quantum device via a sequence of sequential control R_y operations controlled by an increasing number of qubits. A four-qubit example is given in Figure 4.6a. Each n_c -control multiplexor can be decomposed into 2^{n_c} CNOT gates and 2^{n_c} single qubit R_y rotations [215, 216]. Figure 4.6b shows an example of a R_y multiplexor with three control qubits ($n_c = 3$). Any real n quantum state (2^n real amplitudes) can thus be prepared on a quantum computer with the number of single and CNOT gates scaling as [215, 216]:

$$\# \text{ single qubit gates} = \sum_{i=0}^n 2^i = 2^{n+1} - 1 \implies \mathcal{O}(2^n), \quad (4.101a)$$

$$\# \text{ CNOT gates} = \sum_{i=1}^n 2^i = 2^{n+1} - 2 \implies \mathcal{O}(2^n). \quad (4.101b)$$

For unitary partitioning, as the size of R_j^{LCU} scales with the size of the anti-

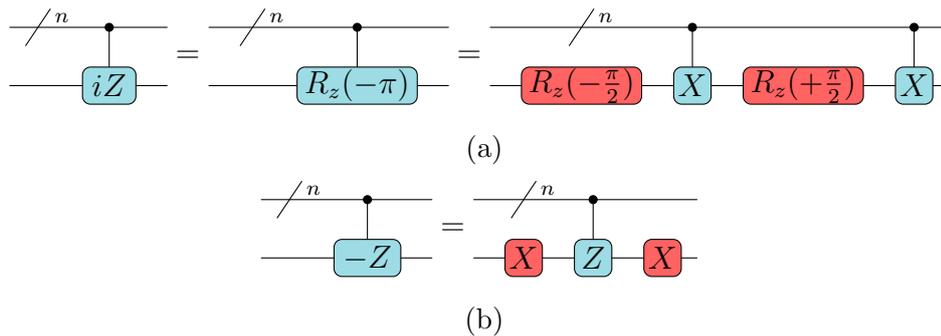


Figure 4.7: Circuit primitives to implement imaginary (a) and negative (b) n -control single qubit Pauli matrices using standard gates [220]. Equation 4.102a and equation 4.102b give the logic behind these operations.

commuting clique, explicitly: $|C_j^U| - 1$ (equation 4.92), the number of ancilla qubits required will scale as $\log_2[|C_j^U| - 1]$ or $\mathcal{O}(\log_2(|C_j^U|))$. As the number of anticommuting operators on n_s qubits must be less than or equal to $2n_s + 1$, the number of prepare (ancillary) qubits required will scale as $\mathcal{O}(\log_2(n_s))$. This is the qubit register that requires being able to implement arbitrary state preparation on. Due to the number of prepare qubits required scaling logarithmically with the number of system qubits, the gate cost for the prepare oracle will have a circuit cost scaling as $\mathcal{O}(2^{\log_2(n_s)}) = \mathcal{O}(n_s)$ single qubit and CNOT gates. This linear gate cost in system qubits makes the technique practical for near-term use.

4.3.2.4 Circuit construction of select operator in unitary partitioning

To implement the select oracle (equation 4.93) via quantum circuits, we need to be able to implement control Pauli gates that have the phase and sign encoded at the operator level. To do this, we can utilize commutation properties within the Pauli group. For the single qubit Pauli matrices, we utilize the following relations:

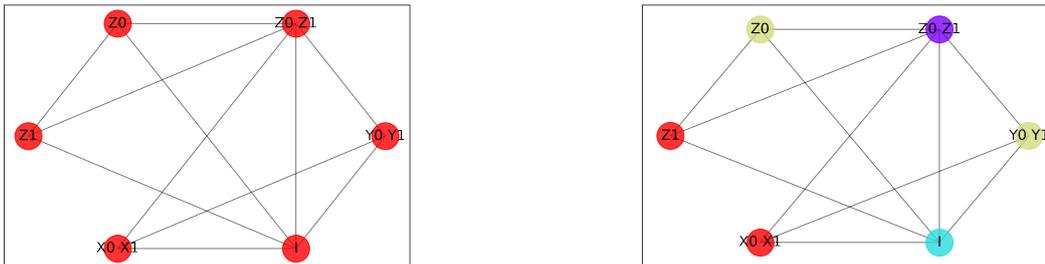
$$-\sigma_k = \sigma_j \sigma_k \sigma_j \quad \forall \sigma_k \neq \sigma_j \text{ and } \sigma_k, \sigma_j \in \{X, Y, Z\}, \quad (4.102a)$$

$$R_{\sigma_k}(\mp\pi) = e^{\mp i \frac{\pi}{2} \sigma_k} = \pm i \sigma_k \text{ where } \sigma_k \in \{X, Y, Z\}. \quad (4.102b)$$

Figure 4.7 shows how to implement these rules within quantum circuits. For now we leave this topic, but in Section 4.3.4 we revisit these circuits and show how they can be improved upon. Next, we report our numerical study on unitary partitioning.

4.3.3 Numerical study

The unitary partitioning measurement reduction is dependent on the problem Hamiltonian. To assess the performance, we applied the method to two chemical Hamilto-



(a) Graph of qubit Hamiltonian of H_2 with commuting Pauli operators joined by edges

(b) Graph colouring of Graph (a), each colour represents a clique.

Figure 4.8: Illustration of graph approach to finding anticommuting sets of P_i operators.

nians. This section summarises these results, which have been published in [221].

4.3.3.1 Method

We consider Hamiltonians for H_2 and LiH molecules employing the *STO-3G* and *STO-6G* basis sets respectively. These were calculated using Openfermion-PySCF and converted into the qubit Hamiltonian using the Bravyi-Kitaev transformation in OpenFermion [222–224]. The raw details on each Hamiltonian may be found online in the Supporting Material of the publication in [221].

NetworkX [225] was used to partition each molecular Hamiltonian, described as a linear combination of Pauli operators, into, anticommuting sets C_j^U . First, a graph of the qubit Hamiltonian was built, where each node is a term in the Hamiltonian. Next edges are put between nodes on the graph that commute. A graph colouring was then performed. The “largest first” colouring strategy in NetworkX was used [225, 226]. Each unique colour represents an anticommuting clique. Fig. 4.8 shows the method applied to H_2 .

The input state $|\psi(\vec{\theta})\rangle$ for all calculations was the exact full configuration interaction (FCI) ground state, found by diagonalizing the Hamiltonian. As our aim was to investigate different implementations of unitary partitioning, this meant the Ansatz optimization step in VQE was not required.

For the simulations performed on IBM’s *ibmqx2* quantum processing unit (QPU), a measurement error mitigation strategy available in Qiskit was utilized and is a simple inversion procedure [227]. The quantum circuits required were generated by the `qiskit.ignis.complete_meas_cal` method and executed alongside each separate *ibmqx2* experiment, with the maximum number of shots (8192). This sampling cost was not included in the number of calls to quantum device. The *CompleteMeasFitter*

4.3. UNITARY PARTITIONING STUDY

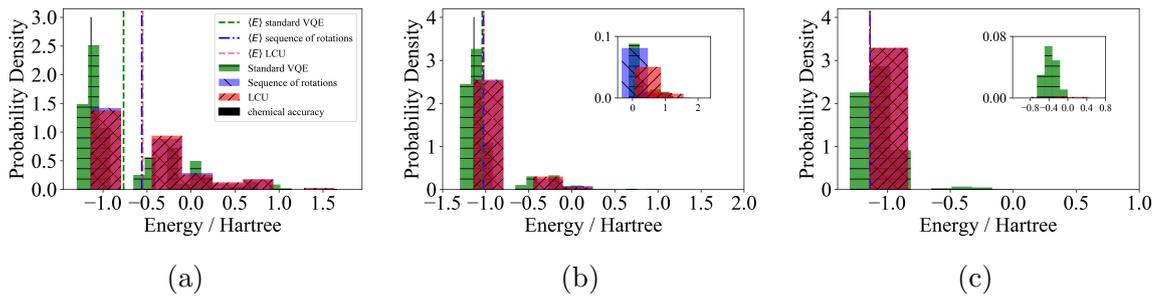


Figure 4.9: Probability density functions of single-shot VQE estimates of the ground state energy e_j of H_2 . A bin is given to every possible energy outcome. Note $E_{FCI} = -1.13728$ Ha. The raw results from `ibmqx2` are given in (a), (b) shows these results with measurement error mitigation applied and (c) gives results from simulation on a noise-free QPU.

method in `qiskit.ignis` [228] was used to generate the calibration matrix required for measurement error mitigation [227, 228].

We denote the number of terms in the original Hamiltonian by t_o and t_p for the unitary partitioned Hamiltonian. For each implementation, we fix M the total number of calls to the QPU. This can be thought of as a measurement budget. The total number of e_j samples - single shot estimates of all n -fold Pauli operators in the Hamiltonian - is $N_o = M/t_o$ for the original Hamiltonian and $N_p = M/t_p$ for the partitioned Hamiltonian. Clearly, because unitary partitioning reduces the number of terms in H_q , more energy samples are obtained for a fixed M .

4.3.3.2 Results

For a given preparation of the true ground state of H_2 , we compare both implementations of measurement reduction by unitary partitioning against a standard VQE calculation on IBM’s open access quantum device (`ibmq 5 Yorktown - ibmqx2`) and Qiskit’s `qasm` simulator [228]. Fig. 4.9 shows the distribution of single-shot energy estimates of all three techniques applied to molecular hydrogen. The average energy is given by $\langle E \rangle = \frac{1}{N} \sum_{j=0}^{N-1} e_j$. To compare each method the measurement budget was fixed to $M = 1.2663 \times 10^6$. A calibration matrix method available in Qiskit was used to mitigate measurement errors by amending the raw outputs from `ibmqx2`.

The qubit Hamiltonian for H_2 has five terms, which is reduced to three by unitary partitioning, not including the identity term. The number of energy estimates e_j obtained was 253260 for standard VQE and 5/3 this for unitary partitioning by the sequence of rotations method. This is because the smaller number of terms allowed a correspondingly larger number of samples to be taken for a fixed M . The total number of e_j samples from `ibmqx2` for these techniques was reduced to 253074 and 421951 after measurement error mitigation was applied. The LCU approach to unitary partitioning is probabilistic and requires post-selection on the all-zero state of the ancilla register.

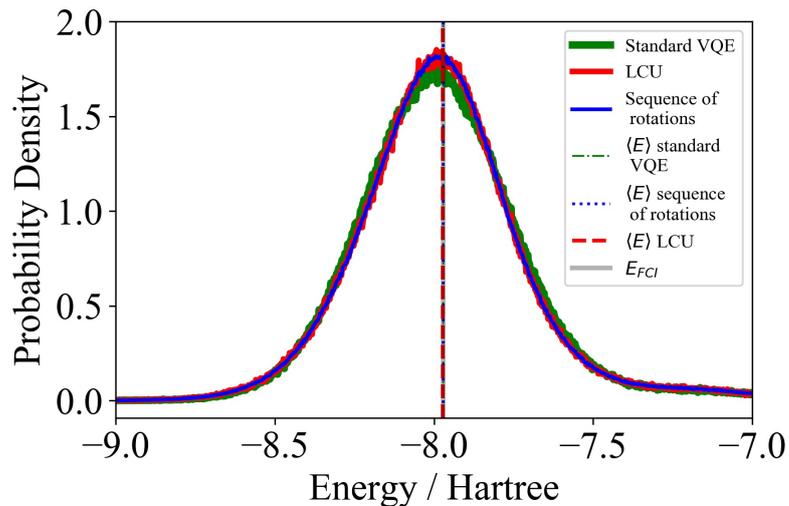


Figure 4.10: Probability density function of single-shot VQE estimates of the ground state energy e_j of LiH, from a noise-free QPU simulation. The number of bins was set to 2500 and the centre of each plotted. Note $E_{FCI} = -7.97118$ Ha.

After post-selection, our simulation of unitary partitioning as a LCU on ibmqx2 gave 333407 raw e_j samples and 332763 e_j samples after measurement error mitigation was applied to the raw output. Our emulation of this method on a noise-free quantum processing unit (QPU) gave 336390 e_j samples after post-selection. The theoretical maximum possible number of samples for LCU would be the same as the sequence of rotations method if all samples obtained were successful.

Figure 4.9 summarises the results of the H_2 study. The reason a normal distribution is not obtained is due to the number of terms in the qubit Hamiltonian being so few. At most only 32 distinct values of e_j are possible for standard VQE and 8 under unitary partitioning, and so we do not expect the central limit argument to apply here.

To investigate the distribution of energies obtained from each method in more detail, we simulated the larger problem of LiH using Qiskit’s statevector simulator [228]. Fig. 4.10 summarizes the results. Again, each data point is an energy estimate found from the weighted measurement outcomes of a single-shot VQE run. The standard qubit Hamiltonian for this problem is made up of 630 terms, which after applying unitary partitioning is then composed of 102 terms. These counts do not include the identity term. The measurement budget was fixed at $M = 1.018521 \times 10^9$. The total number of energy estimates e_j for standard VQE, the sequence of rotations and the LCU methods after post-selection were 1616700, 9985500 and 1447349 respectively.

We performed the Kolmogorov-Smirnov [229] and Shapiro-Wilk tests [230] on the data in Fig. 4.10 to check for normality. The p -values obtained in all cases were smaller than 0.05, and thus we could not assume a normal distribution. This may be

caused by insufficient samples allowing convergence to the central limit or the problem size still being too small. To estimate the statistics of the true distribution we thus employed bootstrap resampling [231].

4.3.3.3 Discussion

In our results, for a fixed measurement budget M we obtain a set of independent identically distributed random energy samples $\{e_1, e_2, \dots, e_N\}$. The standard deviation of this sample σ_e converges to the true standard deviation as the number of samples increases. As the number of samples increases, the error on the sample mean decreases. The standard error of the mean (SEM) is defined:

$$\text{SEM} = \sigma_e / \sqrt{N}, \quad (4.103)$$

for N energy samples and is the square root of equation 4.9. For a fixed measurement budget M , if the SEM is reduced by unitary partitioning over standard VQE then the method has led to an improvement.

To benchmark each method, we compare σ_e and SEM of the ground state energies samples. 95% confidence intervals (CI) were calculated using bootstrapping with 10,000 resamples with replacement. The full statistical analysis is given in Table 4.2.

Qualitatively, the noise-free LiH simulation results in Fig. 4.10 show that VQE with unitary partitioning applied as either a LCU or a sequence of rotations give a similar distribution of energies compared to standard VQE.

In the original work these techniques were proposed, it was shown that the variance of the different methods should be similar [198]. We observe this on both the QPU emulator and quantum device. Quantitatively, the σ_e of ground state energy estimates of LiH for each method were very similar, with the largest difference being 4.8 mHa. We see this in Figure 4.10, with each distribution looking very similar. However, it is important to note the number of data points in each plot is significantly different from one another and therefore the SEM is different for each implementation. Whereas, for the noise-free simulation of H₂ (Fig. 4.9c), the sample standard deviation of e_j from VQE with unitary partitioning applied were an order of magnitude lower than standard VQE. We expect this is due to the small number of distinct e_j outcomes for this specific problem under unitary partitioning.

We calculated the R -score (equation 4.64) for each system and obtained a value of 1.99 and 3.11 for H₂ and LiH respectively. To do this the variance of each Pauli operator was required. These were calculated analytically on classical hardware with reference to the FCI ground state (equation 4.8). Unitary partitioning is therefore expected to give a measurement reduction by a factor of 2 and 3 respectively. For the sequence of rotations method, our noiseless simulations show that the SEM of both

4.3. UNITARY PARTITIONING STUDY

Molecule	Method	Backend	N	$\langle E \rangle / Ha$	$\langle E \rangle$ 95%CI / Ha	σ_e / Ha	σ_e 95%CI / Ha	SEM / Ha	SEM 95%CI / Ha
H ₂	LCU	ibmqx2 - mit	332763	-1.021	[-1.022, -1.020]	3.166e-01	[3.150e-01, 3.182e-01]	5.489e-04	[5.462e-04, 5.516e-04]
H ₂	SeqRot	ibmqx2 - mit	421951	-1.025	[-1.026, -1.024]	3.198e-01	[3.184e-01, 3.213e-01]	4.924e-04	[4.902e-04, 4.946e-04]
H ₂	standard	ibmqx2 - mit	253074	-1.038	[-1.039, -1.037]	3.424e-01	[3.405e-01, 3.442e-01]	6.806e-04	[6.770e-04, 6.843e-04]
H ₂	LCU	ibmqx2 - raw	333407	-0.545	[-0.548, -0.543]	6.529e-01	[6.513e-01, 6.544e-01]	1.131e-03	[1.128e-03, 1.133e-03]
H ₂	SeqRot	ibmqx2 - raw	422100	-0.559	[-0.561, -0.557]	6.503e-01	[6.489e-01, 6.516e-01]	1.001e-03	[9.987e-04, 1.003e-03]
H ₂	standard	ibmqx2 - raw	253260	-0.765	[-0.767, -0.763]	5.663e-01	[5.644e-01, 5.681e-01]	1.125e-03	[1.122e-03, 1.129e-03]
H ₂	LCU	simulator	336390	-1.137	[-1.137, -1.137]	1.566e-02	[1.431e-02, 1.708e-02]	2.699e-05	[2.469e-05, 2.939e-05]
H ₂	SeqRot	simulator	422100	-1.137	[-1.137, -1.137]	1.629e-02	[1.508e-02, 1.753e-02]	2.507e-05	[2.315e-05, 2.697e-05]
H ₂	standard	simulator	253260	-1.137	[-1.138, -1.136]	1.775e-01	[1.763e-01, 1.787e-01]	3.528e-04	[3.504e-04, 3.551e-04]
LiH	LCU	simulator	1447349	-7.972	[-7.972, -7.971]	2.727e-01	[2.721e-01, 2.732e-01]	2.267e-04	[2.262e-04, 2.271e-04]
LiH	SeqRot	simulator	9985500	-7.971	[-7.971, -7.971]	2.729e-01	[2.727e-01, 2.731e-01]	8.637e-05	[8.630e-05, 8.643e-05]
LiH	standard	simulator	1616700.	-7.972	[-7.972, -7.971]	2.682e-01	[2.677e-01, 2.686e-01]	2.109e-04	[2.106e-04, 2.113e-04]

Table 4.2: The mean, standard deviation and standard error on the mean for each method calculating the ground state energies of H₂ and LiH using single-shot VQE. The simulator backend represents a noise-free QPU emulator and ibmqx2 is a real quantum device. Ibmqx2-raw are the raw experimental results from the QPU and ibmqx2-mit with measurement error mitigation applied. 95% confidence intervals (CI) were calculated using bootstrap resampling [231].

H₂ and LiH was an order of magnitude lower than standard VQE. On the other hand, the LCU realization of unitary partitioning is probabilistic. Even though the R -score indicates an improvement should be obtained, post-selection requires some samples to be discarded. We see this in the simulation for LiH, where the LCU approach actually has the fewest e_j samples at 1447349 compared to 1616700 for standard VQE. As the σ_e of all three approaches are similar, the LCU implementation has the highest SEM in this case. The advantage over standard VQE is thus dependent on the success probability, which for each circuit is given by the inverse of the 1-norm squared of the operator to be implemented as a LCU. We re-iterate, post-selection is only performed on the ancilla register.

The experimental results for H₂ on ibmqx2 show that applying the unitary partitioning technique does not appreciably change the performance of VQE, when combined with error mitigation techniques. We suspect this is due to the extra coherent resource required to perform R_l causing an increase in errors, which offsets the improvement of the SEM given by the technique. We expect this to be mitigated as gate fidelities increase.

The experimental execution of R_l by LCU on ibmqx2 performed comparably to the sequence of rotations realization. Ignoring post-selection issues, the LCU algorithm is more complex and requires more qubits to implement. We believe this motivates further examination of the use of more advanced quantum algorithms on NISQ devices.

A particular feature of our results from ibmqx2 (Fig. 4.9) is that the mean ground state energy obtained is overestimated by a seemingly constant amount. We suspect this could be due to two effects. Firstly, our ansatz circuit prepares the ground state. Any coherent errors in this circuit will increase the energy of the state prepared by virtue of the variational principle [18]. Secondly, inspecting the qubit Hamiltonian for H₂ most coefficients are positive. As our results overestimate the energy, it implies

that measurement outcomes of each n -fold Pauli operator are more frequently $+1$ causing each estimate of e_j to be larger. This could be an indication of a higher $|0\rangle$ count on each qubit or $P(0|1) > P(1|0)$.

The single-qubit gate error rates of IBM QPU's have error rates in the range of 0.1%-0.3% and two-qubit gate errors in the range of 2%-5% [232]. The most error-prone operation is measurement and `ibmqx2` on average has a measurement error rate of 4%, but this can be much higher (13%) [232]. This large measurement error is apparent when comparing the raw and measurement error mitigated results from the QPU simulation of H_2 . In future experiments, it would be interesting to improve measurement fidelity, for example by using invert-and-measure designs [232] as well as flipping the qubit encoding ($|0\rangle \mapsto |1\rangle$ and $|1\rangle \mapsto |0\rangle$) as in [106], or by other mitigation schemes [233].

Crucially, when partitioning the qubit Hamiltonian into anticommuting cliques, the greatest measurement reduction is obtained if R is maximised (equation 4.64). However, if non-optimal partitioning is done a measurement reduction can still be obtained. We propose that for practical applications a non-optimal clique cover may be beneficial. By splitting the problem Hamiltonian into pairs of anticommuting operators ($|H_{S_l}| = 2 \forall \{l\}_{l=0,1,\dots,m_c-1}$), the extra coherent resources required to perform R_l are experimentally realistic for current and near-term devices. The choice of partitioning can be motivated by estimating the value of R (equation 4.64). To do this the variance and covariances of the Pauli operators can be estimated on quantum hardware or classically determined using an approximate ground state wavefunction. If this is not possible, then the upper bound of the variance of each Pauli operator can be used (equation 4.8).

4.3.4 LCU circuit optimization

In this section, we provide a different analysis of how to improve the quantum circuits required to implement the select operator in LCU more efficiently. The gate set we consider in this section are arbitrary single qubit gates and CNOT gates. We do not discuss the Prepare part of the algorithm, which we analysed previously in Section 4.3.2.3.

Efficient block encoding strategies are an important primitive in many of the most recent quantum algorithms. This topic is outside the scope of this thesis, but we recommend the review by Martyn *et al.* on the grand unification of quantum algorithms [234].

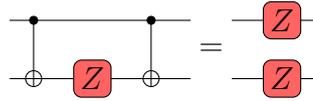


Figure 4.11: Circuit template for $Z \cdot Z$ operation.

4.3.4.1 Select circuit

In Section 4.3.2.3 we showed how the prepare part of the LCU approach to unitary partitioning scales linearly with the number of system qubits and is therefore tractable. The next consideration is the select operator (equation 4.93). The complicated control structure of this operator makes this operation hard to experimentally realise. Here we analyse how to reduce the quantum resources to implement this gate. We report a new approach that improves upon the original work in [221], requiring no extra ancillary qubits.

Naively, each n_c -control Pauli operator can be decomposed into separate n_c -control single qubit Pauli operators on each n_s qubit. Up to a change of basis, these will then require n_c controlled single qubit Z gate decompositions. However, this can be improved. We analyse two distinct methods.

The first method requires the following circuit primitive illustrated in Figure 4.11. We prove this identity as follows. Consider the CNOT gate decomposed in the Pauli basis as

$$\begin{aligned}
 CNOT(i, j) &= \left(|0\rangle\langle 0| \otimes I_j + |1\rangle\langle 1| \otimes X_j \right) \\
 &= \left(\frac{(I_i + Z_i)}{2} \otimes I_j + \frac{(I_i - Z_i)}{2} \otimes X_j \right).
 \end{aligned} \tag{4.104}$$

4.3. UNITARY PARTITIONING STUDY

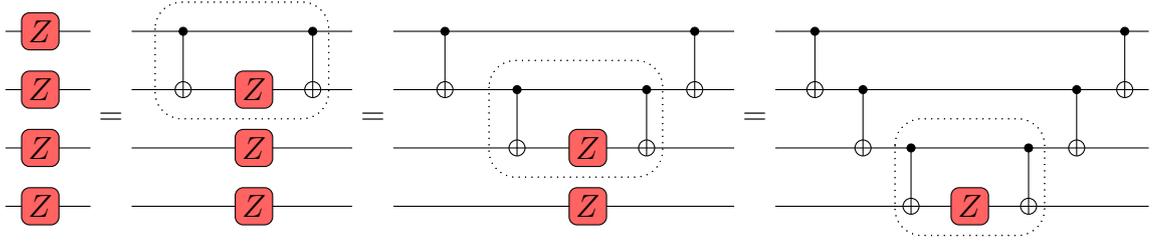


Figure 4.12: Result of recursively applying the template given in Figure 4.11 to a sequence of single qubit Z gates.

Then the left hand side of Figure 4.11 implements the following:

$$\begin{aligned}
 & CNOT(0,1) \cdot I_0 Z_1 \cdot CNOT(0,1) \\
 &= \left(|0\rangle \langle 0| \otimes I + |1\rangle \langle 1| \otimes X \right) \cdot (I \otimes Z) \cdot \left(|0\rangle \langle 0| \otimes I + |1\rangle \langle 1| \otimes X \right) \\
 &= \left(|0\rangle \langle 0| \otimes Z - |1\rangle \langle 1| \otimes iY \right) \left(|0\rangle \langle 0| \otimes I + |1\rangle \langle 1| \otimes X \right) \\
 &= \left(|0\rangle \langle 0| \otimes Z - |1\rangle \langle 1| \otimes Z \right) \\
 &= \frac{1}{2} \left((I + Z) \otimes Z - (I - Z) \otimes Z \right) \\
 &= \frac{1}{2} \left(IZ + ZZ - IZ + ZZ \right) \\
 &= \frac{2(ZZ)}{2} = ZZ.
 \end{aligned} \tag{4.105}$$

We can recursively apply this result on a set of single Pauli Z operators acting on independent qubits - see Figure 4.12. The importance of this circuit is it allows an efficient construction of a multicontrol Pauli operator composed of single qubit Z and I terms. Consider the 3-qubit version of Figure 4.12, controlled on an arbitrary number of n_c qubits. We obtain the circuit illustrated in Figure 4.13. The n_c -controls on each CNOT gate, in the furthest right circuit in Figure 4.13, can be removed due to their trivial action if the n_c -control Z operator does not act. We observe that the three n_c -control Z gates have been converted to a ladder of CNOT gates acting on the system qubits and a single n_c -control Z gate.

In general, we find that an n_c control P operator, where $P \in \{I, Z\}^{\otimes n_s}$, can be decomposed into a ladder of $2(n_s - 1)$ CNOT gates and a single n_c -control single qubit Z gate. A change of basis on each system qubit, using the single qubit gates $\{H, S^\dagger, S\}$, applied on either side of the CNOT ladders can then be used to transform the control Pauli operator made up of single qubit Pauli Z and I terms into a general Pauli operator $P \in \{I, X, Y, Z\}^{\otimes n_s}$. Figure 4.14 shows the type of circuit. Therefore,

4.3. UNITARY PARTITIONING STUDY

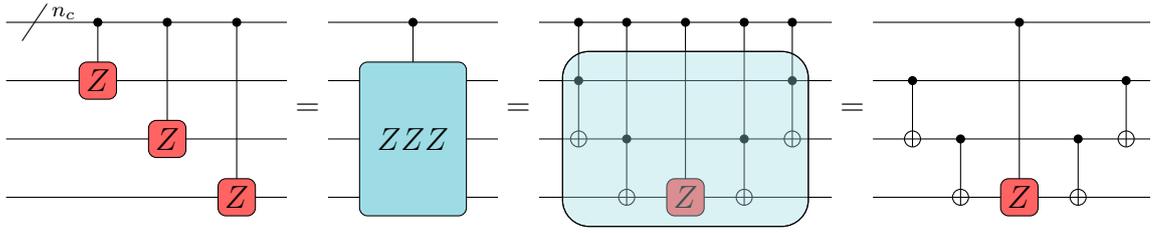


Figure 4.13: Circuit template of a n_c -control P , where $P \in \{I, Z\}^{\otimes 3}$.

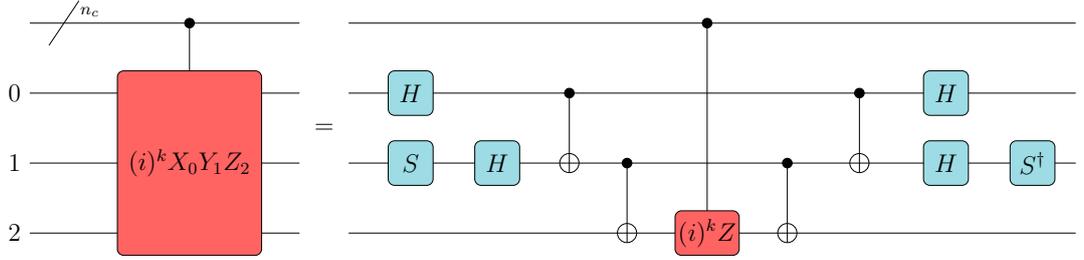


Figure 4.14: Circuit template to implement a multicontrol Pauli operator, where the single qubit $(i)^k Z$ gate for $k \in \{0, 1, 2, 3\}$ can be built via Figure 4.7.

rather than naively requiring a single multicontrol n_c - Z gate per system qubit only a single one is required and a ladder of CNOT gates. This approach was given in [221]; however an interesting circuit template was missed. R_j^{LCU} (equation 4.89) is composed of a real Identity Pauli operator and a linear combination of $P_{k\omega}^{(j)}$ terms that each has an imaginary phase. A control operation for the first term is unnecessary as a controlled identity has no action and only needs to be accounted for in the prepare operation. For the remaining operators, a multicontrol $R_z(\mp\pi) = \pm iZ$ can be used in combination with the template in Figure 4.14, replacing the multicontrol n_c - Z gate. The multicontrol R_z gate can then efficiently be implemented via a quantum multiplexor - see Figure 4.6b and Figure 4.6c. The angles in each single qubit R_z gate should be set to $\theta_a = \frac{(-1)^a(\mp\pi)}{2^{n_c}}$, where a is an index of the rotation angle in $\vec{\theta}$ and n_c is the number of control qubits. Figure 4.15 shows an example with three control qubits. The gate cost for this multiplexor is given in equation 4.101. The cost of implementing the template is illustrated in Figure 4.14, where $k \in 1, 3$, has a cost scaling as $\mathcal{O}(2^{n_c} + n_s)$ CNOT and single qubit gates. Again as discussed in Section 4.3.2.3, the number of control qubits in the LCU method scale as $\mathcal{O}(\log_2(n_s))$ and so the overall cost to implement R_j^{LCU} (equation 4.89) scales as $\mathcal{O}(|C_j^U|n_s)$ CNOT and single qubit gates.

The second approach uses the same process as before, except the single n_c -control Z gate in Figure 4.14 is decomposed in a slightly different manner. The main idea stems from the fact that $P \in \{I, Z\}^{\otimes n_s}$ unitary operators and their controlled versions, are diagonal operators. In [218] Bullock and Markov show that any n -qubit diagonal gate can be implemented in $2^{n+1} - 3$ alternating controlled NOT gates and single qubit

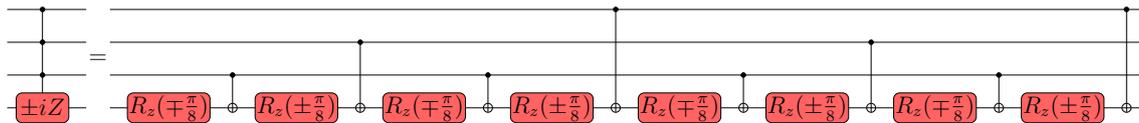


Figure 4.15: Example of a multicontrol $\pm iZ$ gate, implemented via a multicontrol $R_z(\mp\pi)$ multiplexor.

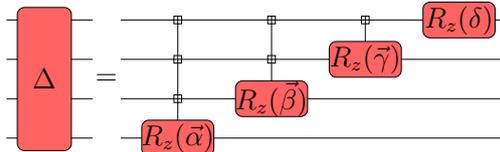


Figure 4.16: Decomposition of a diagonal unitary Δ as a sequence of multicontrol R_z multiplexors [218].

R_z rotations. This is asymptotically optimal [218]. An example is shown in Figure 4.16. Therefore, implementing a single n_c -control $i^k Z$ gate requires $2^{n_c+2} - 3$ single and two qubit gates. Again, as the number of control gates scales logarithmically with the number of system qubits we find this has a CNOT and single qubit gate cost scaling linearly as $\mathcal{O}(n_s)$. Therefore, the overall circuit cost to implement R_j^{LCU} (equation 4.89) scales the same as the first method - $\mathcal{O}(|C_j^U|n_s)$ CNOT and single qubit gates. While this method has a slightly larger circuit cost in practice compared to the previous approach, we report this methodology as the first method requires Pauli operators with an imaginary phase $\pm iP$ ($k \in \{1, 3\}$ in Figure 4.14) whereas this formulation has no such restriction - i.e. $k \in \{0, 1, 2, 3\}$.

4.4 Conclusion

In this chapter, we examined why the $\mathcal{O}(K^4)$ scaling of the second quantized molecular Hamiltonian led to the number of measurements required to achieve a precision of ϵ scaling as $\mathcal{O}(K^8/\epsilon^2)$. Due to the desired precision for most chemistry applications being $\epsilon = 1.6 \times 10^{-3}$ Ha, this meant that even though VQE appears to scale polynomially with problem size, the sheer number of measurement samples required can cause the runtime of the algorithm to take many years. This ignores other problems associated with VQE, such as what Ansatz to use and how to perform the classical optimization. Chapter 3 provides an introduction to some of these concerns. Furthermore, these issues are compounded by the lack of an error-corrected quantum device [235].

We analysed the performance of the unitary partitioning method that in principle decreases the number of Pauli operators requiring separate measurement from $\mathcal{O}(K^4) \mapsto \mathcal{O}(K^3)$. Our numerical results showed that the unitary partitioning tech-

nique can significantly improve the precision of variational calculations. For a fixed measurement budget M , fewer terms required measurement and thus the total number of separate energy estimates was increased. As the sample standard deviation of energies is similar for the different approaches, the standard error of the mean will be lower when unitary partitioning is applied. In general, we expect the R ratio to be greater than one and so a measurement reduction should be obtained. Our results indicate that the deterministic sequence of rotations implementation offers the best improvement, which we find in our noiseless simulation of H_2 and LiH . In contrast, the LCU approach is probabilistic and some measurements must be discarded. The advantage over standard VQE is thus dependent on the success probability. This naive implementation of LCU can be improved by using oblivious and standard amplitude amplification [236–239], which can boost the probability of success. However, further coherent resources would be required and this is unlikely to be applicable for near-term applications.

We note the LCU technique allows non-unitary operations to be implemented probabilistically on a quantum computer. However, for unitary partitioning the operator we block encode R_j^{LCU} (equation 4.91), which is already unitary. This means it can be deterministically implemented as a quantum gate and so should not require a block encoding. The only complication is that it is currently not known how to construct a quantum circuit for a linear combination of Pauli operators that define a unitary operator. Instead, as the operator is already provided as a linear combination of Pauli operators it is natural to use the LCU method. It would be interesting to determine whether an efficient circuit for R_j^{LCU} was possible.

The experimental results obtained using IBM’s NISQ device (ibmqx2) show VQE with unitary partitioning applied performs no worse than conventional VQE for a fixed M , when combined with error mitigation techniques. Even though unitary partitioning requires fewer terms to be combined to give an energy estimate leading to less statistical noise and more energy samples, we suspect the additional coherent resources required causes an increased error accumulation, which offsets the advantages given by the technique. As quantum devices continue to improve, this effect should be reduced and we expect unitary partitioning will benefit many variational quantum algorithms.

Our work shows how precision can be improved for a fixed number of calls to a QPU. An alternate outlook, is how this technique may allow larger problems to be studied. For a given precision, applying unitary partitioning requires fewer samples and thus may allow larger scale simulations to be performed on reasonable timescales. However, we note that the issue of the measurement roadblock for VQE still remains even when using unitary partitioning. The formal scaling in the number of measurements is $\mathcal{O}(K^6/\epsilon^2)$ and thus still poses a major problem. In fact, what is

more important is the R -ratio, which indicates the real improvement achieved by a given measurement reduction. This takes into account variances and covariances, which ultimately determines the overall measurement cost. Therefore, even if other measurement reduction schemes can further reduce the number of terms requiring separate measurement. If each term requires more measurements to determine the expectation value to a fixed precision then the resulting measurement cost can be the same or even worse. However, what makes the discussion slightly ambiguous is the state-dependence of the R -ratio. Absolute statements about the improvement in the number of measurements for different measurement reduction schemes are therefore hard to make in general. One advantage the unitary partitioning technique has is the R -ratio can never be below one and thus can never do worse than if no measurement reduction is applied. On the other hand, the possible improvement is also bounded (equation 4.66) and is related to the size of the anticommuting cliques, which have a maximum size of $2n + 1$. Due to this, we observe that the R -ratio will scale as $\mathcal{O}(\sqrt{n})$ and thus the fundamental issue of the number of measurements required by VQE making the runtime of VQE infeasible still remains when using unitary partitioning.

We note some interesting avenues for future work. (1) It would be interesting to apply the sorted insertion algorithm to partition a qubit Hamiltonian into anticommuting sets. It is likely this method will improve the possible measurement reduction obtained, which could be evaluated from an R score. (2) Given the coefficients of each Pauli operator can affect the measurement reduction obtained it would be interesting to explore methods that can increase the norm of such operators. For example in [240], Koridon and coworkers use orbital localization methods to change the 1-norm of a given molecular Hamiltonian. This could be used to improve the probability of success for the LCU method and the R -ratio. This idea can also be used with other measurement reduction methods. (3) It should be possible to perform commuting, qubit-wise commuting and anticommuting partitioning at once for a given qubit Hamiltonian. How to combine these techniques to achieve the best possible measurement reduction is an interesting question. (4) Given the classical shadow method requires measurement in random bases, the unitary rotations used by unitary partitioning to simultaneously measure an anticommuting clique could be used.

Finally, at a high level our results show that the deterministic sequence of rotations approach to unitary partitioning is better than the probabilistic LCU procedure. While it seems like the LCU construction is unlikely to be useful going forward, we will show in the next chapter that this formulation offers a major algorithmic improvement for the contextual-subspace VQE algorithm.

Chapter 5

The Contextual-Subspace

Variational Quantum Eigensolver

Present-day quantum computers have many constraints such as: lack of error correction, device topology (qubit connectivity) and low numbers of qubits. Consequently, this has given rise to a family of quantum-classical algorithms that leverage as much classical processing as possible to reduce the quantum resources required to solve the problem at hand. As discussed in Chapter 3 one of these is the variational quantum eigensolver (VQE) [94]. Other common examples include the Quantum Approximate Optimization Algorithm (QAOA) [241] and Variational Quantum Linear Solver (VQLS) [242]. A good example of how classical resources can alleviate some of the quantum resources required is the recently proposed entanglement forging method [114], where the electronic structure problem for H₂O was reduced from a 10-qubit problem to multiple 5 qubit problems that were each studied using conventional VQE and classically combined. Another recent novel approach is known as the Quantum-Classical hybrid Quantum Monte Carlo (QC-QMC) method, which was used to unbiased the sign problem in the projector Monte Carlo (PMC) method, which implements imaginary time evolution [243]. At a high level, the accuracy of a constrained PMC calculation is determined by the quality of trial wavefunctions. Quantum computers offer a way to efficiently store highly entangled trial wavefunctions and measure certain overlaps, that would require exponential resources classically. Huggins *et al.* performed QC-QMC simulations of different chemical systems on Google’s Sycamore processor and obtain results competitive with state-of-the-art classical methods [243].

The contextual subspace VQE (CS-VQE) algorithm is another hybrid quantum-classical approach [244]. It gives an approximate simulation method, where the quantum resources required can be varied for a trade-off in accuracy. This allows problems to be studied where the full Hamiltonian would normally be too large to investigate on current NISQ hardware. This was shown in the original CS-VQE paper, where chemical accuracy for various molecular systems was reached using significantly fewer qubits compared with the number required for VQE on the full system [244]. As CS-VQE reduces the number of qubits required for simulation, the number of terms in a Hamiltonian requiring separate measurements is also reduced.

A natural question that arises from this is whether further measurement reduction schemes can be utilized to reduce the overall measurement cost even more [103, 177–179, 185, 187, 188, 190, 193, 195, 196, 198, 201, 245]. The goal of this work was to investigate the possible reductions given by the unitary partitioning strategy [197, 198, 221] and whether chemical accuracy on larger molecules can be reached on currently available NISQ hardware.

In this chapter, we first summarise what contextuality is and where it can occur in qubit Hamiltonians. We then introduce the CS-VQE algorithm, where we introduce changes that improve the original implementation. We then apply the unitary partitioning measurement reduction technique along with CS-VQE on a test bed of different molecular structure Hamiltonians, where the contextual subspace approximation has been employed.

5.1 Contextuality

The foundation of quantum contextuality goes back to the Bell-Kochen-Specker (BKS) theorem [246]. In lay terms, every measurement provides a classical probability distribution (via the spectral theorem) and a joint distribution can be built as a product over all possible measurements [247]. The BKS theorem proves that it is impossible to reproduce the probabilities of every possible measurement outcome for a quantum system as marginals of this joint probability distribution [248]. This is related to how quantum mechanics does not allow models that are locally causal in a classical sense [249]. Contextuality is a generalization of nonlocality [249, 250]. This means that quantum measurement cannot be understood as simply revealing a pre-existing value of some underlying hidden variable [251, 252].

A good example of this phenomenon is the “Peres-Mermin square” [252, 253], where no state preparation is involved and only observables are considered. The example considers nine measurements arranged in a square. In this section, we follow the construction given in [248]. Each measurement has only two possible outcomes (dichotomic) $+1$ and -1 . In a realistic interpretation, performing each measurement on an object reveals whether the property is present ($+1$) or absent (-1), yielding nine properties.

We take three measurements along a column or row to form a “context” - a set of measurements whose values can be jointly measured i.e. the observables commute and thus share a common eigenbasis. Table 5.1 gives an example.

In a classical (noncontextual) model for this system, the nine measurements

$$\{IZ, ZI, ZZ, XI, IX, XX, XZ, ZX, YY\} \tag{5.1}$$

IZ	ZI	ZZ	r_0
XI	IX	XX	r_1
XZ	ZX	YY	r_2
c_0	c_1	c_2	

Table 5.1: Example Peres-Mermin square of nine possible observables for a physical system, where each measurement can be assigned a ± 1 value.

can be assigned a definite value independent of the context the measurement is obtained in. For example, if all measurements are assigned $+1$ in Table 5.1, then $c_0 = c_1 = c_2 = r_0 = r_1 = r_2 = +1$ and six positive products are obtained. If a single entry in Table 5.1 is changed it will affect two products (a row and column product). We consider the following Equation in this setting:

$$\begin{aligned}
 \langle PM \rangle &\equiv \langle IZ \cdot ZI \cdot ZZ \rangle + \langle XI \cdot IX \cdot XX \rangle + \\
 &\quad \langle XZ \cdot ZX \cdot YY \rangle + \langle IZ \cdot XI \cdot XZ \rangle + \\
 &\quad \langle ZI \cdot IX \cdot ZX \rangle - \langle ZZ \cdot XX \cdot YY \rangle \\
 &= r_0 + r_1 + r_2 + c_0 + c_1 - c_2
 \end{aligned} \tag{5.2}$$

and find that classically we get an inequality $\langle PM \rangle \leq 4$. We reiterate that this is the setting of eight $+1$ assignments and a single -1 assignment. This inequality is saturated when the -1 value is assigned to one of the observables in the last column of Table 5.1.

The significance of this inequality is that it can be violated by quantum systems. Thinking of this in a quantum setting, the operators in rows and columns of Table 5.1 commute. If we multiply along the rows and columns we get $+II$ apart from the last column where $c_2 = -II$ (see Table 5.2). This is the case regardless of what quantum state is considered. Using the expectation values of the product of these operators in Equation 5.2, we find $\langle PM \rangle = 6$, violating the classical bound.

Classically Equation 5.2 is bounded as $\langle PM \rangle \leq 4$ due to the assumption that the nine observables of the object can be assigned a value consistently. Violation of this bound implies that either the value assignment must depend on which context (row or column) the observable appears in or there is no value assignment. This phenomenon is known as quantum contextuality [248].

IZ	ZI	ZZ	$\langle +II \rangle = +1$
XI	IX	XX	$\langle +II \rangle = +1$
XZ	ZX	YY	$\langle +II \rangle = +1$
$\langle +II \rangle = +1$	$\langle +II \rangle = +1$	$\langle -II \rangle = -1$	

Table 5.2: Example Peres-Mermin square of nine Hermitian operators, all with ± 1 eigenvalues - representing observables.

In VQE, a Hamiltonian is defined by a linear combination of Pauli operators. The expectation value is obtained by measuring each Pauli operator in a separate experiment and combining the results. Different groups of commuting operators form contexts. In general there will be incompatible contexts where it is impossible to consistently assign joint outcomes. In other words, different inference relations will lead to contradictions. Outcomes assigned to individual measurements are therefore context-dependent and the problem is contextual. If not, then the problem is non-contextual and a noncontextual (classical) hidden variable model can be used to solve such systems. Colloquially, for a noncontextual problem it is possible to assign deterministic outcomes to observables simultaneously without contradiction; however, for a contextual problem this is not possible [244].

The contextual subspace VQE algorithm uses the idea of contextuality to motivate different approximations decreasing the quantum resources required to study a given problem. The next Section introduces this algorithm in full.

5.2 Contextual-subspace VQE

The contextual-Subspace VQE algorithm is based on partitioning any qubit Hamiltonian:

$$\begin{aligned} H_{\text{full}} &= \sum_a c_a P_a = \sum_a c_a \left(\bigotimes_{j=0}^{n-1} \sigma_j^{(a)} \right) \\ &= \sum_a c_a (\sigma_0^{(a)} \otimes \sigma_1^{(a)} \otimes \dots \otimes \sigma_{n-1}^{(a)}), \end{aligned} \tag{5.3}$$

into two disjoint components. One is noncontextual and the other is contextual, explicitly [244]:

$$H_{\text{full}} = H_{\text{con}} + H_{\text{noncon}}. \tag{5.4}$$

Here c_a are real coefficients. Each Pauli operator P_a is made up of an n -fold tensor product of single qubit Pauli matrices $\sigma_j \in \{\mathcal{I}, X, Y, Z\}$, where j indexes the qubit the operator acts on.

At a high level, CS-VQE starts by solving the noncontextual problem H_{noncon} and then solves the contextual part H_{con} constrained by the solution of H_{noncon} . We will see later that this constraint manifests from a subspace of states defined by the noncontextual solution. The method projects the contextual part of this problem into a subspace consistent with these states and so any new solution is compatible with the noncontextual result. We note that this projection in general will result in a loss of information and thus leads to approximations. However, as there is flexibility in how the projection is done, the technique allows different levels of approximations from a

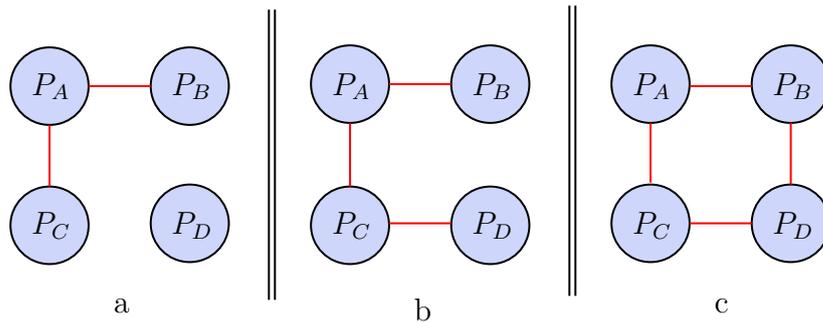


Figure 5.1: Compatibility graphs for contextual sets of four Pauli operators [254]. Edges are between Pauli operators that commute.

completely noncontextual approximation to no approximation (standard VQE). We will see that this allows problems to be significantly simplified, while still maintaining enough accuracy to be meaningful. The subsequent subsections introduce each part of the CS-VQE algorithm starting with how the Hamiltonian is partitioned.

5.2.1 Partitioning the Hamiltonian

Let $\mathcal{S}^{H_{\text{full}}}$ be the set of Pauli operators, in the full system Hamiltonian, requiring measurement in a VQE experiment. It was shown in [206, 254], that a set of four Pauli operators $\{P_A, P_B, P_C, P_D\}$ is strongly contextual if any of the compatibility graphs shown in Figure 5.1 are present. Alternatively, for a general set of Pauli operators, once globally commuting operators are removed if commutation is transitive then the set is noncontextual and if commutation is not transitive then the set is contextual.

Given an arbitrary set of Pauli operators \mathcal{P} , this gives an algorithm to check for contextuality [254]. We focus on the graph illustrated in figure 5.1a. To check if \mathcal{P} is contextual, all we need to do is check certain commutativity relations between operators in the set \mathcal{P} . A pseudo algorithm is given in algorithm 1. First, an $\mathcal{O}(|\mathcal{P}|^2)$ routine is used to remove operators in \mathcal{P} that commute with all operators. These terms represent symmetries of the problem and are put into the set \mathcal{Z} . The leftover terms are put into the set \mathcal{T} , mathematically we write this as: $\mathcal{T} = \mathcal{P} \setminus \mathcal{Z}$. The remaining procedure takes $\mathcal{O}(|\mathcal{T}|^3)$ steps to determine whether \mathcal{P} is contextual. This goes through unique triples of Pauli operators in \mathcal{T} and checks for the structure in Figure 5.1a (where P_D need not be considered). If any triple is found where $[P_A, P_B] = 0$, $[P_A, P_C] = 0$ and $\{P_A, P_C\} \neq 0$ the problem is contextual. This check for contextuality is implemented in OpenFermion [224].

We note two edge cases. All one qubit problems are noncontextual and any Hamiltonian H with $|\mathcal{T}| \leq 3$ must be noncontextual.

To begin CS-VQE, we first need to define the contextual and noncontextual parts. The task of finding the largest noncontextual subset of Pauli operators in $\mathcal{S}^{H_{\text{full}}}$ is a

Algorithm 1 Test for strong contextuality in a given set of Pauli operators [254]

Input: $\mathcal{P} = \{P_0, P_1, P_2, \dots\}$ \triangleright Input \mathcal{P} is a set of Pauli operators.**Output:** *contextual* (*True/False*) \triangleright Whether the set \mathcal{P} is strongly contextual. $\mathcal{Z} \leftarrow \{\}$ $\mathcal{T} \leftarrow \{\}$ *contextual* \leftarrow *False***for** $i = 0$ to $|\mathcal{P}| - 1$ **do** **if** $[P_i, P_j] = 0 \forall j \neq i$ where $j = 0$ to $|\mathcal{P}| - 1$ **then** $\mathcal{Z} \leftarrow \mathcal{Z} \cup \{P_i\}$ **else** $\mathcal{T} \leftarrow \mathcal{T} \cup \{P_i\}$ **end if****end for****for** $i = 0$ to $|\mathcal{T}| - 3$ **do** **for** $j = i + 1$ to $|\mathcal{T}| - 2$ **do** **for** $k = j + 1$ to $|\mathcal{T}| - 1$ **do** **if** $[P_i, P_j] = 0, [P_i, P_k] = 0$ and $\{P_j, P_k\} = 0$ **then** *contextual* \leftarrow *True* **return** *contextual* **else** **continue** **end if** **end for** **end for****end for****return** *contextual*

generalization of the disjoint clique problem [255, 256], which is NP-complete. However, different heuristics can be used to approximately solve this problem. To date, VQE experiments have mainly focused on chemistry Hamiltonians, where Hartree-Fock accounts for most of the energy. Such Hamiltonians contain Pauli operators that l_1 norms are dominated by diagonal terms - i.e. Pauli operators made up of tensor products of single qubit I and Pauli Z matrices. To find a noncontextual set in such a scenario a greedy heuristic can be used. Here, we enumerate through the Pauli operators of a given Hamiltonian and add operators to a set so long as the set remains noncontextual via Algorithm 1. This gives a noncontextual set containing mainly diagonal terms, with some additional operators [256]. Alternative procedures to find the largest noncontextual subsets remain an open question for the CS-VQE

algorithm.

5.2.2 Solving the noncontextual Hamiltonian

Once the noncontextual Hamiltonian H_{noncon} is determined, we can define the set $\mathcal{S}^{H_{\text{noncon}}}$ to be the Pauli operators in H_{noncon} . We split $\mathcal{S}^{H_{\text{noncon}}}$ into two subsets \mathcal{Z} and \mathcal{T} - representing the set of universally commuting Pauli operators \mathcal{Z} and their complement respectively [244, 256]:

$$\mathcal{S}^{H_{\text{noncon}}} = \mathcal{Z} \cup \mathcal{T} = \left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{Z}}}^{|\mathcal{Z}|-1} P_i \right\} \cup \left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{T}}}^{|\mathcal{T}|-1} P_i \right\}. \quad (5.5)$$

Slight modifications to Algorithm 1 achieve this - where \mathcal{P} should be set to $\mathcal{S}^{H_{\text{noncon}}}$ and both \mathcal{Z}, \mathcal{T} should be returned.

The operators in \mathcal{Z} are noncontextual, as by definition they are universally commuting and represent symmetries of H_{noncon} . For the overall superset $\mathcal{S}^{H_{\text{noncon}}}$ to be noncontextual, the remaining operators in \mathcal{T} must be made up of N disjoint cliques C_j [256], where operators within a clique all pairwise commute and operators between cliques pairwise anticommute. This is because commutation forms an equivalence relation on \mathcal{T} if and only if $\mathcal{S}^{H_{\text{noncon}}}$ is noncontextual [206, 244, 256]. We can write \mathcal{T} as:

$$\mathcal{T} = \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{T}}}^{|\mathcal{T}|-1} P_i = \bigcup_{j=0}^{N-1} C_j = \bigcup_{j=0}^{N-1} \left(\bigcup_{\substack{k=0 \\ \text{where} \\ [P_k, P_l]=0 \\ \forall P_k, P_l \in C_j}}^{|C_j|-1} P_k^{(j)} \right). \quad (5.6)$$

We re-define each clique C_j using the identity operation defined by the first operator of the j th clique, $P_0^{(j)} P_0^{(j)} = \mathcal{I}$, which represents the first operator in each of the N cliques. We write the j th clique as [256]:

$$C_j = \bigcup_{\forall P_k \in C_j} P_k^{(j)} = \bigcup_{\forall P_k \in C_j} P_k^{(j)} \underbrace{P_0^{(j)} P_0^{(j)}}_{\mathcal{I}} = \bigcup_{\forall P_k \in C_j} \left(P_k^{(j)} P_0^{(j)} \right) P_0^{(j)} = \bigcup_{k=0}^{|C_j|-1} A_k^{(j)} P_0^{(j)}. \quad (5.7)$$

The new operators $A_k^{(j)} = P_k^{(j)} P_0^{(j)}$ are just Pauli operators up to a complex phase. The new operators $A_k^{(j)}$ must still commute with the universally commuting operators in \mathcal{Z} , but now must also commute with all the other terms in the $N - 1$ cliques C_j [256]. Using this, the noncontextual set (Equation 5.5) can be rewritten as:

$$\mathcal{S}^{H_{\text{noncon}}} = \left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{Z}}}^{|\mathcal{Z}|-1} P_i \right\} \cup \left\{ \bigcup_{j=0}^{N-1} C_j \right\} \quad (5.8a)$$

$$= \underbrace{\left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{Z}}}^{|\mathcal{Z}|-1} P_i \right\}}_{\mathcal{Z}} \cup \underbrace{\left\{ \bigcup_{j=0}^{N-1} \left(\bigcup_{\substack{k=0 \\ \text{where} \\ [P_k, P_l]=0 \\ \forall P_k, P_l \in C_j}}^{|C_j|-1} A_k^{(j)} P_0^{(j)} \right) \right\}}_{\mathcal{T}}. \quad (5.8b)$$

So far we have considered the noncontextual set of Pauli operators $\mathcal{S}^{H_{\text{noncon}}}$, which in general will be a dependent set. By this we mean that some operators in the set can be written as a product of other commuting operators in the set. We need to reduce this set $\mathcal{S}^{H_{\text{noncon}}}$ to an independent set of Pauli operators, where all operators in the noncontextual Hamiltonian can be inferred from the values of other operators in the set under the Jordan product.

To obtain an independent set from $\mathcal{S}^{H_{\text{noncon}}}$, we first take the completely commuting Pauli operators:

$$\begin{aligned} \mathcal{G}' &\equiv \mathcal{Z} \cup \left\{ \bigcup_{j=0}^{N-1} \{A_k^{(j)} | k = 1, 2, \dots, |C_j| - 1\} \right\} \\ &\equiv \left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{Z}}}^{|\mathcal{Z}|-1} P_i \right\} \cup \left\{ \bigcup_{j=0}^{N-1} \{A_k^{(j)} | k = 1, 2, \dots, |C_j| - 1\} \right\}, \end{aligned} \quad (5.9)$$

and using the procedure in [124] find an independent subset \mathcal{G} :

$$\mathcal{G} \equiv \{P_i | i = 0, 1, \dots, |\mathcal{G}| - 1\}. \quad (5.10)$$

The set \mathcal{G} represents a generating set for the symmetry operators of H_{noncon} . In other words, different products of operators in \mathcal{G} can be used to generate any Pauli operator that by definition must commute with H_{noncon} .

Finally, we need to consider the N pairwise anticommuting $P_0^{(j)}$ operators defined by the N anticommuting cliques. As the operators in \mathcal{G} universally commute with all operators in the noncontextual Hamiltonian, each operator (clique representative) in the set $\{P_0^{(j)} | j = 0, 1, \dots, (N - 1)\}$ must be independent [256]. We combine the N anticommuting operators into the observable [256]:

$$A(\vec{r}) = \sum_{j=0}^{N-1} r_j P_0^{(j)}. \quad (5.11)$$

We denote the set of Pauli operators making up this operator as $\mathcal{A} \equiv \{P_0^{(j)} | j = 0, 1, \dots, N-1\}$.

Next, we combine the \mathcal{A} and \mathcal{G} into an independent set \mathcal{R} [256]:

$$\begin{aligned} \mathcal{R} &\equiv \mathcal{A} \cup \mathcal{G} \\ &\equiv \{P_0^{(j)} | j = 0, 1, \dots, N-1\} \cup \mathcal{G}. \end{aligned} \tag{5.12}$$

Inspecting the properties of \mathcal{R} , one can bound its size. The set \mathcal{G} has size at most $n-1$, as n independent commuting Pauli operators form a complete commuting set of observables for n qubits. In other words, as \mathcal{G} is a universally commuting set if its size was n (or more) then taking \mathcal{G} and one operator $P_0^{(j)}$ (the set $\mathcal{G} \cup \{P_0^{(j)}\}$) is also a fully commuting set - and would be a commuting set of size $n+1$ (or more) [256]. The maximum number of independent anticommuting operators on n qubits was shown in [206] to be $2n+1$. This actually bounds the size of \mathcal{R} , which occurs when the set $\mathcal{G} = \{\}$ (and thus \mathcal{Z} is empty) and $|\mathcal{A}| = 2n+1$ [256].

Looking at the noncontextual set of Pauli operators in equation 5.8, making up H_{noncon} , we see that the subset $\mathcal{G} \subseteq \mathcal{R}$ (Equation 5.12) includes all the generators for the terms in \mathcal{Z} and each Pauli $A_k^{(j)}$ operator. Any operator in \mathcal{Z} and each Pauli $A_k^{(j)}$ operator can therefore be found by a finite combination of operators in \mathcal{G} . Each operator in \mathcal{T} can also be generated by a combination of one $P_0^{(j)}$ operator and some combination of operators in \mathcal{G} . Again \mathcal{R} (Equation 5.12) contains all the operators required. To summarise, the set \mathcal{R} contains all the required terms to reproduce the expectation value of any operator in $\mathcal{S}^{H_{\text{noncon}}}$ under the Jordan product. The Jordan product is defined as: $P_a \circ P_b = \frac{\{P_a, P_b\}}{2}$ and is equal to the regular matrix product if the operators commute, and equal to zero if the operators anticommute. This ensures a product of anticommuting Pauli operators cannot be taken, as a simultaneous value assignment of anticommuting Pauli operators is not possible.

Next, as all terms in H_{noncon} can be simultaneously assigned a definite value without contradiction, a phase space description of its eigenspace can be used [244, 256, 257]. The phase space points are the possible joint value assignments to a set of observables in \mathcal{R} . The eigenstates of H_{noncon} are probability distributions over this phase space [244]. This is a quasi-quantized model which is a classical phase-space model, with an uncertainty relation imposed upon the allowed probability distributions (sometimes called epistemic states) on the phase space [244, 257, 258].

The states for this quasi-quantized model we denote as “noncontextual” states. These noncontextual states are probability distributions over the phase space points i.e. the eigenvalue assignments of each Pauli operator in \mathcal{R} . Probability distributions corresponding to valid quantum states must obey an uncertainty relation [257, 258]. Kirby *et al.* in [256], show that a sufficient condition is that the symmetry generators $G_i \in \mathcal{G}$ take definite ± 1 values and the expectation value of each clique representative

$\langle P_0^{(j)} \rangle$ form a unit vector [256]. If we let each noncontextual state be defined by the parameters (\vec{q}, \vec{r}) then:

$$\langle G_i \rangle = q_i = \pm 1, \quad (5.13a)$$

$$\langle P_0^{(j)} \rangle = r_j, \quad (5.13b)$$

where:

$$\begin{aligned} (\vec{q}, \vec{r}) &= (q_0, q_1, \dots, q_{|\mathcal{G}|-1}, r_0, r_1, \dots, r_{|\mathcal{A}|-1}), \\ \text{and } \left(\sum_{i=0}^{|\mathcal{A}|-1} |r_i|^2 \right)^{1/2} &= 1. \end{aligned} \quad (5.14)$$

The joint probability distribution (classical) associated with the set of expectation values given in equation 5.13 is supplied in [256]. A discussion on this is outside the scope of this thesis.

In summary, with respect to the phase-space model [256] a valid noncontextual state (\vec{q}, \vec{r}) sets the expectation value of the operators in \mathcal{R} (Equation 5.12). The expectation value of all the operators in $\mathcal{S}^{H_{\text{noncon}}}$ are generated from some finite combination of terms in \mathcal{R} under the Jordan product. Explicitly, let $P_i^{\mathcal{Z}} \in \mathcal{Z} \subseteq \mathcal{S}^{H_{\text{noncon}}}$ then if we let $\mathcal{J}_{P_i^{\mathcal{Z}}}^{\mathcal{G}}$ be the set of indices such that $P_i^{\mathcal{Z}} = \prod_{i \in \mathcal{J}_{P_i^{\mathcal{Z}}}^{\mathcal{G}}} G_i$; then [256]:

$$\langle P_i^{\mathcal{Z}} \rangle = \prod_{i \in \mathcal{J}_{P_i^{\mathcal{Z}}}^{\mathcal{G}}} \langle G_i \rangle = \prod_{i \in \mathcal{J}_{P_i^{\mathcal{Z}}}^{\mathcal{G}}} q_i. \quad (5.15)$$

In words, we combine the expectation value of some finite set of Pauli operators in the independent set \mathcal{G} - given by $\mathcal{J}_{P_i^{\mathcal{Z}}}^{\mathcal{G}}$ - to reproduce the expectation value for $\langle P_i^{\mathcal{Z}} \rangle$.

Similarly, the expectation value for each $A_k^{(j)} P_0^{(j)} \in \mathcal{T} \subseteq \mathcal{S}^{H_{\text{noncon}}}$ (Equation 5.8) term is given by:

$$\langle A_i^{(j), \in \mathcal{G}} P_0^{(j), \in \mathcal{T}} \rangle = \left(\prod_{i \in \mathcal{J}_{A_i^{(j)}}^{\mathcal{G}}} \langle G_i \rangle \right) r_j = \left(\prod_{i \in \mathcal{J}_{A_i^{(j)}}^{\mathcal{G}}} q_i \right) r_j, \quad (5.16)$$

where $\mathcal{J}_{A_i^{(j)}}^{\mathcal{G}}$ are the set of indices such that $\langle A_i^{(j)} \rangle = \prod_{i \in \mathcal{J}_{A_i^{(j)}}^{\mathcal{G}}} \langle G_i \rangle$ and $\langle P_0^{(j)} \rangle = r_j$ [256].

We can write the noncontextual Hamiltonian as:

$$H_{\text{noncon}} = \left(\sum_{i=0}^{|\mathcal{Z}|-1} c_i P_i^{\mathcal{Z}} \right) + \sum_{j=0}^{N-1} \left[\sum_{k=0}^{|\mathcal{C}_j|-1} a_k A_k^{(j)} P_0^{(j)} \right] \quad (5.17)$$

and find the energy of a particular noncontextual state (\vec{q}, \vec{r}) :

$$\langle H_{\text{noncon}} \rangle = E_{\text{noncon}}(\vec{q}, \vec{r}) = \left(\sum_{i=0}^{|\mathcal{Z}|-1} \beta_i \langle P_i^{\mathcal{Z}} \rangle \right) + \sum_{j=0}^{N-1} \left[\sum_{k=0}^{|\mathcal{C}_j|-1} \beta_k \langle A_k^{(j)} P_0^{(j)} \rangle \right], \quad (5.18)$$

where β_i and β_k are real coefficients and each expectation value is given by equation 5.15 and 5.16 [256].

To find the ground state of H_{noncon} , we minimize Equation 5.18 via a brute-force search as described in [256]. Algorithm 2 summarises the steps. First, a trial \vec{q} is defined. This is a set of ± 1 expectation values for each G_j . An initial guess of the amplitudes r_i of the unit vector \vec{r} is made and the energy (Equation 5.18) is minimized over this continuous parameterization of \vec{r} for a fixed trial \vec{q} , until the energy converges to a minimum. These steps are repeated for all the $2^{|\mathcal{G}|}$ assignments of \vec{q} . The (\vec{q}, \vec{r}) combination that gives the lowest overall energy represents the noncontextual ground state of the physical system. We denote this parameterization as (\vec{q}_0, \vec{r}_0) . Note for a fixed \vec{q} , we optimize over \vec{r} . This can be thought of as optimizing a function defined on a hypersphere. Currently, we haven't explored the properties of this function.

It remains an open question for the CS-VQE algorithm if alternate optimization strategies are possible, for example using chemical intuition during optimization. This brute force approach of searching over all $2^{|\mathcal{G}|}$ possibilities for \vec{q} may not be necessary. In the next section, we discuss how to map the contextual problem into a subspace consistent with a defined noncontextual state (\vec{q}, \vec{r}) .

5.2.3 Noncontextual stabilizers

Thus far, we have solved the classical part of CS-VQE. Now we look at solving the contextual component that has so far been neglected. At a high level this section involves solving H_{con} while keeping the solution consistent with H_{noncon} . We use the stabilizer framework to achieve this, where H_{con} can be constrained by a set of stabilizers based on $\mathcal{R} \equiv \mathcal{A} \cup \mathcal{G}$ (equation 5.12) and the noncontextual state.

To start this process, we first need to consider the observable $A(\vec{r})$ (equation 5.11). For a given noncontextual state (\vec{q}, \vec{r}) we find the expectation value of $A(\vec{r})$ is fixed as:

$$\langle A(\vec{r}) \rangle = \sum_{j=0}^{N-1} r_j \langle P_0^{(j)} \rangle = \sum_{j=0}^{N-1} r_j r_j = \sum_{j=0}^{N-1} |r_j|^2 = +1. \quad (5.19)$$

Using this result, a noncontextual state (\vec{q}, \vec{r}) is equivalent to the joint expectation value assignment of $\langle G_i \rangle = q_i = \pm 1$ and $\langle A(\vec{r}) \rangle = +1$. This defines a set of “stabilizer-

Algorithm 2 Brute force method to solve noncontextual problem

$\mathcal{Q} \leftarrow \{q_0, q_1, \dots, q_{|\mathcal{G}|-1}\}^{2^{|\mathcal{G}|}}$ \triangleright Set \mathcal{Q} contains all possible \vec{q} vectors, where $q_i \in \{+1, -1\}$.

$\vec{q}_0 \leftarrow \{\}$
 $\vec{r}_0 \leftarrow \{\}$
 $E_{\text{noncon}}^0 \leftarrow 0$

for \vec{q}_{test} in \mathcal{Q} **do**

$\vec{r}_{\text{opt}}, E_{\text{noncon}}^{\text{opt}} \leftarrow \underset{\vec{r}}{\text{argmin}}[E_{\text{noncon}}(\vec{q}_{\text{test}}, \vec{r})]$ \triangleright for a given \vec{q}_{test} , minimize the energy

(Equation 5.18)with respect to \vec{r} .

if $E_{\text{noncon}}^{\text{opt}} < E_{\text{noncon}}^0$ **then**

$\vec{q}_0 \leftarrow \vec{q}_{\text{test}}$

$\vec{r}_0 \leftarrow \vec{r}_{\text{opt}}$

$E_{\text{noncon}}^0 \leftarrow E_{\text{noncon}}^{\text{opt}}$

else

continue

end if

end for

return $\vec{q}_0, \vec{r}_0, E_{\text{noncon}}^0$

like” operators:

$$\mathcal{W}_{\text{all}} \equiv \underbrace{\{q_i Q_i | i = 0, 1, \dots, |\mathcal{G}| - 1\}}_{G_i \in \mathcal{G} \text{ and } q_i \in \vec{q}} \cup A(\vec{r}), \quad (5.20)$$

which by definition must stabilize that noncontextual state (\vec{q}, \vec{r}) or more precisely, the subspace of quantum states corresponding to it. The reason we call these operators “stabilizer-like” is due to $A(\vec{r})$ not being a stabilizer, as it is not an element of the Pauli group, but is unitary equivalent to one. This stems from unitary partitioning, where a linear combination of normalized anticommuting Pauli operators can be rotated onto a single Pauli via a unitary operator. Section 4.3 in the previous chapter covers this process in full. We can consider equation 5.20 under this unitary transformation, where \mathcal{W}_{all} becomes:

$$\begin{aligned} \mathcal{W}_{\text{all}} \mapsto \mathcal{W}'_{\text{all}} &\equiv \{q_i R Q_i R^\dagger | i = 0, 1, \dots, |\mathcal{G}| - 1\} \cup R A(\vec{r}) R^\dagger \\ &\equiv \{q_0 G_0, q_1 G_1, \dots, q_{|\mathcal{G}|-1} G_{|\mathcal{G}|-1}, \xi P_0^{(k)}\}. \end{aligned} \quad (5.21)$$

This defines a regular set of stabilizers for the noncontextual state (\vec{q}, \vec{r}) that defines a subspace of quantum states. Here $\xi = \pm 1$ and is determined by $R A(\vec{r}) R^\dagger = \xi P_0^{(k)}$

and can always be chosen to be +1, which we do throughout this chapter. Here R can be constructed as a LCU R^{LCU} (Section 4.3.2.2) or a sequence of rotations R_S (Section 4.3.1) and $P_0^{(k)} \in \mathcal{A}$. As the symmetry generators universally commute with the operators R (due to R being composed of Pauli operators in \mathcal{A}), we find: $RG_iR^\dagger = RR^\dagger G_i = G_i \forall G_i \in \mathcal{G}$.

5.2.4 Mapping to contextual subspace

In CS-VQE, the expectation value of the full Hamiltonian (equation 5.4) is obtained by first solving the noncontextual problem yielding a noncontextual state (\vec{q}, \vec{r}) - normally the ground state (\vec{q}_0, \vec{r}_0) . Next, the contextual Hamiltonian is projected into the subspace of allowed quantum states consistent with the defined noncontextual state. This constraint is imposed via: $H_{\text{full}} \mapsto H'_{\text{full}} = Q_{\mathcal{W}}^\dagger U_{\mathcal{W}}^\dagger H_{\text{full}} U_{\mathcal{W}} Q_{\mathcal{W}}$, where the expectation value is then found on a quantum device. The ground state of H'_{full} is then found, which is an approximate ground state of H_{full} . In this section, we define what $U_{\mathcal{W}}$ (a unitary) and $Q_{\mathcal{W}}$ (a projector) are.

The unitary operation $U_{\mathcal{W}}$ is defined by the set of contextual stabilizers $\mathcal{W} \subseteq \mathcal{W}_{\text{all}}$ (equation 5.20) that eigenvalue we fix according to the noncontextual state. If $A(\vec{r}) \in \mathcal{W}$, meaning $\langle A(\vec{r}) \rangle$ is fixed to be +1, then the steps summarised in Equation 5.21 must first be performed to reduce $A(\vec{r})$ to a single Pauli operator. Clifford operators $V_i(P)$ are then used to map each $P \in \mathcal{W}$ to a single-qubit Z operator. Each V_i is made up of at most two $\frac{\pi}{2}$ Clifford rotations, generated by Pauli operators, per element in \mathcal{W} . In [244] it was shown that at most there will be $2n$ of these rotations, where n is the number of qubits the problem is defined on [86]. We provide details on this operation in Appendix A.1. We can write the overall operator as:

$$U_{\mathcal{W}}^\dagger(\vec{q}, \vec{r}) = \begin{cases} \prod_{P_i \in \mathcal{W} \subseteq \mathcal{W}_{\text{all}}} V_i(P_i), & \text{if } A(\vec{r}) \notin \mathcal{W} \\ \left(\prod_{P_i \in \mathcal{W} \subseteq \mathcal{W}_{\text{all}}} V_i(P_i) \right) R, & \text{if } A(\vec{r}) \in \mathcal{W} \end{cases}. \quad (5.22)$$

Applying $U_{\mathcal{W}}^\dagger \mathcal{W} U_{\mathcal{W}} = \mathcal{W}^Z$ results in a set of single-qubit Z Pauli operators. An implementation note is that each operator $V_i(P_i)$ in $U_{\mathcal{W}}$ depends on each other - this can be seen by expanding $U_{\mathcal{W}}^\dagger \mathcal{W} U_{\mathcal{W}}$. Therefore, each V_i operator is dependent on the stabilizers in \mathcal{W} and in what order they occur. We recursively define each V_i as follows:

1. Set $\mathcal{W} = R\mathcal{W}R^\dagger$ iff $A(\vec{r}) \in \mathcal{W}$.
2. Find the unitary V_0 mapping the first Pauli operator $P_0 \in \mathcal{W}$ to a single qubit Pauli operator.
3. Apply this operator to each operator in the set: $V_0 \mathcal{W} V_0^\dagger = \mathcal{W}^{(0)}$.

4. Find the unitary V_1 mapping $V_0 P_1 V_0^\dagger \in \mathcal{W}^{(0)}$ to a single-qubit Z Pauli operator.
5. Apply this operator to all operators in the set: $V_1 \mathcal{W}^{(0)} V_1^\dagger = \mathcal{W}^{(1)}$
6. Repeat this procedure from step (3) until all the operators are mapped to single qubit Z Pauli operators: $\mathcal{W} \mapsto \mathcal{W}^Z$.

Finally, the eigenvalue of the single-qubit Z Pauli stabilizers in \mathcal{W}^Z are defined by the vector \vec{q} of the noncontextual ground state (\vec{q}, \vec{r}) , note $\langle A(\vec{r}) \rangle$ is fixed to $+1$ and thus \vec{r} isn't important here. $U_{\mathcal{W}}$ can flip the sign of these assignments, but it is efficient to classically determine by tracking how $U_{\mathcal{W}}$ affects the sign of the operators in \mathcal{W} .

To project the Hamiltonian into the subspace consistent with the noncontextual state, we first perform the following rotation $H_{\text{full}} \mapsto H_{\text{full}}^{\mathcal{W}} = U_{\mathcal{W}}^\dagger H_{\text{full}} U_{\mathcal{W}}$. As this is a unitary transform, the resultant operator has the same spectrum as before. We then restrict the rotated Hamiltonian to the correct subspace by enforcing the eigenvalue of the operators in \mathcal{W}^Z - where the outcomes are defined by the noncontextual state. As each operator in \mathcal{W}^Z only acts nontrivially on a unique qubit, each stabilizer fixes the state of that qubit to be either $|0\rangle$ or $|1\rangle$. We write this state as:

$$|\psi_{\text{fixed}}\rangle = \bigotimes_{P_v \in \mathcal{W}^Z} |i\rangle_v \begin{cases} i = 0, & \text{if } \langle P_v \rangle = +1 \\ i = 1, & \text{if } \langle P_v \rangle = -1 \end{cases}, \quad (5.23)$$

where v indexes the qubit a given single-qubit stabilizer acts on and $\langle P_v \rangle$ is defined by the noncontextual state. We can write the projector onto this state as:

$$Q_{\mathcal{W}} = |\psi_{\text{fixed}}\rangle \langle \psi_{\text{fixed}}| \otimes \mathcal{I}_{(n-|\mathcal{W}^Z|)}, \quad (5.24)$$

where $\mathcal{I}_{(n-|\mathcal{W}^Z|)}$ is the identity operator acting on the $(n - |\mathcal{W}^Z|)$ qubits not fixed by the single-qubit P_v stabilizers. The action on a general state $|\phi\rangle$ is:

$$Q_{\mathcal{W}} |\phi\rangle = |\psi_{\text{fixed}}\rangle \langle \psi_{\text{fixed}}| \phi\rangle \otimes |\phi\rangle_{(n-|\mathcal{W}^Z|)}, \quad (5.25)$$

where $Q_{\mathcal{W}}$ has only fixed the state of qubits v and thus each stabilizer P_v removes one qubit from the problem. As the states of these qubits are fixed, the expectation value of the single-qubit Pauli matrices indexed on qubits v are known. Thus the Pauli operators in the rotated Hamiltonian $H_{\text{full}}^{\mathcal{W}} = U_{\mathcal{W}}^\dagger H_{\text{full}} U_{\mathcal{W}}$ acting on these qubits can be updated accordingly and the Pauli matrices on qubits v dropped. Any term in the rotated Hamiltonian that anticommutes with a fixed generator P_v is forced to have an expectation value of zero and can be completely removed from the problem Hamiltonian. The resultant Hamiltonian acts on $|\mathcal{W}^Z|$ fewer qubits. We denote this operation as $H_{\text{full}} \mapsto H_{\text{full}}^{\mathcal{W}'} = Q_{\mathcal{W}}^\dagger U_{\mathcal{W}}^\dagger H_{\text{full}} U_{\mathcal{W}} Q_{\mathcal{W}}$. The noncontextual approximation

will be stored in the identity term of the problem and therefore doesn't need to be tracked separately.

The choice of which stabilizer eigenvalues to fix (i.e. what is included in \mathcal{W}) and which to allow to vary remains an open question of the CS-VQE algorithm. The number of possible stabilizer combinations will be $\sum_{i=1}^{|\mathcal{W}_{\text{all}}|} \binom{|\mathcal{W}_{\text{all}}|}{i} = 2^{|\mathcal{W}_{\text{all}}|} - 1$. Rather than searching over all $2^{|\mathcal{W}_{\text{all}}|} - 1$ combinations of stabilizers to fix, in this work we use the heuristic given in [244]. This begins at the full noncontextual approximation, where \mathcal{W} contains all possible stabilizers. We then add a qubit to the quantum correction, by removing an operator from \mathcal{W} and greedily choosing each pair that gives the lowest ground state energy estimate [244]. Alternative strategies on how to do this remains an open question for CS-VQE. A possible way to approach this problem is to look at the priority of different terms in H_{con} [259]. Note that the quality of the approximation is sensitive to which stabilizers are included in \mathcal{W} . When fewer stabilizers are considered (included in \mathcal{W}), the resultant rotated Hamiltonians will act on more qubits and approximate the true ground state energy better.

Kirby *et al.* in [244] construct R as a sequence of rotations (exponentiated Pauli operators) defined by $A(\vec{r})$ as defined by the unitary partitioning method [197, 198]. We denote this operation R_S . Section 4.3.1 gives the full definition of this operator. If R_S is considered as just an arbitrary sequence of rotations generated by exponentiated Pauli operators, then the transformation $H_{\text{full}} \mapsto H_{\text{full}}^{R_S} = R_S H_{\text{full}} R_S^\dagger$ can result in an operator $H_{\text{full}}^{R_S}$ composed of more Pauli operators. In fact the number of terms increases by a factor of $\mathcal{O}(2^N)$, where N is the number of cliques defined from \mathcal{T} [244]. This presented a possible roadblock for the CS-VQE algorithm, as classically precomputing $U_{\mathcal{W}}^\dagger H_{\text{full}} U_{\mathcal{W}}$ could cause the number of terms to exponentially increase. We give a further analysis of this operation in Appendix A.2 and show that additional structure between R_S and H_{full} can make the base of the exponent slightly lower; however, the scaling still remains exponential in the number of qubits n . The only case in which there is not an exponential increase in terms is for the trivial instance that R_S commutes with H_{full} . In the next section, we provide an alternative construction of R via a linear combination of unitaries that results in only a quadratic increase in the number of terms of the Hamiltonian when transformed. This avoids the need to apply the unitary partitioning operator R (via a sequence of rotations) coherently in the quantum circuit after the ansatz circuit, which was proposed in [244].

5.2.5 Linear combination of unitaries construction of R

In the unitary partitioning method [197, 198], it was shown that R could also be built as a linear combination of Pauli operators [198, 221]. We denote the operator as R_{LCU} (equation 4.91). Rotating a general Hamiltonian H_{full} by this operation R_{LCU} results

in:

$$\begin{aligned}
 R_{LCU} H_{\text{full}} R_{LCU}^\dagger &= \sum_i^{|H_{\text{full}}|} (\mu_i) P_i + \\
 &\sum_j^{|A|-1} \sum_{\substack{i \\ \forall \{P_j, P_k, P_i\}=0}}^{|H_{\text{full}}|} \mu_{ij} P_j P_k P_i + \\
 &\sum_j^{|A|-1} \sum_i^{|H_{\text{full}}|} \sum_{\substack{l > j \\ \forall \{P_i, P_j, P_l\}=0}}^{|A|-1} \mu_{ijl} P_i P_j P_l.
 \end{aligned} \tag{5.26}$$

The Pauli operators P_j , P_k and P_l are operators in \mathcal{A} , further details are covered in Appendix A.3. Overall, this unitary transformation causes the number of terms in the Hamiltonian to scale as $\mathcal{O}(|H_{\text{full}}| \cdot |\mathcal{A}|^2)$. This scaling is quadratic in the size of \mathcal{A} and as $|\mathcal{A}| \leq 2n + 1$ [256], the number of terms in the rotated system will at worst scale quadratically with the number of qubits n . In a different context, this scaling result was also obtained for involutory linear combinations of entanglers [260]. Overall, unlike the sequence of rotations approach, this non-Clifford operation does not cause the number of terms in a Hamiltonian to increase exponentially.

The transformation given in Equation 5.26 $H_{\text{full}} \mapsto H_{\text{full}}^{LCU'} = R_{LCU}^\dagger H_{\text{full}} R_{LCU}$ is performed classically in CS-VQE. This is efficient to do because it just involves Pauli operator multiplication, which can be done symbolically or via a symplectic approach [261]. This operation could be applied within the quantum circuit. However, contrary to the deterministic sequence of rotations approach, this implementation would be probabilistic as it requires post-selection on an ancillary register [198, 209, 210, 221]. Amplitude amplification techniques could improve this but would require further coherent resources [236–239]. Doing this transformation in a classical pre-processing step, therefore, reduces the coherent resources required and at worst increases the number of terms needing measuring quadratically with respect to the number of qubits.

5.2.6 CS-VQE implementation

In the original CS-VQE proposal [244], $U_{\mathcal{W}}(\vec{q}, \vec{r})$ was fixed to include all the stabilizers of the noncontextual ground state $\mathcal{W} \equiv \mathcal{W}_{\text{all}}$ (Equation 5.20), rather than possible subsets $\mathcal{W} \subseteq \mathcal{W}_{\text{all}}$. The whole Hamiltonian was mapped according to $H_{\text{full}} \mapsto H_{\text{full}}^{\mathcal{W}_{\text{all}}} = U_{\mathcal{W}_{\text{all}}}^\dagger H_{\text{full}} U_{\mathcal{W}_{\text{all}}}$. In general $A(\vec{r}) \in \mathcal{W}$ and $U_{\mathcal{W}_{\text{all}}}$ will therefore normally include the unitary partitioning operator R . The problem with this approach is the unitary R is not a Clifford operation and the transformation can cause the number of terms in the Hamiltonian to increase. This increase is exponential if R_S is used and quadratic if

R_{LCU} is employed. As this step can generate more terms, R should only be included in $U_{\mathcal{W}}$ if the eigenvalue of $A(\vec{r})$ is fixed to $+1$; otherwise, it is a redundant operation as the spectrum of the operator rotated by R is unchanged. We therefore modify the CS-VQE algorithm to construct $U_{\mathcal{W}}$ from the CS-VQE noncontextual generator eigenvalues that are fixed. This means $\mathcal{W} \subseteq \mathcal{W}_{all}$ and ensures that the number of terms can only increase if the eigenvalue of $A(\vec{r})$ is fixed.

5.3 Numerical results

We split our results into two subsections. First, we explore a toy problem, showing the steps of the CS-VQE algorithm. We show how classically applying R without fixing the eigenvalue of $A(\vec{r})$ to $+1$ can unnecessarily increase the number of terms in a Hamiltonian without changing its spectrum. Finally, in Section 5.3.3 we apply measurement reduction combined with CS-VQE to a set of electronic structure Hamiltonians and show that this can significantly reduce the number of terms requiring separate measurement. These results have been published in [262].

5.3.1 Method

We investigated the same electronic structure Hamiltonians considered in the original CS-VQE paper [244]. All molecules considered had a multiplicity of 1 and thus a singlet ground state. The same qubit tapering was performed to remove the \mathbb{Z}_2 symmetries [124]. For each tapered Hamiltonian, we generate a set of reduced Hamiltonians $\{Q_{\mathcal{W}}^\dagger U_{\mathcal{W}}^\dagger H_{full} U_{\mathcal{W}} Q_{\mathcal{W}}\}$ where the size of \mathcal{W} varies from 1 to $|\mathcal{W}_{all}|$, representing differing noncontextual approximations, as summarised in Section 5.2.6. To generate the different CS-VQE Hamiltonians, we modify the original CS-VQE source code used in [244, 263]. The code was modified to implement the unitary partitioning step of CS-VQE if and only if the eigenvalue of $A(\vec{r})$ was fixed. This ensured the number of terms in the rotated Hamiltonian didn't increase unnecessarily, as described in Section 5.2.6.

For each electronic structure Hamiltonian generated in this way, we then apply the unitary partitioning measurement reduction scheme to further reduce the number of terms requiring separate measurement [197, 198, 221]. Partitioning into anticommuting sets was performed using NETWORKX [225]. A graph of the qubit Hamiltonian is built, where nodes represent Pauli operators and edges are between nodes that commute. A graph coloring can be used to find the anticommuting cliques of the graph. This searches for the minimum number of colors required to color the graph, where no neighbors of a node can have the same color as the node itself. The "largest first" coloring strategy in NETWORKX was used in all cases [225, 226].

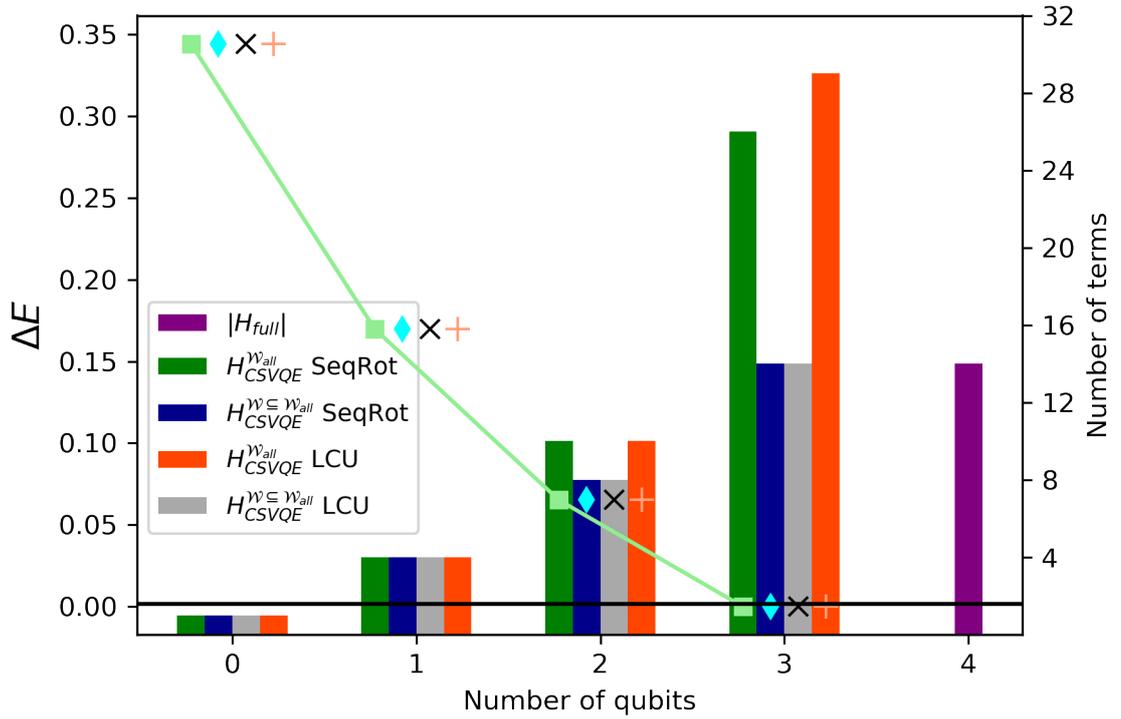


Figure 5.2: Summary of different contextual subspace approximations for the toy problem given in equation 5.27. Each Hamiltonian has been transformed as $Q_{\mathcal{W}}^{\dagger} U_{\mathcal{W}}^{\dagger} H U_{\mathcal{W}} Q_{\mathcal{W}}$, apart from the 4 qubit case, which is the full H . The scatter plot is associated with the left-hand y-axis and gives the energy error of the ground-state as: $\Delta E = |E_{approx} - E_{true}|$. The bar chart gives the number of terms in each Hamiltonian and is associated with the right-hand y-axis. From left to right the following generators are fixed: $\{YIYI, IXYI, IIIZ, \mathcal{A}(\vec{r}_0)\}$, $\{IXYI, IIIZ, \mathcal{A}(\vec{r}_0)\}$, $\{IXYI, IIIZ\}$, $\{IIIZ\}$ and $\{\}$. The $\mathcal{W} = \{\}$ case represents standard full VQE over the full problem. The 0 qubit case presents the scenario where the problem is fully noncontextual and no quantum correction is made. The full details on how each Hamiltonian is built is provided in Appendix A.4. The horizontal black line indicates an absolute error of 1.6×10^{-3} . SeqRot, sequence of rotations

We calculate the ground state energy of each Hamiltonian in this study by directly evaluating the lowest eigenvalues. This was achieved by diagonalising them on a conventional computer.

5.3.2 Toy example

We consider the qubit Hamiltonian:

$$\begin{aligned}
 H = & 0.6 I I Y I + 0.7 X Y X I + 0.7 X Z X I + 0.6 X Z Z I + \\
 & 0.1 Y X Y I + 0.7 Z Z Z I + 0.5 I I I Z + 0.1 X X X I + \\
 & 0.5 X X Y I + 0.2 X X Z I + 0.2 Y X X I + 0.2 Y Y Z I + \\
 & 0.1 Y Z X I + 0.1 Z Y Y I,
 \end{aligned} \tag{5.27}$$

and use it to exemplify the steps of the CS-VQE algorithm. The results are reported to three decimal places and full numerical details can be found in Appendix A.4.

Following the CS-VQE procedure [244], we first split the Hamiltonian into its contextual and noncontextual parts (Equation 5.4):

$$\begin{aligned}
 H_{\text{noncon}} = & 0.5 \underbrace{III Z}_{\mathcal{Z}} + \\
 & 0.7 XZ XI + 0.7 ZZ ZI + \\
 & 0.1 YXYI + 0.6 IYI + \\
 & \underbrace{0.7 XY XI + 0.6 XZ ZI}_{\mathcal{T}},
 \end{aligned} \tag{5.28a}$$

$$\begin{aligned}
 H_{\text{con}} = & 0.1 XXXI + 0.5 XXYI + 0.2 XXZI + \\
 & 0.2 YXXI + 0.2 YYZI + 0.1 YZ XI + \\
 & 0.1 ZY YI.
 \end{aligned} \tag{5.28b}$$

Each row after the first in Equation 5.28a, is a clique of \mathcal{T} . From here, we define the set \mathcal{R} (Equation 5.12):

$$\mathcal{R} = \underbrace{\{YIYI, IXYI, III Z\}}_{\mathcal{G}} \cup \underbrace{\{XZ XI, YXYI, XY XI\}}_{\{P_0^{(j)} | j=0,1,\dots,N-1\}}. \tag{5.29}$$

Note how different combinations of the operators in Equation 5.29 allow all the operators in H_{noncon} (Equation 5.28a) to be inferred under the Jordan product, defined as: $P_a \circ P_b = \frac{\{P_a, P_b\}}{2}$.

The expectation value for H_{noncon} can be induced, by setting the expectation values of operators in \mathcal{R} (Equation 5.29), as the Pauli operators in H_{noncon} are generated by \mathcal{R} under the Jordan product. The expectation value of each operator in H_{noncon} can therefore be inferred without contradiction. To find the ground-state of H_{noncon} , we checked all possible ± 1 expectation values for each G_j ($2^3 = 8$ possibilities). For each possible ± 1 combination, the energy was minimized with respect to the unit vector \vec{r} , which sets the expectation value for each $\langle P_0^{(j)} \rangle = r_j$. The vector (\vec{q}, \vec{r}) that was found to give the lowest energy defines the noncontextual ground-state. In this case the ground-state is:

$$\left(\underbrace{-1, +1, -1}_{\vec{q}_0}, \underbrace{+0.253, -0.658, -0.709}_{\vec{r}_0} \right). \tag{5.30}$$

This noncontextual state defines the operator $\mathcal{A}(\vec{r}_0)$:

$$\mathcal{A}(\vec{r}_0) = 0.253 YXYI - 0.658 XY XI - 0.709 XZ XI. \tag{5.31}$$

5.3. NUMERICAL RESULTS

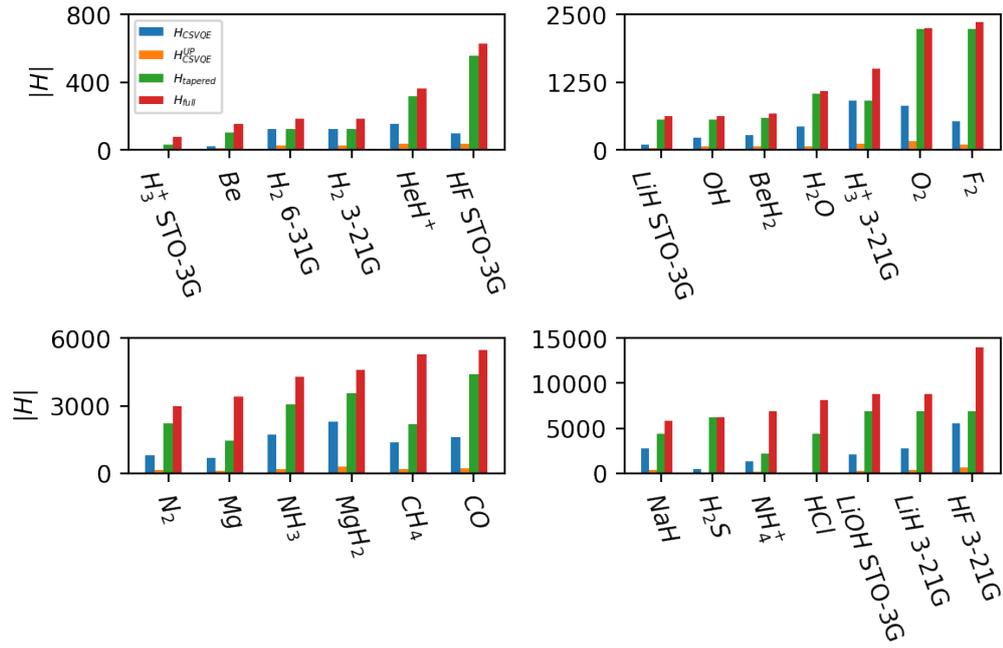


Figure 5.3: Number of Pauli operators requiring separate measurement to determine the ground-state energy of a particular molecular Hamiltonian to within chemical accuracy. For each molecule, the full Hamiltonian, tapered Hamiltonian, CS-VQE and CS-VQE with unitary partitioning measurement reduction applied are given. Full numerical details of each are provided in Appendix A.5. The size of the Hamiltonian for LiH (3-21G singlet) with measurement reduction applied is different for the sequence of rotations and LCU unitary partitioning methods. This is an artifact of the graph color heuristic finding different anticommuting cliques in the CS-VQE Hamiltonian.

From this we can write \mathcal{W}_{all} (Equation 5.20):

$$\mathcal{W}_{all} \equiv \{A(\vec{r}_0)\} \cup \{-YIYI, IXYI, -IIIZ\} \quad . \quad (5.32)$$

To map $A(\vec{r}_0)$ to a single Pauli operator we use unitary partitioning [197, 198, 221]. The required unitary can be constructed as either a sequence of rotations [198],

$$R_S = e^{-1i \cdot 0.788 \cdot ZYZI} \cdot e^{+1i \cdot 1.204 \cdot ZZZI}, \quad (5.33)$$

or linear combination of unitaries [198],

$$R_{LCU} = 0.792 IIII + 0.416i ZZZI - 0.448i ZYZI. \quad (5.34)$$

These operators perform the following reduction: $R_S A(\vec{r}_0) R_S^\dagger = R_{LCU} A(\vec{r}_0) R_{LCU}^\dagger = YXYI$.

If the eigenvalue of $A(\vec{r}_0)$ is fixed, then we should consider \mathcal{W}_{all} (Equation 5.32)

under the unitary transform R_{LCU} or R_S (Equation 5.20):

$$\begin{aligned}\mathcal{W}'_{all} &= R_{S/LCU} A(\vec{r}) R_{S/LCU}^\dagger \cup \{-YIYI, IXYI, -IIIZ\} \\ &= \{YXYI\} \cup \{YIYI, IXYI, IIIZ\}.\end{aligned}\tag{5.35}$$

Next, we define different $U_{\mathcal{W}}$ (Equation 5.22), depending on which stabilizers \mathcal{W} we wish to fix. For this problem we found the optimal ordering of which stabilizers to fix to be:

1. $\{-1 YIYI, +1 IXYI, +1 \mathcal{A}(\vec{r}_0), -1 IIIZ\}$
2. $\{+1 IXYI, -1 IIIZ, +1 \mathcal{A}(\vec{r}_0)\}$
3. $\{+1 IXYI, -1 IIIZ\}$
4. $\{-1 IIIZ\}$.

This ordering was found by a brute force search over all $\sum_{i=1}^{|\mathcal{W}_{all}|} \binom{|\mathcal{W}_{all}|}{i} = 2^4 - 1 = 15$ possibilities for \mathcal{W} .

The members of the resulting set of four different \mathcal{W} each represent different non-contextual approximations. These give four different $U_{\mathcal{W}}$ built according to Equation 5.22. The full definition of each operator is given in Appendix A.4.

Taking a specific example, for $\mathcal{W} = \{+IXYI, -IIIZ, +\mathcal{A}(\vec{r}_0)\}$ we define $U_{\mathcal{W}}^\dagger$ (Equation 5.22). This operator transforms \mathcal{W} as $\mathcal{W}^Z = U_{\mathcal{W}}^\dagger \mathcal{W} U_{\mathcal{W}} = \{+IZII, -IIIZ, +IIZI\}$. The eigenvalues of the operators in \mathcal{W}^Z are fixed by the noncontextual state to be $\langle IZII \rangle = +1$, $\langle IIZI \rangle = +1$, $\langle IIIZ \rangle = -1$. This defines the projector:

$$\begin{aligned}Q_{\mathcal{W}} &= \left(\underbrace{|0\rangle\langle 0| + |1\rangle\langle 1|}_{\mathcal{I}_{(n-|\mathcal{W}^Z|)}} \right) \otimes \underbrace{|0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes |1\rangle\langle 1|}_{|\psi_{\text{fixed}}\rangle} \\ &= I \otimes |001\rangle\langle 001|.\end{aligned}\tag{5.36}$$

The reduced Hamiltonian is therefore

$$\begin{aligned}H \mapsto H_{\mathcal{W}}^{LCU} &= Q_{\mathcal{W}}^\dagger U_{\mathcal{W}}^\dagger (LCU) H_{\text{full}} U_{\mathcal{W}} (LCU) Q_{\mathcal{W}} \\ &= -1.827 I - 0.414 X - 0.292 Z + 0.648 Y.\end{aligned}\tag{5.37}$$

Appendix A.4 gives further details about this operation and provides the specifics for the other projected Hamiltonians.

Overall four Hamiltonians are generated, representing different levels of approximation, that act on 0, 1, 2 and 3 qubits respectively. The 4-qubit case represents the standard VQE on the full Hamiltonian. Figure 5.2 summarizes the error ΔE of each of these compared with the true ground-state energy (scatter plot). The number of terms in each Hamiltonian is given by the bar chart. The green and orange results have $\mathcal{W} \equiv \mathcal{W}_{all}$ for all cases and represent the old CS-VQE implementation. For these

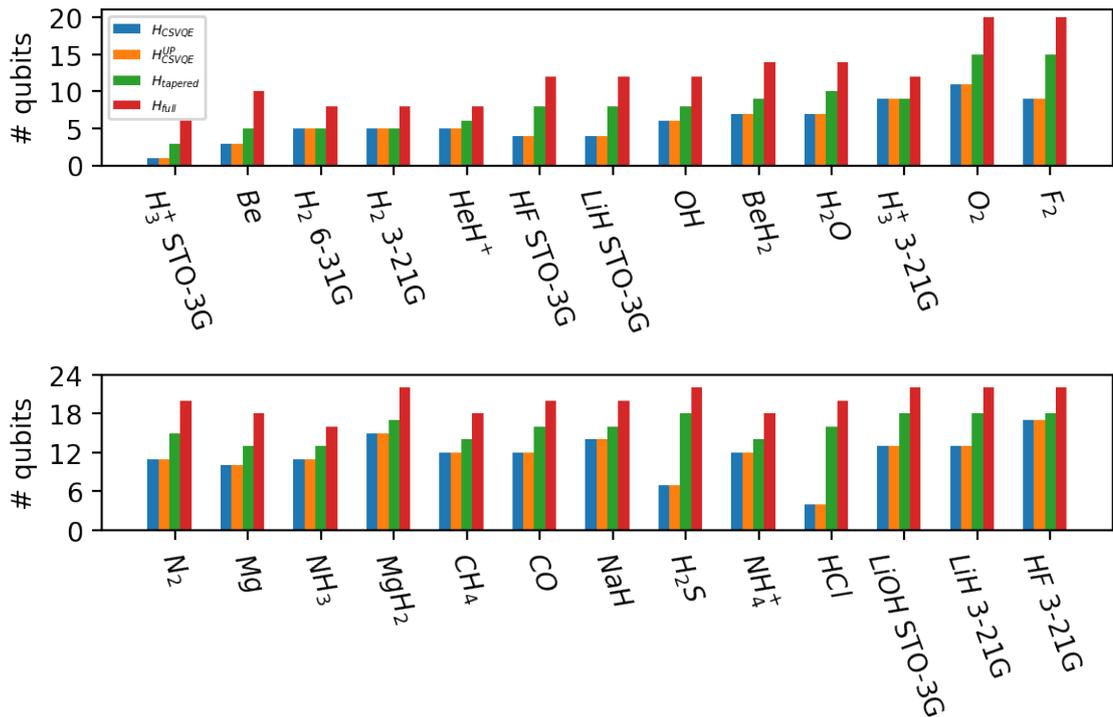


Figure 5.4: Number of qubits required to simulate different electronic structure Hamiltonians in order to achieve chemical accuracy. For each molecule the full Hamiltonian, tapered Hamiltonian, CS-VQE and CS-VQE with unitary partitioning measurement reduction applied are given. Numerical details for each result are provided in Appendix A.5.

results, in the 3 and 4 qubit Hamiltonians have an increased number of terms due to $R_{S/LCU}$ being implemented, even though the eigenvalue of $\mathcal{A}(\vec{r}_0)$ is not being fixed to +1. On the other hand, the gray and blue results in Figure 5.2 build $U_{\mathcal{W}}$ according to Equation 5.22, where $\mathcal{W} \subseteq \mathcal{W}_{all}$. This approach ensures that $R_{S/LCU}$ is only applied when necessary.

5.3.3 Measurement reduction

Figures 5.3 and 5.4 summarize the results of applying the unitary partitioning measurement reduction strategy to a set of electronic structure Hamiltonians. We report the number of terms and number of qubits in each Hamiltonian required to achieve chemical accuracy compared with the original problem. Appendix A.5 gives further information about each result, where the different levels of noncontextual approximation are shown. As previously discussed in [244], even though CS-VQE in general is an approximate method, chemical accuracy can still be achieved using significantly fewer qubits. Applying unitary partitioning on-top of the reduced CS-VQE Hamiltonians required to achieve chemical accuracy can further reduce the number of terms by roughly an order of magnitude. This is consistent with the previous results in

Molecule	Basis	Number of gates for R_S
BeH ₂	(STO-3G)	[90, 72]
Mg	(STO-3G)	[189, 162]
H ₃ ⁺	(3-21G)	[209, 176]
O ₂	(STO-3G)	[184, 160]
OH ⁻	(STO-3G)	[104, 80]
CH ₄	(STO-3G)	[325, 286]
Be	(STO-3G)	[14, 8]
NH ₃	(STO-3G)	[299, 260]
H ₂ S	(STO-3G)	[120, 96]
H ₂	(3-21G)	[66, 48]
HF	(3-21G)	[735, 672]
F ₂	(STO-3G)	[133, 112]
HCl	(STO-3G)	[36, 24]
HeH ⁺	(3-21G)	[88, 64]
MgH ₂	(STO-3G)	[403, 364]
CO	(STO-3G)	[325, 286]
LiH	(STO-3G)	[36, 24]
N ₂	(STO-3G)	[207, 180]
NaH	(STO-3G)	[493, 442]
H ₂ O	(STO-3G)	[120, 96]
H ₃ ⁺	(STO-3G)	[3, 0]
LiOH	(STO-3G)	[378, 336]
LiH	(3-21G)	[459, 408]
H ₂	(6-31G)	[66, 48]
NH ₄ ⁺	(STO-3G)	[325, 286]
HF	(STO-3G)	[36, 24]

Table 5.3: Gate requirements to implement R as a sequence of rotations in the unitary partitioning measurement reduction step. The square tuple gives the upper bound on the number of single qubit and CNOT gates required - [*single*, *CNOT*]. These resource requirements are based on the largest anticommuting clique of each Hamiltonian, as these have the largest circuit requirements for R_S .

[221].

To actually obtain a measurement reduction, one needs to show that the number of measurements required to measure the energy of a molecular system, to a certain precision ϵ , is reduced. Currently, Figure 5.3 only shows that we have reduced the number of Pauli terms being measured. We have not commented on the variance. In Section 4.2.6, we prove that simultaneous measurement of normalized anticommuting cliques can never do worse than performing no measurement reduction and will more often than not give an improvement. The proof given is state independent. There are other measurement strategies based on grouping techniques, such as splitting a Hamiltonian into commuting or qubit-wise commuting cliques [103, 190, 195, 196, 201]. The measurement reduction obtained from these methods is more complicated, as the covariance of operators within a clique must be carefully accounted for [95,

190]. This is one of the reasons we do not analyse the performance of these strategies in this work. Many other measurement methods have also been proposed [163, 172, 177–179, 185, 187, 188, 193, 245] and their effect on the number of measurements would be interesting to investigate.

In Table 5.3, we report the upper bound on the gate count required to implement measurement reduction as a sequence of rotations. The LCU method would require ancilla qubits and analysis of the circuit depth is more complicated. Further analysis can be found in [221]. The number of extra coherent resources required to implement unitary partitioning measurement reduction is proportional to the size of each anti-commuting clique a Hamiltonian is split into [198, 221]. The sequence of rotations circuit depth scales as $\mathcal{O}(N_s(|C| - 1))$ single qubit and $\mathcal{O}(N_s(|C| - 1))$ CNOT gates, where N_s is the number of system qubits and $|C|$ is the size of the anticommuting clique being measured. Table 5.3 reports the gate count upper bound for the largest anticommuting clique of a given CS-VQE Hamiltonian. We do not consider possible circuit simplifications, such as gate cancellations. To decrease the depth of the quantum circuit required for practical application, we suggest finding nonoptimal clique covers; for example, if anticommuting cliques are fixed to a size of 2, the resources required to perform R_S are experimentally realistic for current and near-term devices, as only $\mathcal{O}(N_s)$ single qubit and $\mathcal{O}(N_s)$ CNOT gates are required [221].

The heuristic used to determine the operators in H_{noncon} selected terms in the full Hamiltonian greedily by coefficient magnitude, while keeping the set noncontextual [256]. The Hamiltonians studied here had weights dominated by diagonal Pauli operators, as the Hartree-Fock approximation accounts for most of the energy. This heavily constrains the operators allowed in \mathcal{A} . For the electronic structure Hamiltonians considered in this study, we found in all cases that $|\mathcal{A}| = 2$. In general, we do expect more commuting terms in H_{noncon} than anticommuting terms. This is because there are more possible commuting Pauli operators defined on n qubits compared with anticommuting operators (2^n vs $2n + 1$). \mathcal{G} will therefore in general be the larger contributor to superset \mathcal{R} (Equation 5.12).

In Figure 5.3, the CS-VQE bars have not been split into two for the case when R is constructed as R_{LCU} or R_S . This is due to $|\mathcal{A}|$ being 2 in all cases, which is the special case when these operators (R_{LCU} and R_S) end up being identical. In this instance R has the form $R = \alpha I + i\beta P$ and thus the number of terms will only increase for every term in the Hamiltonian that P anticommutes with. However, in general $|\mathcal{A}|$ will be greater than 2 and the effect of R can dramatically affect the number of terms in the resultant rotated Hamiltonian. We observe this in Fig. 5.2 of the toy example, where the 2 and 3 qubit CS-VQE Hamiltonians have had $U_{W_{all}}$ applied to them even though the eigenvalue of $A(\vec{r})$ is not fixed. In that example, for the 3 qubit approximation the sequence-of-rotations rotated Hamiltonian (green) actually

has fewer terms than the LCU rotated operator (orange). This result is an artifact of the small problem size. In Appendix A.3 we show that the scaling will favor the LCU implementation, where the number of terms in a Hamiltonian can only increase quadratically, not exponentially, when performing the unitary partitioning rotation as a LCU rather than a sequence of rotations.

In Appendix A.5, we show the convergence of CS-VQE at different noncontextual approximations. The results illustrate that CS-VQE can converge to below chemical accuracy well before the case when no noncontextual approximation is made (full VQE). Results beyond convergence are included to show the different possible levels of approximation. In practice, knowledge of the true ground-state energy is not known *a priori* and so using chemical precision to motivate the noncontextual approximation will not be possible. In this setting, a way to approach quantum advantage is to note that CS-VQE is a variational method. The quantum resources required can be expanded until the energy obtained by CS-VQE is lower than that coming from the best possible classical method. At this point, either the algorithm can be terminated or further contextual corrections can be added until the energy converges, at which point the algorithm should be stopped.

5.4 Conclusion

The work presented in this chapter shows that combining the unitary partitioning measurement reduction strategy with the CS-VQE algorithm can further reduce the number of terms in the projected Hamiltonian requiring separate measurement by roughly an order of magnitude for a given molecular Hamiltonian. The number of qubits needed to achieve chemical accuracy in most cases was also dramatically decreased, for example the H₂S (STO-3G singlet) problem was reduced to 7 qubits from 22.

We also improve two parts of the CS-VQE algorithm. First, we avoid having to apply the unitary partitioning operator R after the ansatz which averts the potential exponential increase in the number of Pauli operators of the CS-VQE Hamiltonian caused by classically computing the non-Clifford rotation of the full Hamiltonian when R is defined as a sequence of rotations [198, 244]. We show that applying this operation as a linear combination of unitaries [198]: $H_{\text{full}} \mapsto H_{\text{full}}^{LCU'} = R_{LCU}^\dagger H_{\text{full}} R_{LCU}$, results in the number of terms at worst increasing quadratically with the number of qubits. This result makes classically precomputing this transformation tractable and R no longer needs to be performed coherently after the ansatz. Secondly, we define the unitary $U_{\mathcal{W}}$, which maps each stabilizer in \mathcal{W}_{all} (equation 5.21) to a distinct single-qubit Pauli matrix, according to which stabilizer eigenvalues are fixed by the noncontextual state. This ensures that the non-Clifford rotation required by CS-VQE is only applied

when necessary and also reduces the number of redundant Clifford operations that are classically performed.

There are still several open questions for the CS-VQE algorithm. We summarize a few here. (1) What is the best optimization strategy to use when minimizing the energy over (\vec{q}, \vec{r}) in the classical noncontextual problem? (2) What heuristic is best to construct the largest $|H_{\text{noncon}}|$? (3) How can we efficiently determine which non-contextual stabilizers to fix while maintaining low errors? In this work, the size of each electronic structure problem allowed us to classically compute the ground-state energies at each step, but if this is not possible then VQE calculations would be required. However, as each run requires fewer qubits and decreases the number of terms requiring separate measurement this approach may overall still be less costly than performing VQE over the whole problem, especially when combined with further measurement reduction strategies. (4) What are the most important terms to include in H_{con} or equivalently in H_{noncon} ? Currently, it is not known whether $|H_{\text{noncon}}|$ should be maximized or whether selecting high-priority terms [259] from the whole Hamiltonian results in a better approximation for a given problem. We leave these questions to future work. We have written an open-source CS-VQE code [264] that includes all the updated methodologies discussed in this chapter - in particular using the LCU approach to unitary partitioning and constructing $U_{\mathcal{W}}$ (equation 5.22) according to which operators in \mathcal{W} have their eigenvalue fixed. We hope to further expand this codebase in the future.

Chapter 6

Multiscale Quantum Simulation via Projection-Based Embedding

Quantum embedding schemes for electronic structure problems aim to reduce the computational cost of a problem by dividing a molecular problem into smaller (and so less costly) subsystems [265]. There are many different embedding approaches that each divide a problem in different ways. Most embedding methods focus on dividing a given problem into what makes the most chemical sense, for example avoiding breaking bonds, which can be important to a given property being simulated [265]. The important aspect of embedding techniques is different levels of theory can be “mixed and matched”. We restrict this chapter to only consider “QM:QM” embedding, which involves embedding a quantum mechanical (QM) system in another quantum mechanical system. We do not discuss “QM:MM”, which involves a quantum mechanical calculation embedded in a molecular mechanics (MM) simulation. The importance of these methods is evidenced by the 2013 Nobel Prize in chemistry awarded to Martin Karplus, Michel Levitt, and Arieh Warshel “for the development of multiscale models for complex chemical systems” [266].

For QM:QM approaches, the main embedding methods are [267]: (1) Green’s function embedding [268–272], (2) density matrix embeddings [273–276] and (3) density functional embeddings. Due to the sheer breadth and number of different embedding techniques, in this chapter we restrict our background to only cover density functional embeddings. We then introduce the projection-based embedding (PBE) model and provide a numerical investigation of the resource reductions obtained for a test bed of large molecules. The importance of this approach is it allows larger systems to be studied on current NISQ hardware.

6.1 Density functional embeddings

Embedding methods based on DFT are often used as they can account for electron exchange and correlation in a computationally tractable fashion. Standard DFT calculations are commonly used to study systems of hundreds of atoms giving properties such as energetics, conformational properties and magnetic properties among many

6.1. DENSITY FUNCTIONAL EMBEDDINGS

Citation	Date first proposed	Method
Cortona [277]	1991	Subsystem DFT (sDFT)
Wesolowski <i>et al.</i> [278], Wesolowski <i>et al.</i> [279]	1993	Frozen Density Embedding Theory (FDET)
Govind <i>et al.</i> [280], Klüner <i>et al.</i> [281], Libisch <i>et al.</i> [282]	1999	Density Functional Embedding Theory (DFET)
Elliot <i>et al.</i> [283], Elliot <i>et al.</i> [284]	2009	Partition Density Functional Theory (PFET)
Huang <i>et al.</i> [285]	2011	Potential-Functional Embedding Theory (PFET)
Manby <i>et al.</i> [286, 287]	2012	Projection-Based Embedding (PBE)
Mosquera <i>et al.</i> [288] Mosquere <i>et al.</i> [289]	2018	Locally Coupled Open Subsystems (LCOS)
Mosquere <i>et al.</i> [290]	2019	Domain Separated DFT (DS-DFT)

Table 6.1: A brief history of different DFT embedding methods.

others. For further details see [291] and references therein. Embedding approaches seek to utilize these techniques with other approximations.

To our knowledge, the early ideas of DFT embedding were first discussed by Parr *et al* [292]. Cortona [277] and closely after Wesolowski and Warshel [278] formalized these ideas forming the groundwork of subsequent methods.

Consider the ground-state density ρ_A of a molecular system A , adjusted through an external field v_A^{emb} , where the field comes from the environment. Through the Coulomb interaction and indirectly the Pauli exclusion principle, the energy of A is modified by its interactions with the environment. These energetic contributions to the ground state are contained within a ΔE term, where the embedding DFT framework provides rigorous foundations for v_A^{emb} and ΔE as functionals [267].

To understand the embedding process, we first need to consider the exact energy density functional for the full system (equation 2.96). For readability, we repeat this here:

$$E[\rho] = F[\rho] + \int d^3\vec{r} v_{ext}(\vec{r})\rho(\vec{r}). \quad (6.1)$$

Consider this functional written in terms of the Kohn-Sham kinetic, Coulomb, external, and exchange-correlation terms (equation 2.101). The Euler equation for this

problem is given by equation 2.104, which again we reproduce here as:

$$\begin{aligned}\mu &= \frac{\partial F[\rho]}{\partial \rho(\vec{r})} + v_{ext}(\vec{r}) \\ &= \frac{\partial T_s[\rho]}{\partial \rho(\vec{r})} + v_{eff}(\vec{r}),\end{aligned}\tag{6.2}$$

where:

$$\begin{aligned}v_{eff}(\vec{r}) &= \frac{\partial J[\rho]}{\partial \rho(\vec{r})} + \frac{\partial E_{xc}[\rho]}{\partial \rho(\vec{r})} + v_{ext}(\vec{r}) \\ &= \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' + v_{xc}(\vec{r}) + v_{ext}(\vec{r}) \\ &= v_H(\vec{r}) + v_{xc}(\vec{r}) + v_{ext}(\vec{r}).\end{aligned}\tag{6.3}$$

As discussed in Section 2.9 , $E_{xc}[\rho]$ and v_{xc} account for the nontrivial correlation effects and must be approximated.

In DFT embedding, the total energy $E[\rho]$ (equation 6.1) is split into two parts based on the fragment energy and its remainder [267]:

$$E[\rho] = E[\rho_A] + \Delta E[\rho, \rho_A],\tag{6.4}$$

where

$$\frac{\partial}{\partial \rho_A} \left(E[\rho_A] + \Delta E[\rho, \rho_A] \right) - \mu = 0.\tag{6.5}$$

This result is the same as the Euler equation for A placed in an external field v_A^{emb} , by choosing [267]:

$$v_A^{emb} = \frac{\partial \Delta E[\rho, \rho_A]}{\partial \rho_A}.\tag{6.6}$$

This definition for v_A^{emb} gives the exact embedding potential that yields the exact subsystem density ρ_A .

In DFT embedding, it is common to divide the total electron density as:

$$\rho = \rho_{total} = \rho_A + \rho_B.\tag{6.7}$$

It is standard practice to refer to ρ_A as the density of the embedded subsystem and ρ_B as the environment density. An advantage of this methodology is if ρ_A and ρ_B are N -representable then so is ρ_{total} [267, 293].

Written in terms of subsystems and the environment, v_A^{emb} can be written in its

components [267]:

$$\begin{aligned} v_A^{emb} &= \frac{\partial}{\partial \rho_A} \left(T_s[\rho_{total}] - T_s[\rho_A] \right) + v_H \underbrace{[\rho_{total} - \rho_A]}_{\rho_B} + \frac{\partial}{\partial \rho_A} \left(E_{xc}[\rho_{total}] - E_{xc}[\rho_A] \right) \\ &= v_s^{nadd} + v_H^B + v_{xc}^{nadd}, \end{aligned} \quad (6.8)$$

where v_s^{nadd} is the nonadditive kinetic potential, v_H^B is the environment Coulomb potential and v_{xc}^{nadd} is the nonadditive exchange-correlation potential [267]. The nonadditive kinetic potential (NAKE) is the largest contribution and includes the exclusion principle that forces electrons in the active subsystem to occupy states orthogonal to those in the environment [267, 277, 278]. Explicitly, we can write the nonadditive kinetic energy as [278]:

$$T_s[\rho_{total}] = T_s[\rho_A + \rho_B] = T_s[\rho_A] + T_s[\rho_B] + T_s^{nadd}[\rho_A, \rho_B]. \quad (6.9)$$

The nonadditive exchange-correlation term can be obtained in the same way. These nonadditive terms account for the interaction between subsystems.

Given this formalism, we give a brief chronology of the development of DFT embeddings, which is summarised in Table 6.1 and use this timeline as a natural flow of our discussion. Subsystem DFT was first proposed by Cortona [277]. The main idea is each atom is treated as a subsystem and assigned a particular electron density [265, 277]. The total electron density is then simply a sum of all these atomic like densities, where the intratomic kinetic energy is obtained exactly and the interatomic kinetic energy approximated. The exchange-correlation functional is also approximated by the local density approximation (LDA). A modern account of this method is presented by Jacob and Neugebauer in [294]. They relate the approximations of this model back to the non-additive contribution arising from the exchange-correlation functional and the non-additive kinetic energy.

Wesolowski and Warshel then introduced Frozen Density Embedding theory (FDET) [278]. They split a system into two fragments. One is kept frozen (or fixed) and used to generate an effective potential for the active system, which is then solved self-consistently. Explicitly, this would involve keeping ρ_B in equation 6.7 fixed and solving for ρ_A . This technique was then extended by Wesolowski to iteratively solve subsystem ρ_A keeping ρ_B fixed and then solving ρ_B while keeping the updated ρ_A density fixed. This procedure is repeated until both densities converge and is commonly known as “freeze-and-thaw” iterations. It is possible to further extend this theory to multiple subsystems. A richer discussion on this is outside the scope of this thesis, but a review of FDET and its extensions may be found in [279].

Then in 1999, Govind *et al.* introduced density functional embedding theory

(DFET). This combined DFT with first principle correlation techniques allowing post-SCF methods to be applied to the active system. The energy is found according to [280–282]:

$$\begin{aligned} E_{total}^{emb} &= E_A^{DFT} + E_B^{DFT} + E_{nonadd}^{DFT} + (E_A^{WFT} - E_A^{DFT}) \\ &= E_{total}^{DFT} + (E_A^{WFT} - E_A^{DFT}), \end{aligned} \quad (6.10)$$

where *WFT* denotes post-SCF wavefunction calculations of the active system of interest. This requires the Hamiltonian H of the active system to be modified to $H_{emb} = H + V_A^{emb}$ [282]. Region A is not treated in isolation, but rather in the presence of an embedding potential that is generated from a DFT calculation on the environment fragments. This allows a user to systematically improve the description of energetics in a local region. It should be noted that this subtractive method, requires a well-converged reference density and reference energy for the total system [265], as seen by the form of equation 6.10. Again, the kinetic-energy contribution to $v_A^{emb}(\vec{r})$ is the most challenging component [280]. A problem with all wavefunction in DFT embeddings can be seen in equation 6.10, where the error in the nonadditive energy of the DFT calculation requires exactly cancelling the wavefunction description of the system - i.e. the $E_A^{WFT} - E_A^{DFT}$ term requires exact error cancellation in E_{nonadd}^{DFT} . Incomplete cancellation is denoted as “double counting” [267].

Klüner *et al.* extended the DFET method by improving descriptions of the ground state and treating excited states. It is noted in [280] that this scheme is exact in theory, but in practice there is ambiguity coming from the kinetic and exchange-correlation contribution to the embedding potential.

A conceptual problem that arises for DFET is the problem of infinitely many ways to obtain a partitioning of the total electron density (equation 6.7) [295]. Carter and coworkers devised a way to avoid the non-uniqueness of the density partitioning by requiring that the active and environment subsystems share the same embedding potential, $v_A^{emb} = v_B^{emb}$, where the number of electrons in each subsystem is fixed [296]. This follows from similar ideas used in Partition Density Functional Theory (PDFT).

Density Functional Partition Theory (DFPT) was introduced in 2009 and later revised to Partition Density Functional Theory (PDFT) by Elliot *et al.* [283, 284]. The technique involves dividing the total density $\rho_{total}(\vec{r})$ into fragments, while ensuring the sum of fragment densities matches the molecular density. The system of interacting fragments is mapped to an effective system of non-interacting fragments that share a common embedding potential [285, 297]. In the context of PDFT this is known as the partition potential and is found as the Lagrange multiplier in a functional minimization of the sum of fragment energies [294]. Unlike DFET, PDFT is formulated without constraints on having integer electron numbers in each subsystem, allowing fractional occupation to occur in each subsystem. The only constraint is the

total number of electrons is fixed [284, 294, 296].

Both DFPT and PDFT do not consider changes in the environment due to the wavefunction description of the active system [282]. A more accurate theory should update the embedding potential from the wavefunction description of the active system. Carter and Huang in [285] developed potential-functional embedding theory (PFET) to do this, where rather than optimizing the electron densities, the embedding potential is optimized instead. This formalism is based on the fact that the embedding potential is a unique property of the total system for a given partitioning [282].

A deficiency of many subsystem based embedding techniques is if an embedding potential strongly delocalises the density of the embedded system over to the environment or vice versa. Under such regimes, these subsystem methods are likely to have trouble describing charge-transfer processes [265]. Mosquera *et al.* introduced tools to deal with this issue.

In 2018, Mosquera *et al.* proposed the Locally Coupled Open Subsystem (LCOS) method [288, 289]. This uses an auxiliary wavefunction expressed as a linear combination of tensor products of subsystem states, in which one can use fully correlated wavefunctions or Kohn-Sham Slater determinants. For this system, the auxiliary Hamiltonian is expressed over subsystem Hamiltonians plus a coupling operator that allows electron transfer between subsystems. This coupling operator, which the authors denote the “coupling potential”, can be estimated using density functional approximations or machine learning approaches [289]. If the charge-transfer operator in the DFT version of LCOS is not included then it becomes equivalent to PDFT and can be implemented as an extension of other methods such as FDET, subsystem DFT and PFET [265]. In fact, the authors recommend integration of the LCOS approach with such fragment-based formalisms [289].

Chan and Sun in [267] comment on how one of the major disadvantages of DFT embedding is that when examining ρ_A alone it is difficult to distinguish between the scenarios of A bonded to the environment and one that isn't. This is due to the density by definition not containing off-diagonal density matrix correlations describing entanglement. In principle, all these effects are given by the exact density functional; however, in all practical instances this has to be approximated and so the lack of the off-diagonal information can pose difficulties for density functional approximations used in practice. We note that the LCOS approach may be able to correct some of these deficiencies; however, further investigation is required.

Another embedding approach is denoted as Domain Separated Density Functional Theory (DS-DFT). The main idea behind this method is using the electronic repulsion operator to motivate decomposing a problem into different volume elements. This partitioning allows users to apply different levels of theory to different regions of space.

In other words, different regions of a molecule may be studied by different types of electron interaction operators (and thereby different levels of theory). Practical implementation of this method requires density functional approximations to subtract the global Hartree exchange-correlation energy associated with the regions that are to be computed with nonlocal interactions or explicit wavefunction methods [290].

In many of the DFT embeddings methods introduced, most suffer from approximating the nonadditive kinetic potential v_s^{nadd} (equation 6.8) and energy $T_s^{nadd}[\rho_A, \rho_B]$ (equation 6.9). A problem arises because the KS molecular orbitals of the whole system are not available [295]. One approach involves approximating T_s using a density functional approximation, such as the Thomas-Fermi functional [277, 278]. However, this strategy is limited by approximation of the kinetic energy term [267]. More recently, there have been proposals to calculate the nonadditive kinetic potential v_s^{nadd} and $\frac{\partial T_s}{\partial \rho_A}$ numerically [298–300]. As the noninteracting ground state density can be found from an external potential $v_s \mapsto \rho$ with a cost $\mathcal{O}(N^3)$ where N is the number of electrons [267], inversion of this mapping is tractable [267, 301]. Such an inversion determines a KS potential and orbitals corresponding to a given electron density. Common KS inversion methods are Zhao-Morrison-Parr (ZMP) [302] and Wu-Yang (WY) [301]. The KS-pies python library offers a useful open-source implementation of these methods [303]. However, in practice there can be numerical issues leading to nonphysical potentials with wild oscillations that lead to poor approximations of v_s [304, 305].

The issue of the nonadditive kinetic potential and energy under certain conditions is not a problem. Wesolowski and Weber in Appendix A of [306] prove that the nonadditive kinetic energy is zero for subsystems electron densities ρ_A and ρ_B that do not overlap for all points in space. The projection-based embedding (PBE) framework proposed by Manby and coworkers doesn't have the issue of nonadditive kinetic energy due to this [286]. As will be discussed in the next section, this is due to a level-shifting projector keeping the orbitals of different subsystems orthogonal to one another.

6.2 Theory of projection-based embedding

Just as the other embedding methods discussed in the previous section, in projection-based embedding (PBE) the total system density is written as a sum of two subsystems - i.e. $\rho_{total} = \rho_A + \rho_B$ (equation 6.7). The Kohn-Sham DFT energy is then written as a combination of subsystem energies and a nonadditive term [286]:

$$E[\rho_A + \rho_B] = E[\rho_A] + E[\rho_B] + \Delta E[\rho_A, \rho_B], \quad (6.11)$$

where subsystem A defines the active system and B defines the environment subsystem. As discussed in the previous section, the $\Delta E[\rho_A, \rho_B]$ term contains the exchange-correlation effects between subsystems [286]:

$$\Delta E_{xc}[\rho_A, \rho_B] = E_{xc}[\rho_A + \rho_B] - E_{xc}[\rho_A] - E_{xc}[\rho_B]. \quad (6.12)$$

An advantage of the PBE method is the nonadditive kinetic energy (equation 6.9) is zero and does not need to be approximated. We will see later that this is due to a projector term keeping the molecular orbitals between subsystems orthogonal - i.e. $\langle \psi_i^A | \psi_j^B \rangle = 0$.

The projection-based embedding procedure can be broken down into three major steps: a global calculation, localization step and partitioning step. The following subsections summarise each part.

6.2.1 Global calculation

The first step in the PBE procedure is to perform a low-level (cheap) self-consistent field calculation (SCF) of the whole molecular system under consideration. We call this the global calculation. For both Hartree-Fock and Kohn-Sham DFT calculations, this yields a set of spatial molecular orbitals (MOs) $\{\psi_i(\vec{r}) | i = 1, 2, \dots, \eta\}$. Each MO is formed from a linear combination of K known atomic orbital (AO) basis functions $\{\phi_j(\vec{r}) | j = 1, 2, \dots, K\}$:

$$\psi_i(\vec{r}) = \sum_{j=1}^K C_{ji} \phi_j(\vec{r}), \quad (6.13)$$

where C is a matrix of canonical MO coefficients - just as equation 2.57. In general, the AO basis functions $\phi_j(\vec{r})$ are not orthonormal. We can see this by the $(K \times K)$ overlap matrix:

$$S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle = \int d\vec{r} \phi_\mu(\vec{r})^* \phi_\nu(\vec{r}). \quad (6.14)$$

If S is the identity matrix then all the AO basis functions are orthonormal; however, in general this is not the case. However, linear combinations of these non-orthogonal AOs, given by the columns of C , construct orthogonal MOs $\psi_i(\vec{r})$ - i.e. we need $C^\dagger S C = I$.

In order to partition the molecular problem into an active and environment part, these canonical MOs ψ_i must first be localized and assigned to a subsystem. We introduce how this is done next.

6.2.2 Localization of molecular orbitals

In [307], Lehtola and Jónsson state “Hartree-Fock and KS-DFT is invariant under a unitary transformation of the occupied-occupied and virtual-virtual blocks”. This is true because the Hartree-Fock and KS-DFT wavefunctions are approximated as a single Slater determinant. From linear algebra, it is well known that the determinant of a matrix product is given by the product of their determinants - i.e. for general matrices $\det(AB) = \det(A)\det(B)$. Using this property and the fact that the determinant of the identity matrix is $\det(I) = 1$. Given a unitary matrix V , where $V^\dagger V = I$, we obtain the following:

$$1 = \det(I) = \det(V^\dagger V) = \det(V^\dagger)\det(V) = \det(V)^*\det(V) = |\det(V)|^2, \quad (6.15)$$

which implies that the determinant of any unitary matrix must have a value of $e^{i\theta}$, as $|e^{i\theta}| = 1$. Therefore, acting with a unitary will leave the Slater determinant unchanged up to a global phase and so observable quantities of the wavefunction will be unchanged. Thus, the solution to an SCF problem can be described by a set of different (unitarily) rotated orbitals. Such a unitary rotation U can be used to spatially localize each MO ψ_i as much as possible. The form of U is defined by a particular localization procedure and there are many methods based on different localization criteria. We denote these orbitals as localized molecular orbitals (LMOs) or ψ_i^{LMO} . The matrix of orbital coefficients for these localized orbitals are given by the columns of C^{LMO} defined as [240]:

$$C^{LMO} = CU. \quad (6.16)$$

This construction ensures the orthonormality condition of each molecular orbital is still conserved - i.e. $(C^{LMO})^\dagger SC^{LMO} = I$. We see this via the following proof:

$$\begin{aligned} (C^{LMO})^\dagger SC^{LMO} &= U^\dagger C^\dagger SCU \\ &= U^\dagger IU = IU^\dagger U \\ &= I. \end{aligned} \quad (6.17)$$

This construction allows for U to be determined from C and C^{LMO} :

$$\begin{aligned} (C^{LMO})^\dagger SC^{LMO} &= U^\dagger C^\dagger SC^{LMO} = I \\ \implies U &= C^\dagger SC^{LMO} \end{aligned} \quad (6.18)$$

by multiplying on the left with U . The reason we include equation 6.18 is sometimes quantum chemistry packages only return C^{LMO} without U .

In summary, we get the following mapping:

$$\psi_i(\vec{r}) = \sum_{j=1}^K C_{ji} \phi_j(\vec{r}) \mapsto \psi_i^{LMO} = \sum_{j=1}^K [CU]_{ji} \phi_j(\vec{r}) = \sum_{j=1}^K C_{ji}^{LMO} \phi_j(\vec{r}). \quad (6.19)$$

Next, we show how the charge density remains unchanged.

For a closed shell molecule, described by a single determinant wavefunction, each MO ψ_i contains two electrons and thus the total charge density is [18]:

$$\begin{aligned} \rho(\vec{r}) &= 2 \sum_{i=1}^{\eta/2} \psi_i^*(\vec{r}) \psi_i(\vec{r}) \\ &= 2 \sum_{i=1}^{\eta/2} \left(\underbrace{\sum_{\nu=1}^K C_{\nu i}^* \phi_{\nu}^*(\vec{r})}_{\psi_i^*(\vec{r})} \underbrace{\sum_{\mu=1}^K C_{\mu i} \phi_{\mu}(\vec{r})}_{\psi_i(\vec{r})} \right) \\ &= \sum_{\mu=1}^K \sum_{\nu=1}^K \left[2 \sum_{i=1}^{\eta/2} C_{\mu i} C_{\nu i}^* \right] \phi_{\nu}^*(\vec{r}) \phi_{\mu}(\vec{r}) \\ &= \sum_{\mu=1}^K \sum_{\nu=1}^K \gamma_{\mu\nu} \phi_{\nu}^*(\vec{r}) \phi_{\mu}(\vec{r}). \end{aligned} \quad (6.20)$$

Here the square brackets define the density matrix $\gamma_{\mu\nu}$ (defined in the AO basis):

$$\gamma_{\mu\nu} = 2 \sum_{i=1}^{\eta/2} C_{\mu i} C_{\nu i}^*, \quad (6.21)$$

that for a set of basis function $\{\phi_j(\vec{r}) | j = 1, 2, \dots, K\}$ fully specifies the charge density $\rho(\vec{r})$ [18]. The sum runs over $\eta/2$, as these are the occupied MOs of a closed shell calculation. The whole matrix can be obtained as $\gamma^{total} = 2C_{occ}(C_{occ})^\dagger$, where *occ* denotes only using the occupied columns of the C matrix (the first $\eta/2$ columns, indexed by i in Equation 6.21). In the localized basis, the density matrix remains unchanged as:

$$\begin{aligned} \gamma^{total} &= 2C_{occ}(C_{occ})^\dagger \\ &= 2[C_{occ}^{LMO}U] [U^\dagger(C_{occ}^{LMO})^\dagger] \\ &= 2C_{occ}^{LMO}(C_{occ}^{LMO})^\dagger. \end{aligned} \quad (6.22)$$

Given a set of localised molecular orbitals, we partition them into two subsystems denoted *act* (active) and *env* (environment). There are different methods to do so and we discuss this later in Section 6.3. Overall we generate a set of (occupied) LMO indices \mathcal{K} and \mathcal{L} for the active and environment subsystems respectively. The

resulting charge density for each subsystem can then be written as:

$$\gamma_{\mu\nu}^{act} = 2 \sum_{k \in \mathcal{K}} C_{\mu k}^{LMO} (C_{\nu k}^{LMO})^\dagger, \quad (6.23a)$$

$$\gamma_{\mu\nu}^{env} = 2 \sum_{l \in \mathcal{L}} C_{\mu l}^{LMO} (C_{\nu l}^{LMO})^\dagger, \quad (6.23b)$$

for closed-shell calculations. The set $\mathcal{K} \cup \mathcal{L}$ contains all the occupied molecular orbital indices.

The total system electron density is written as a sum of subsystem densities:

$$\begin{aligned} \gamma^{total} &= \gamma^{act} + \gamma^{env} \\ &= 2C_{\mathcal{K}}^{LMO} (C_{\mathcal{K}}^{LMO})^\dagger + 2C_{\mathcal{L}}^{LMO} (C_{\mathcal{L}}^{LMO})^\dagger \\ &= 2C_{occ}^{LMO} (C_{occ}^{LMO})^\dagger. \end{aligned} \quad (6.24)$$

The number of electrons will also be split according to $n_e^{total} = n_e^{act} + n_e^{env} = \text{Tr}(S\gamma^{act}) + \text{Tr}(S\gamma^{env}) = \text{Tr}(S\gamma^{total})$, where Tr denotes the trace operation.

The energy of the full system can be found from its components via [308]:

$$\begin{aligned} E[\gamma^{act}, \gamma^{env}] &= \underbrace{\text{Tr}(\gamma^{act} h_{core}) + g(\gamma^{act})}_{\text{energy of isolated } act \text{ system}} + \\ &\quad \underbrace{\text{Tr}(\gamma^{env} h_{core}) + g(\gamma^{env})}_{\text{energy of isolated } env \text{ system}} + \\ &\quad \underbrace{g(\gamma^{act}, \gamma^{env})}_{\text{nonadditive two-electron energy}}. \end{aligned} \quad (6.25)$$

Here h_{core} is the one-electron core Hamiltonian and g groups the two-electron terms - Coulomb and exchange for Hartree-Fock and exchange-correlation for DFT. The nonadditive two-electron energy is given by:

$$g(\gamma^{act}, \gamma^{env}) = g(\gamma^{act} + \gamma^{env}) - g(\gamma^{act}) - g(\gamma^{env}), \quad (6.26)$$

and accounts for the interaction between subsystems [308].

6.2.3 Projection

Next, we want to solve the active system using a higher (more accurate) level of theory. The effect of the interaction between the active and environment subsystems is accounted for by additional terms in the core Hamiltonian. The Fock matrix for

the active system embedded in the environment system is [286]:

$$\begin{aligned} F_{emb}^{act} &= h_{core} + V_{emb} + P_{proj}^{env} + g(\gamma_{emb}^{act}) \\ &= h_{emb} + g(\gamma_{emb}^{act}), \end{aligned} \quad (6.27)$$

where:

$$\begin{aligned} V_{emb} &= g(\gamma^{act} + \gamma^{env}) - g(\gamma^{act}) \\ &= \left(V_{xc}[\gamma^{act} + \gamma^{env}] - V_{xc}[\gamma^{act}] \right) + \left(J[\gamma^{act} + \gamma^{env}] - J[\gamma^{act}] \right) \end{aligned} \quad (6.28a)$$

$$h_{emb} = h_{core} + V_{emb} + P_{proj}^{env}. \quad (6.28b)$$

Here the matrices J , V_{xc} and h_{core} are obtained according to equations 2.113b, 2.113c and 2.113a respectively. Note the definition of V_{emb} is consistent with the previous section (see equation 6.8, with $v_s^{nadd} = 0$). The embedding potential V_{emb} describes all the interactions (nonadditive part) between the active and environment subsystems [309]. Due to the subsystem densities (Equation 6.24) being constructed from disjoint subsets of orthogonal orbitals, the normally difficult-to-evaluate nonadditive kinetic potential (NAKP) terms [298] are exactly zero [286, 309, 310].

P_{proj}^{env} is a projection operator that enforces inter-subsystem (orbital) orthogonality. There are different ways to define this operator and we consider two in this work. The first definition was proposed by the Manby and Miller groups [286]. They use a parameter μ to shift the orbital energies of the environment to high energies - effectively meaning they will never be occupied. This projector is defined as:

$$\begin{aligned} (P_{\mu}^{env})_{ij} &= \mu \langle \psi_i^{LMO} | P^{env} | \psi_j^{LMO} \rangle \\ &= \mu [S \gamma^{env} S]_{ij}, \end{aligned} \quad (6.29)$$

where μ is some large integer, S is the AO overlap matrix. P^{env} is a projector defined as:

$$P^{env} = \sum_{l \in \mathcal{L}} |\psi_l^{LMO}\rangle \langle \psi_l^{LMO}|. \quad (6.30)$$

Here we use the notation $l \in \mathcal{L}$ to mean the sum over the set of occupied MO indices for the environment orbitals. The work in [286, 308] shows μ is numerically robust and can usually be set to $\mu = 10^6$. In the limit that $\mu \rightarrow \infty$ this method is exact, as the environment orbitals will be pushed to infinite energy and thus will never be occupied. The action of this operator with the Fock matrix is:

$$(F + P_{\mu}^{env}) |\psi_k^{LMO}\rangle = \epsilon_k^{act} |\psi_k^{LMO}\rangle, \quad (6.31a)$$

$$(F + P_\mu^{env}) |\psi_l^{LMO}\rangle = (\epsilon_l^{env} + \mu) |\psi_l^{LMO}\rangle \approx +\mu |\psi_l^{LMO}\rangle. \quad (6.31b)$$

Again, k and l represent occupied LMOs of the active and environment subsystems respectively. Qualitatively the orbital energies of the active system are left unchanged and the orbitals for the environment are pushed to very high energies as $\mu \gg \epsilon_i^{env}$ - effectively suppressing transitions to these states and stopping hybridisation.

The second approach, proposed by Kallay *et al.* [311], is to use the Huzinaga projector [312, 313]:

$$\begin{aligned} P_{huz}^{env} &= -(F P^{env} + P^{env} F) \\ &= -\frac{1}{2} (F \gamma^{env} S + S \gamma^{env} F). \end{aligned} \quad (6.32)$$

Note that the $-\frac{1}{2}$ prefactor is needed for closed-shell systems. This operator enforces orthogonality of the occupied orbitals of each subsystem [314]. The form of this operator increases the orbital energy for the occupied environment orbitals and leaves the active system unchanged. We write its action formally as:

$$(F + P_{huz}^{env}) |\psi_k^{LMO}\rangle = \epsilon_k^{act} |\psi_k^{LMO}\rangle, \quad (6.33a)$$

$$\begin{aligned} (F + P_{huz}^{env}) |\psi_l^{LMO}\rangle &= (\epsilon_l^{env} - 2\epsilon_l^{env}) |\psi_l^{LMO}\rangle \\ &= -1\epsilon_l^{env} |\psi_l^{LMO}\rangle. \end{aligned} \quad (6.33b)$$

It is common for occupied orbitals to have negative energy and so equation 6.33b shows how the environment ϵ_l^{env} orbitals are usually shifted to positive energies (assuming negative ϵ_l^{env}) and thus will not be filled. If any occupied ϵ_l^{env} are positive, then equation 6.32 can be modified to treat such systems correctly- this operator is known as the ‘‘Fermi-shifted Huzinaga operator’’ [315]. This formalism guarantees that $[P_{huz}^{env}, F_{emb}^{act}] = 0$ and removes the need for the μ parameter shift [316].

The energy of the active system embedded in the environment is given by:

$$\begin{aligned} E[\gamma_{emb}^{act}; \gamma^{act}, \gamma^{env}] &= \mathcal{E}[\gamma_{emb}^{act}] + E[\gamma_{env}] + g(\gamma^{act}, \gamma^{env}) \\ &\quad + \text{Tr} \left((\gamma_{emb}^{act} - \gamma^{act})(V_{emb} + P_{proj}^{env}) \right), \end{aligned} \quad (6.34)$$

colloquially denoted as a DFT-in-DFT calculation. We use the same notation as Claudino and coworkers [308], where \mathcal{E} differs from E as it allows for different functionals to be applied and is computed from the embedded density matrix of the active system. Note that Equation 6.27 is solved self-consistently to give γ_{emb}^{act} . Equation 6.34 reduces to Equation 6.25 for the case that the active and environment regions are treated at the same level of theory [308]. Importantly, $\mathcal{E}[\gamma_{emb}^{act}] = \text{Tr}(\gamma_{emb}^{act} h_{core}) + g(\gamma_{emb}^{act})$ does not involve V_{emb} or P_{proj}^{env} . The final term in Equation 6.34 is a first-order correction

that accounts for the difference between $g(\gamma^{act}, \gamma^{env})$ and $g(\gamma_{emb}^{act}, \gamma^{env})$, and corrects for the fact that in general $\gamma^{act} \neq \gamma_{emb}^{act}$ [317].

This projection-based embedding approach then allows for the active system to be treated using some wavefunction level of theory and therefore can be studied using a quantum computer. The electronic energy for this is given by [308]:

$$E[\Psi_{emb}^{act}; \gamma^{act}, \gamma^{env}] = \langle \Psi_{emb}^{act} | H_{emb} | \Psi_{emb}^{act} \rangle + E[\gamma^{env}] + g(\gamma^{act}, \gamma^{env}) - \text{Tr} \left(\gamma^{act} (V_{emb} + P_{proj}^{env}) \right). \quad (6.35)$$

Importantly $H_{emb} = h_{emb} + g(\Psi_{emb}^{act})$, where $g(\Psi_{emb}^{act})$ is the two-electron operator for a given wavefunction method and h_{emb} is the embedded core Hamiltonian (Equation 6.28b) which depends on γ^{act} and γ^{env} [318]. As the embedding terms have been included in H_{emb} , the final correction term is therefore slightly different to Equation 6.34 [317]. The wavefunction calculation in Equation 6.35 includes contributions from $(V_{emb} + P_{proj}^{env})$ - similar to: $\text{Tr}(\gamma_{emb}^{act} (V_{emb} + P_{proj}^{env}))$. The correction therefore only requires subtracting $\text{Tr}(\gamma^{act} (V_{emb} + P_{proj}^{env}))$, unlike in Equation 6.34, where \mathcal{E} does not use $(V_{emb} + P_{proj}^{env})$ to calculate the energy of the active system.

6.3 Molecular orbital localization

In this Section, we investigate different molecular orbital localization schemes. We focus on five: The Foster-Boys (FB) [319], Edmiston-Ruedenberg (ER) [320], Pipek-Mezey (PM) [321], intrinsic bonding orbitals (IBO) [322] and subsystem Projected AO DEcomposition (SPADE) approach [308]. Except for the SPADE method, in each subsection we present the cost function $L(U)$ that is either maximized or minimized in order to find the optimal unitary transformation matrix U (equation 6.16). We do not discuss how U is updated in each optimization step such that it remains a unitary operator. Further details on this can be found in [323–327]. Each cost function gives a measure of the locality of the molecular orbitals. These differ because there is ambiguity regarding the choice of this function, hence the different localization methods.

6.3.1 Foster-Boys

The Foster-Boys (FB) localization scheme minimizes the spread of an orbital, by minimizing the square of the distance separating two electrons $r_{12}^2 = |r_1 - r_2|^2$ [319]. The set of LMOs is obtained by minimizing [240]:

$$L_{FB}(U) = \left[\sum_i^K \int \int |\psi_i^{LMO}(\vec{r}_1)|^2 r_{12}^2 |\psi_i^{LMO}(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2 \right], \quad (6.36)$$

where each ψ_i^{LMO} is given by equation 6.19 and is where the unitary U appears. It can be shown this is equivalent to maximizing [30]:

$$L_{FB}(U) = \left[\sum_i^K (\langle \psi_i^{LMO} | r | \psi_i^{LMO} \rangle)^2 \right]. \quad (6.37)$$

Currently this formalism still requires a description in molecular orbitals, not atomic orbitals. In chapter nine of [30], it is shown that the dipole integrals in the molecular basis may be obtained from the corresponding AO integrals:

$$\langle \psi_i | r | \psi_j \rangle = \sum_\alpha^M C_{\alpha i} \left(\sum_\beta^M C_{\beta j} \langle \phi_i | r | \phi_j \rangle \right). \quad (6.38)$$

The overall computational cost of this localization therefore scales as $\mathcal{O}(K^3)$ multiplied by the number of times the cost function is called [30].

6.3.2 Edmiston-Ruedenberg

The Edmiston-Ruedenberg localization scheme minimizes the inverse distance between two electrons $1/r_{12} = 1/|r_1 - r_2|$ [320], which is proportional to the two-body electronic repulsion operator. The LMOs are obtained by maximising [240]:

$$L_{ER}(U) = \left[\sum_i^K \int \int |\psi_i^{LMO}(\vec{r}_1)|^2 \frac{1}{r_{12}} |\psi_i^{LMO}(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2 \right], \quad (6.39)$$

where each ψ_i^{LMO} is given by equation 6.19 and is where the unitary U appears. Again, as with the FB scheme, currently this formalism requires a description in molecular orbitals, not atomic orbitals. Unlike before, the trick of using the AO dipole integrals to obtain the MO integrals is not possible. The cost of this scheme therefore currently scales as $\mathcal{O}(K^5)$ [30], due to the atomic-to-molecular-orbital transformation required (see Section 2.7). Work by Head-Gordon *et al.* showed that this scaling can be reduced to $\mathcal{O}(K^3)$ [30, 328], an analysis of this resource reduction is outside the scope of this thesis. The overall cost will be this scaling multiplied by the number of times the cost function is called [240].

6.3.3 Pipek-Mezey

The Pipek-Mezey scheme involves maximizing the Mulliken charge of each orbital [321]. This is done by maximising [240]:

$$L_{PM}(U) = \left[\sum_J^M \sum_i^K (Q_J^i)^2 \right], \quad (6.40)$$

where

$$Q_J^i = \sum_{\mu \in J} \sum_{\nu}^K C_{\mu i}^{LMO} S_{\mu\nu} C_{\nu i}^{LMO}. \quad (6.41)$$

Here M is the number atoms, Q_J^i is the contribution of orbital i to the Mulliken charge of atom J and each ψ_i^{LMO} is given by equation 6.19. Equation 6.19 gives the definition of C^{LMO} and is where the unitary U appears. The scaling of PM is $\mathcal{O}(K^3)$ multiplied by the number of times the cost function is called [240].

6.3.4 Intrinsic bonding orbitals

Intrinsic Bonding Orbitals (IBOs) are constructed as a linear combination of intrinsic atomic orbitals (IAOs), where the number of atoms the orbital charge is spread over is minimized. The localization performed is similar to the Pipek-Mezey approach.

Taking the canonical orbitals defined as:

$$|\psi_i\rangle = \sum_{j=1}^K C_{ji} |\phi_j\rangle, \quad (6.42)$$

where $|\phi_j\rangle \in B_1$ are basis functions from a basis set B_1 . A key problem with this form is each MO $|\psi_i\rangle$ is hard to interpret, as each atomic orbital's (AO) basis functions $|\phi_j\rangle$ cannot be associated with a given atom [322]. Normally MOs are highly delocalized and each $|\phi_j\rangle$ will contribute where it is needed most. Thus, we would like to expand the MOs over another minimal basis B_2 of free-atom atomic orbitals for each atom, which we write as $\{|\tilde{\phi}\rangle\} \in B_2$. This would make the wavefunction easy to interpret, but would be inaccurate and even incorrect, as free-atom AOs contain no polarization due to the molecular environment. The IAO method aims to combine the best properties of these.

First, a molecular SCF wavefunction $|\Phi\rangle$ is calculated that defines a set of molecular orbitals $|\psi_i\rangle$. Then a set of polarized atomic orbitals $\{|\phi^{(IAO)}\rangle\} \notin B_2$ are formed that can express $|\Phi\rangle$ s in the occupied MOs $|\psi_i\rangle$. To do this, projectors onto the occupied and virtual MOs are defined: $P = \sum_i |\psi_i\rangle \langle \psi_i|$ and $Q = I - P$ respectively [329]. The AO projectors onto the bases B_1 and B_2 are also defined as [322]:

$$P_{12} = \sum_{|\phi\rangle_i, |\phi\rangle_j \in B_1} |\phi\rangle_i S_{ij}^{-1} \langle \phi|_j, \quad (6.43a)$$

$$P_{21} = \sum_{|\tilde{\phi}\rangle_k, |\tilde{\phi}\rangle_l \in B_2} |\tilde{\phi}\rangle_k \tilde{S}_{kl}^{-1} \langle \tilde{\phi}|_l. \quad (6.43b)$$

Here S_{ij}^{-1} and \tilde{S}_{kl}^{-1} are the inverse overlap matrices in the bases B_1 and B_2 . The set

of depolarized MOs is then given by [322, 329]:

$$\{|\tilde{\psi}_i\rangle\} = \{P_{12}P_{21}|\psi_i\rangle\}. \quad (6.44)$$

These depolarized molecular orbitals are then used to define the projectors: $\tilde{P} = \sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|$ and $\tilde{Q} = I - \tilde{P}$. The IAOs are finally given by:

$$|\phi^{(IAO)}\rangle = (P\tilde{P} + Q\tilde{Q})P_{12}|\tilde{\phi}\rangle. \quad (6.45)$$

The IAO basis is then orthonormalized [322, 329]. Overall, each IAO is obtained by a simple set of matrix operations. The utility of IAOs stems from the fact that they are directly associated with atoms and can be used to define atomic properties like partial charges and are basis set independent, unlike the Pipek-Mezey approach.

Often we want to know about molecular bonding rather than atomic properties. Knizia showed that by combining IAOs with a localization in the spirit of Pipek-Mezey, one can obtain intrinsic bonding orbitals (IBOs). The IBOs are found by maximizing [322]:

$$L_{IBO}(U) = \left[\sum_J^M \sum_i^\eta [n_J(i)]^4 \right]. \quad (6.46)$$

Here $n_J(i)$ is the number of electrons from ψ_i^{LMO} (equation 6.19) located on the IAOs $\{|\phi^{(IAO)}\rangle\}$ of atom J . Explicitly, we write this as [322]:

$$n_J(i) = 2 \sum_{|\phi^{(IAO)}\rangle \in J} \langle \psi_i^{LMO} | \phi^{(IAO)} \rangle \langle \phi^{(IAO)} | \psi_i^{LMO} \rangle. \quad (6.47)$$

This construction minimizes the number of atoms upon which an orbital is centred [322].

6.3.5 Subsystem projected AO decomposition

The Subsystem projected AO decomposition (SPADE) involves first having defined occupied MOs $\{\psi_i\}$ in terms of the AO basis (see equation 2.57) obtained from an SCF calculation. To start, Löwdin's method of symmetric orthogonalization is used to generate orthogonal MOs [330, 331]. This is performed as:

$$\tilde{C} = S^{1/2}C, \quad (6.48)$$

where S is the AO overlap matrix and C is the matrix of orbital coefficients (equation 6.13) [308]. This generates a set of orthogonal MOs that have a shape closest to the original AOs in the least square sense [240, 332]. In this basis, the projector Q_A is

defined [308]:

$$Q_A = \sum_{\tilde{\phi} \in A} |\tilde{\phi}_i\rangle \langle \tilde{\phi}_i|, \quad (6.49)$$

where $\{\tilde{\phi}\}$ are the set of orthogonal AOs associated with atoms in subsystem A . Next, if we let \tilde{C}_{occ}^A be the MO coefficients from AOs from the atoms in subsystem A , we have [308]:

$$Q_A = \sum_{\tilde{\phi} \in A} |\tilde{\phi}_i\rangle \langle \tilde{\phi}_i| \implies Q_A \tilde{C}_{occ} = \tilde{C}_{occ}^A, \quad (6.50)$$

where *occ* denotes that only the occupied orbitals are considered.

New orbitals can be obtained by rotating the original MOs (C_{occ}) according to the right-singular vectors of \tilde{C}_{occ}^A , where the corresponding singular values signal their relative importance of the associated vectors. Taking the singular value decomposition (SVD) of \tilde{C}_{occ}^A given by:

$$\tilde{C}_{occ}^A = U^A \Sigma^A (V^A)^\dagger. \quad (6.51)$$

The column vectors of V^A are identical to the eigenvectors of $(\tilde{C}_{occ}^A)^\dagger \tilde{C}_{occ}^A$ [308]. A set of localized MOs can be obtained by rotating the original MOs given by C_{occ} . This procedure is SPADE, which is implemented as [308]:

$$C_{occ}^{SPADE} = C_{occ} V^A. \quad (6.52)$$

Using the appropriate atoms from each subsystem, A and B , the resulting orbital subspaces emerge from the distribution of singular values $\{\sigma\}$. Taking the difference of successive singular values gives and finding the largest difference gives the most adequate partition of the set of MOs. We write this as:

$$i_{max} = \operatorname{argmax}_i [\Delta\sigma_i] = \operatorname{argmax}_i [\sigma_i - \sigma_{i+1}], \quad (6.53)$$

where the index i_{max} represents the largest singular value difference. The indices from $[0, i_{max}]$ define the occupied orbital indices for subsystem A and $(i_{max}, \eta - 1]$ for subsystem B ¹. Here η is the number of electrons, aka the indices go over occupied orbitals.

There are two main benefits of the SPADE approach for PBE. First, the singular values are similar to occupation numbers. This is due to the similarity with a natural population analysis, which has been shown to avoid some of the flaws of the Mul-

¹Closed intervals using square brackets include endpoints. Parentheses refer to intervals that do not include endpoints.

liken analysis [308, 333]. Secondly, the fact that \tilde{C}_{occ}^A (equation 6.50) only contains coefficients from atoms in subsystem A , the corresponding rotated orbitals C_{occ}^{SPADE} (equation 6.52) will mostly be constrained to their corresponding subsystems [308]. This differs from most localization schemes that attempt to localize orbitals in atoms or bonds, whereas SPADE orbitals are local only in the sense that they remain in their native subsystems which is one of the requirements for successful embedding [308].

6.3.6 Partitioning into subsystems

To assign these localized molecular orbitals to the active and environment subsystems we calculate the percentage of the i -th LMO over the active atoms as:

$$p_i^A(C) = \frac{\sum_{\nu \in A} C_{i\nu}^2}{\sum_{\nu=1}^K C_{i\nu}^2}, \quad (6.54)$$

where $\nu \in A$ are the atomic orbital indices for the atoms defined in the active region. The denominator includes all the AOs of the i -th MO. This is the approach given in Equation 10 in the work of Koch *et al.*[334]. Any $p_i^A > 95\%$ we associate to the active subsystem. Our open source implementation of PBE uses this metric [335], where any C matrix (preferably localized) can be used. The codebase can run any localization strategy supplied by PySCF and the SPADE approach. Users can also define their own localization schemes if desired.

6.4 Numerical study

In this Section, we present our numerical analysis of the PBE model used to generate embedded qubit Hamiltonians. These results can be found in the preprint [336].

6.4.1 Method

We studied the performance of our wavefunction projection-based embedding method on a selected set of molecular systems. We have developed a python package, Nbed [335], that utilizes the PySCF and Openfermion quantum chemistry packages to build each embedded model [222–224]. The package outputs a qubit Hamiltonian for the wavefunction portion of an embedded problem and the classical energy corrections from density functional theory. This is freely available for use on GitHub [335].

For all calculations presented, the minimal *STO-3G* basis set was employed. Each global DFT calculation performed, prior to orbital localisation, used the *B3LYP* functional. The Intrinsic Bonding Orbitals (IBO) or Subsystem Projected AO Decomposition (SPADE) localisation procedures are used to isolate the molecular or-

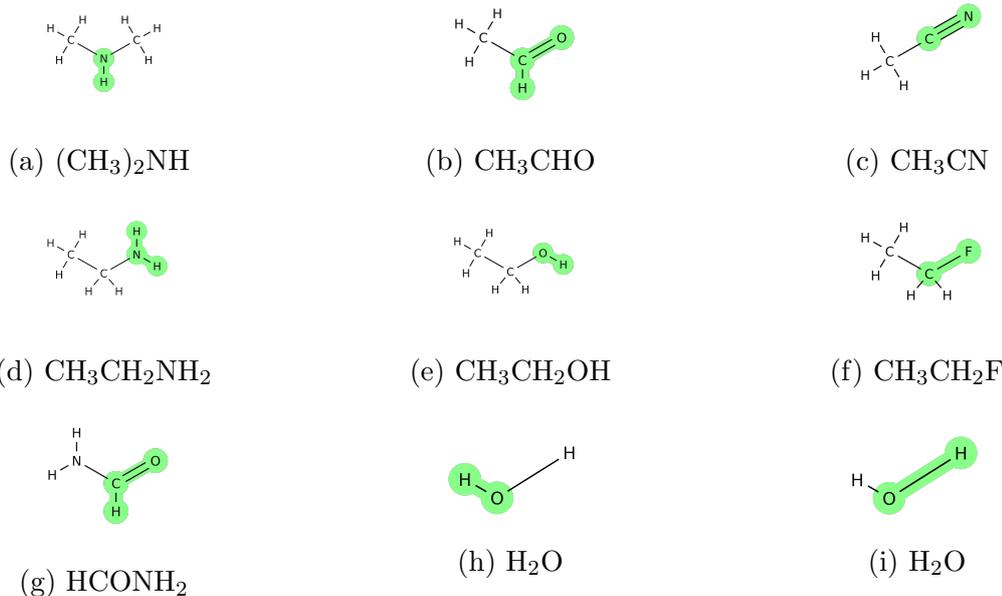


Figure 6.1: Planar representations of the molecules used in embedding calculations. Atoms shaded in green were selected as active for localisation procedures. Images were generated using MolView [337]. 6.1a N-methylmethanamine; 6.1b acetaldehyde; 6.1c acetonitrile; 6.1d ethanamine; 6.1e ethanol; 6.1f flouroethane; 6.1g formamide; 6.1h water (fixed bond active); 6.1i water (stretching bond active)

bitals to the active and environment subsystem from pre-selected active atoms [308, 322]. For the IBO procedure, in order to assign the active and environment molecular orbitals, we calculate the percentage of the i -th LMO over atoms a user defines as the active subsystem. Any LMO that has a percentage higher than a predefined threshold (we used 95 %) is assigned to the active region (see equation 6.54). We performed both the μ -shift and Huzinaga methods for each. A Hartree-Fock calculation for the active system, using the modified core Hamiltonian (equation 6.28b), was performed for each molecular system. The second quantized molecular Hamiltonian was then constructed with Openfermion and converted to a qubit Hamiltonian using the in-built Jordan-Wigner transformation [123]. Post-Hartree-Fock methods were performed with PySCF. The frozen core approximation was not used and all virtual orbitals were included in the wavefunction calculations. Only the occupied environment molecular orbitals were removed from the wavefunction calculations of the active systems. To achieve this, the columns of C (the matrix of MO coefficients) associated with the environment were masked and not considered by further post-Hartree-Fock (HF) methods on the embedded active system (the WF part of a WF-in-DFT calculation). The removal of these (occupied) environment orbitals is what gives a qubit reduction (when constructing the second quantized molecular Hamiltonian). We note that this removal approach is slightly different to the implementation of Goodpaster and coworkers [316], where these orbitals are left in the embedded calculation. Our

approach is justified as the environment MOs have been projected out of the SCF problem. More qualitatively, in Goodpaster’s approach the results of the Huzinaga and μ -shifted approaches are very similar [316]. For the μ -shifted approach, the occupied environment orbitals are shifted to such high energies that they remain unoccupied in subsequent post-HF calculations. This is not the case for the Huzinaga method but, as it gives similar results to the μ -shifted technique, removing the associated (occupied environment) orbitals of the Huzinaga method follows inline with them not being able to be occupied in the μ -shifted approach.

For the single point electronic structure calculations we perform a CCSD-in-DFT calculation (active subsystem treated at CCSD level). Each result is then compared to full system CCSD(T) calculations. Each molecular geometry was obtained from PubChem [338]. The potential energy surface of an OH bond stretching in water was calculated using FCI-in-DFT, where the embedded molecular Hamiltonian at each geometry was diagonalized to find the ground state energy of the active system. This was compared to a full system FCI calculation at each step. As the PBE model requires a full system DFT calculation to determine the active and environment subsystems, we also report these DFT results.

6.4.2 Results and discussion

In the following subsections, we apply our model to different molecular systems.

6.4.2.1 Molecular ground state energy

In order to assess the ability of the embedding procedure, we selected a test bed of molecular structures - which are summarised in Figure 6.1. The active atoms considered at a more expensive level of theory are highlighted in green. The choice of these molecules was motivated by selecting compounds commonly encountered by chemists. To date, most quantum computing studies consider only the smallest molecular systems (often H_2 , LiH , BeH_2) [36, 94, 103], due to current quantum computing constraints - low numbers of qubits numbers and high error rates. The goal of this study is to show this embedding approach will allow larger systems to be studied on such devices. Figure 6.2 reports the results for SPADE localized CCSD-in-DFT embedding molecular ground state energy calculations for the molecules in Figure 6.1. The numerical values of these results are available in Appendix B. The results for the same calculations using IBO localized orbitals can also be found there.

The embedded second quantized Hamiltonians (describing the active region) output using both localisation methods were significantly reduced in the number of Pauli operators and qubit counts compared to the full system second quantized Hamiltonian. However, they still exceed the limit of what is practical to exactly solve using

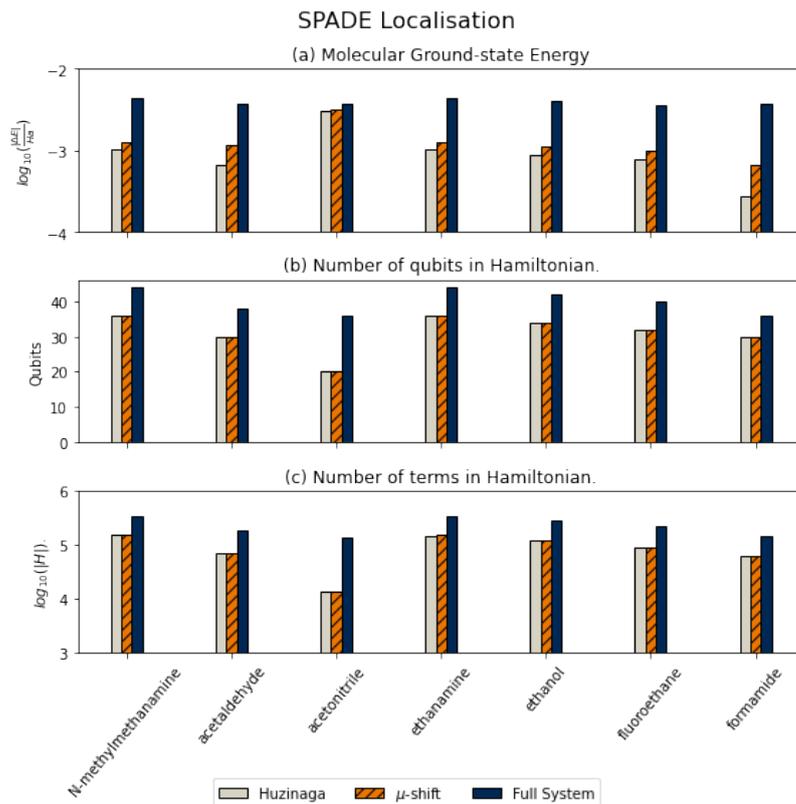


Figure 6.2: Results for WF-in-DFT calculation of small molecules (Figure 6.1), where the SPADE localisation method has been used. Bar chart (a) reports the ground state energy error for small molecules compared to full system CCSD(T): $|\Delta E| = |E_{exp} - E_{CCSD(T)}|$. For the blue result E_{exp} is the full-system DFT (*B3LYP*) ground state energy, for the orange result E_{exp} gives the μ -shift CCSD-in-DFT embedding energy and for the grey result E_{exp} is the Huzinaga CCSD-in-DFT embedding energy. Plot (b) shows the number of qubits needed to describe the embedded Hamiltonian, with the reference showing the number required for the full system second quantized Hamiltonian. Plot (c) reports the number of terms in the Jordan-Wigner encoded embedded FCI-in-DFT Hamiltonian for each molecule with the blue bar representing the number of terms in the second quantized Hamiltonian of the full system.

classical computers. We therefore performed classical CCSD-in-DFT calculations, the results of which are given in Figure 6.1(a). Our results show increased accuracy in CCSD-in-DFT calculated molecular ground state energies, compared to full system DFT. We benchmarked these approaches compared to full system CCSD(T), as full system FCI was not possible. The following metric was used $|\Delta E| = |E_{exp} - E_{CCSD(T)}|$ to approximate the true error $|\Delta E_{true}| = |E_{exp} - E_{FCI}|$, where E_{exp} is the ground state energy calculated via different procedures as specified and $E_{CCSD(T)}$ is a full system CCSD(T) reference ground state energy that is used to approximate each (full system) FCI ground state energy E_{FCI} . Our results show that PBE embedding gives ground state energies closer to the reference value, namely the full system CCSD(T) energy. It is known that different DFT functionals will give different ground state

densities and thus energies [339]; however, there will always be a true ground state wavefunction and energy. The results in Figure 6.1 show that we can improve on the results of a DFT calculation by getting closer to the true ground state energy - which in our case we approximate with CCSD(T). Qualitatively, we attribute the improvement of our PBE method to it including different correlation effects not captured by the DFT calculation. Further evidence of this will be seen in the next section on strong correlation.

Typically, results for the μ -shift and Huzinaga projectors are very similar. This is expected as the active and environment subsystems were described in the supermolecular basis in our implementation of PBE [311, 316]. If the subsystems were described in an alternate basis, it has been shown that the Huzinaga operator outperforms the μ -shift approach [316]. In our results, the Huzinaga projector usually produces marginally more accurate energies compared to the μ -shifted implementation. We attribute this to the Huzinaga approach being based on a formally exact embedding, while the μ -shift embedding is approximate due to a finite shift value being used.

The number of qubits describing the embedded second quantized Hamiltonian (of an FCI-in-DFT problem) will be the same between the embedding methods - Figure 6.1b. This is due to the number of qubits depending on how many spin orbitals are considered in the embedded active system. In this work, only the occupied molecular orbitals of the environment are removed from the embedded active subsystem WF calculations (which leads to a qubit reduction). To further reduce the qubit count, virtual (unoccupied) molecular orbitals should also be removed from the embedded active subsystem calculations. This requires partitioning of the virtual space between the active and environment subsystems. Recent work by Yuan and coworkers showed that truncating the virtual space can still give reliable estimates of both energies and molecular properties and we anticipate that this result will also be found if virtual environment orbitals are removed from the active embedded subsystem [340]. We leave this to future work, but note that it could lead to a significant further reduction in the number of qubits.

The number of terms in the Jordan-Wigner encoded qubit second quantized Hamiltonian of the embedded WF problem (FCI-in-DFT problem), $|H|$, is typically very similar between the two projection methods, as shown in Figure 6.1c. This is expected as the number of molecular orbitals used to describe the embedded problem is the same between the different projection methods.

In comparing the two localisation methods, we find that for acetonitrile and formamide, SPADE and IBO partition the active system in a similar way. This results in a similar number of active MOs and hence the ground state estimation and resource requirements are very similar for these systems. For the majority of the molecules we study, SPADE includes more MOs, resulting in significantly more accurate ground

state energies while still reducing the size of the Hamiltonian. However, by reducing the threshold for assigning the localized MOs from IBO to the active region, additional MOs could be included giving a similar result to SPADE. See Appendix B for further details.

6.4.2.2 Strong correlation

The impact of active region selection is demonstrated by our results shown in Figure 6.3. We consider the bond dissociation of an OH bond in water, where at high bond lengths, a correlated state is created [114, 341]. We perform projection-based FCI-in-DFT calculations, at different molecular geometries, for two different active regions. One has the atoms in the fixed OH bond set active and the other has the atoms in the changing OH bond set active. We show this pictorially in Fig. 6.1h and 6.1i. These results are compared to full system FCI calculations.

For the FCI-in-DFT results given in Figure 6.3b, we note that the error in the embedded calculation is actually higher than the global DFT calculation at an OH bond length of 4.0 Å. There is also a significant difference in the number of qubits and Hamiltonian terms at OH bond lengths of 0.4, 0.6 and 4.0 Å. The variation, compared with the SPADE results (Figure 6.3a), is mainly due to the localization method giving different numbers of active MOs. We repeated the IBO calculation using a different active region threshold (using a minimum setting of 90%) and obtained similar results, as the number of active MOs then matched that of the SPADE calculation. Figure B.2 in Appendix B summarises this result.

At near equilibrium bond lengths, we observe in Figure 6.3 a similar performance between the different active systems. This is due to the symmetrical structure of H₂O, hence at low bond lengths there is little difference between the two active regions. In fact, the third data point gives results for the scenario where both OH bonds are the same length and consequently is why the results for the different active regions are the same here. However, in the correlated regime - at large bond lengths - selecting the active region to encompass the stretched atoms leads to significant improvements in energy calculation over DFT alone. This is due to the correlation being effectively captured in the wavefunction calculation. In contrast, the full DFT calculation is plagued by deficiencies in current approximate exchange-correlation functionals [72, 342]. We see in Figure 6.3 that the global DFT calculation overestimates the bond dissociation energy. This problem is attributed to static correlation [72]. As there is no systematic way to improve the approximate exchange-correlation functionals, the way forward in describing such systems is hybrid quantum-classical embedding. Here quantum processors could be exploited most effectively by application to only those regions of a molecule that are highly correlated.

6.4. NUMERICAL STUDY

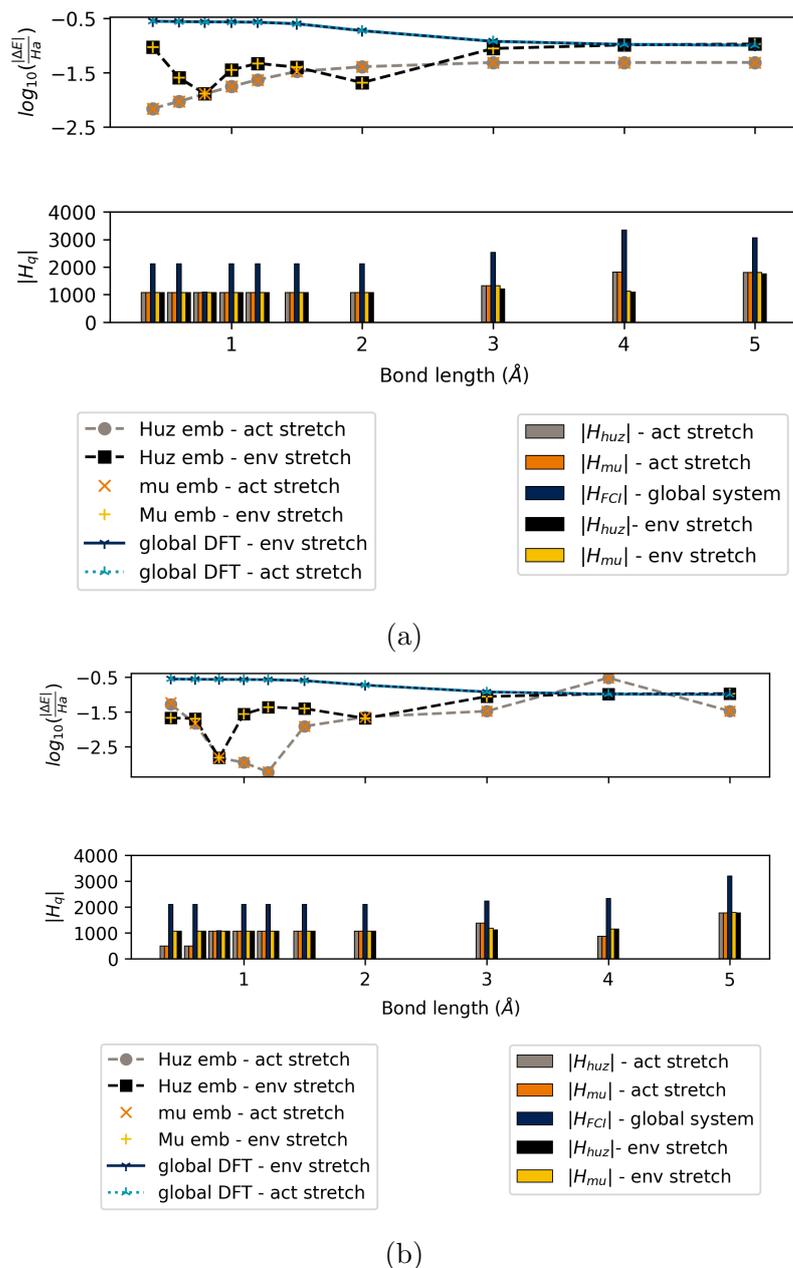


Figure 6.3: Potential energy curve for H_2O , with changing OH bond length. *Active stretch* result has the changing OH bond as the active region and *environment stretch* result has the fixed OH bond selected as the active region. Figure 6.3a uses SPADE localization and Figure 6.3b uses IBO localization (**95%** threshold). For each data set the full problem is reduced from 14 to 12 qubits, with the number of active MOs being four in all cases. In each subfigure, the top plot reports the log base 10 error with respect to the exact FCI ground state energy (E_{FCI}) of the whole system, where $|\Delta E| = |E_{exp} - E_{FCI}|$. Here E_{exp} is obtained from an FCI-in-DFT calculation. The bottom plot reports the number of terms in each Jordan-Wigner encoded qubit Hamiltonian. The blue result gives the size of the full system Hamiltonian, the orange and yellow results are for μ -shifted embedded Hamiltonians while the grey and black results are for the Huzinaga embedded Hamiltonians. Numerical details are available in Appendix B

6.5 Conclusion

For a select set of molecules, too large to study completely (full system) on currently available quantum hardware, we have shown that the PBE method allows a smaller active system to be studied using fewer resources on a quantum computer and the calculated energies of such embedding calculations are closer to the “gold standard” CCSD(T) of the full system compared to full system DFT. Furthermore, we have shown its ability to capture the effects of strong correlation by investigating the bond dissociation of H₂O.

We use the projection-based embedding technique [286] to reduce the size of an electronic structure calculation studied at the wavefunction level. The molecular problem is split into active and environment parts, each solved using different levels of theory. The active part is treated using a wavefunction approach and an embedded qubit Hamiltonian is generated. Solving this provides $E_{act}^{WF} = \langle \Psi_{emb}^{act} | H_{emb} | \Psi_{emb}^{act} \rangle$. The whole system and environment are treated using density functional theory and the overall electronic energy is found via an additive procedure [316–318]. What is included in the active region can be modified and thus the size of the quantum problem varied. This allows users to tune their problem to available hardware and so it is possible to simulate large molecular problems on small quantum devices.

As this approach generates an embedded qubit Hamiltonian, it is agnostic to the quantum algorithm used to solve H_{emb} . NISQ friendly approaches such as the VQE algorithm can therefore be used, but also fault-tolerant methods such as quantum phase estimation (QPE) [89, 94].

Moreover, as our method outputs a qubit Hamiltonian, different resource reduction techniques can be used in conjunction with it; for example, the contextual-subspace approach of Kirby and coworkers, [244] or the entanglement forging approach of Ed-dins [114]. Similarly, the \mathbb{Z}_2 -symmetries of the problem can also be removed via qubit tapering [124].

As our method does not rely on imposing constraints on the system studied or costly parameter fitting, it may be reasonably combined with other hybridisation techniques which do [343, 344].

We note a few avenues for future work. (1) As significant resource reduction is achieved by localisation of only the occupied orbitals, virtual orbital localisation could lead to a greater reduction in computational resources [308]. In the context of this work, if virtual LMOs are included in the active and environment subsystems respectively, then the number of qubits will reduce by how many are included in the environment. This will also have the effect of decreasing the total number of Pauli operators in the associated embedded qubit Hamiltonian. (2) A minor defect of this method is that the subsystems do not interact in a given calculation. It would be

6.5. CONCLUSION

interesting to see if the Locally Coupled Open Subsystem (LCOS) method [288, 289] could correct for some of these deficiencies, where off-diagonal information can be included.

Chapter 7

Conclusions and Outlook

In this thesis, we have analysed a set of tools that aid studying electronic structure problems on quantum computers. Classical treatment of such systems quickly becomes intractable due to the exponential scaling of the Hilbert space dimension with K (the number of spin orbitals), which can naturally be represented on a linear number of qubits. However, due to the limitations of current quantum devices - lack of error correction, device topology (qubit connectivity) and low qubit numbers - it is not possible to implement fault-tolerant quantum algorithms, such as quantum phase estimation [76, 77]. Instead, variational quantum-classical algorithms, such as the variational quantum eigensolver [94], are used to study these problems. These algorithms are more NISQ friendly; however, suffer from many drawbacks. For VQE, one major obstacle to its success is the sheer number of measurements required to obtain the energy of a molecular system to within chemical precision. Estimates into the runtime of VQE for simple molecules has been put in the millions of years [175]. In chapter 4, we investigated the unitary partitioning measurement reduction strategy that decreases the measurement overhead of VQE and thus the algorithm’s runtime. We numerically found that for two molecular examples, the unitary partitioning measurement reduction strategy improves the standard error on the mean by an order of magnitude if applied as a sequence of rotations. As the $SEM \propto \frac{1}{\sqrt{M}}$, this is a 100 fold measurement reduction. Theoretically, it was proven that unitary partitioning can always reduce the number of terms requiring separate measurement linearly with respect to the number of spin orbitals from $\mathcal{O}(K^4) \mapsto \mathcal{O}(K^3)$ [198]. However, as discussed in chapter 4, what is more important to show is that for a fixed precision ϵ the total number of measurements required is reduced. Therefore, even though alternate measurement reduction approaches can reduce the number of terms by a greater amount - e.g. the “Basis Rotation Grouping ” strategy [179] $\mathcal{O}(K^4) \mapsto \mathcal{O}(K)$, what is important to investigate is the R -ratio [201]. This takes into account variances and covariances which is the critical thing to consider when analysing the measurement reduction obtained for a given problem. Gonthier *et al.* investigated this in the context of quantum chemistry and wrote: “work on improving Hamiltonian transformations to reduce the Hamiltonian variance” is needed [163]. This is a very interesting point; however, they did not outline how this could be done. We believe one avenue to

achieve this is to use orbital localization schemes such as those discussed in chapter 6, to transform a molecular Hamiltonian into an alternate basis, where the Pauli coefficients should be updated to improve the measurement reduction obtained. This could be done by maximising the R -ratio. While the unitary partitioning measurement reduction approach does not solve the measurement roadblock of VQE, this idea can be used in combination with many other reduction methods. We leave this to future investigation.

In chapter 5, we analysed the contextual subspace VQE algorithm [244], which gives an approximate simulation method where the quantum resources for a problem can be reduced. This allows one to approximately solve any given qubit problem on current and near term quantum computers. We deconstructed each step of the CS-VQE algorithm and showed that classically applying the unitary partitioning rotation step as a linear combination of unitaries will at worst increase the number of Pauli operators in an n -qubit Hamiltonian by $\mathcal{O}(n^2)$. Importantly, this makes mapping the problem into the contextual subspace classically tractable and gives an exponential improvement upon the $\mathcal{O}(2^n)$ scaling in the originating work [244]. We then applied the unitary partitioning measurement reduction strategy to different CS-VQE Hamiltonians to investigate further possible measurement reductions. As the R -ratio for unitary partitioning is lower bounded at 1 and the conditions to meet this bound are highly unlikely, for example when no partitioning is done, a measurement reduction will be obtained. While we focused on quantum resource reduction, in the future we plan to look into if there is any structure in the contextual terms of a molecular Hamiltonian and if anything can be said about where they arise from. Furthermore, we suspect that the contextual subspace could be a good tool for motivating an active space approximation for a given molecular problem. What is interesting about using this method here is it approaches the problem from a quantum foundations angle, rather than using a “chemist’s intuition”. Which is better is unknown and we leave an investigation for future work. Importantly, the contextual subspace approach does not necessarily require being solved via VQE, for instance quantum phase estimation can be used to solve the contextual problem (H_{con}). This algorithm will therefore be useful even when fault-tolerant quantum devices are available.

As it stands, the CS-VQE method doesn’t utilize classical quantum chemistry techniques beyond Hartree-Fock. This is common in most of the quantum computing approaches to quantum chemistry, where problems only begin from the Hartree-Fock solution; however, many post-Hartree-Fock methods can still be taken advantage of. Such approximate classical solutions offer a wealth of information. Something I feel is lacking in the quantum computing community is not utilizing this information. With respect to this, we are currently investigating how the noncontextual Hamiltonian can be motivated by a classical solution to a given electronic structure problem. This

is ongoing work, but the core idea is to construct stabilizers from a conventional chemistry solution (such as a classically obtained CCSD wavefunction), that are then used to motivate the contextual subspace. We believe this could be a way to achieve quantum advantage, as by the variational principle any decrease in ground state energy measured results in a better approximation of the true ground state. The utility of this approach is it uses the best-in-class classical methods to help motivate the quantum problem.

In a similar vein to the idea of using classical quantum chemistry methods to help study a given electronic structure problem, in chapter 6 we introduced the projection-based embedding model in the context of generating an embedded qubit Hamiltonian [286]. We showed this model can dramatically decrease the resources required to study a given molecular problem on quantum hardware. Importantly, this tool allows any molecular problem to be studied on any sized available quantum hardware. Thus even when fault-tolerant quantum computers are available, such multi-scale approaches can still be used. For instance, a protein can be made of millions of atoms and a full quantum mechanical description of such a system is unforeseeable in the near term future. We see such problems through the lens of an embedding problem, where an active system should be grown (either in number of active atoms or basis set) to fit on the available quantum hardware. Any remaining part of the problem should then be solved via classically tractable methods using any necessary approximations. By targeting the best resources to areas of interest in a molecular problem, such as specific functional groups, we hope this will improve our understanding of different materials and molecules.

Finally, it is my view that quantum computers are unlikely to ever fully replace classical computers in the study of quantum chemistry. I believe integrating PBE, CS-VQE and measurement reduction techniques in a combined quantum-classical hierarchy will offer a powerful tool to study molecular problems utilizing the best classical and quantum computers available. As we have open-source code for each of these methods [264, 335], we hope this approach will allow new insights into quantum chemistry problems.

Bibliography

- [1] T. S. Kuhn, *Black-Body Theory and the Quantum Discontinuity, 1894-1912* (Oxford University Press, 1987).
- [2] C. A. Gearhart, *Physics in Perspective* **4**, 170–215 (2002).
- [3] B. Schwarzschild, *Physics Today* **59**, 15–17 (2006).
- [4] B. Odom, D. Hanneke, B. D’Urso, and G. Gabrielse, *Physical Review Letters* **97**, 030801 (2006).
- [5] G. Gabrielse, *Physics Today* **66**, 64–65 (2013).
- [6] P. Dirac, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **123**, 714–733 (1929).
- [7] R. P. Feynman, *International Journal of Theoretical Physics* **21**, 467–488 (1982).
- [8] I. Buluta and F. Nori, *Science* **326**, 108–111 (2009).
- [9] I. M. Georgescu, S. Ashhab, and F. Nori, *Reviews of Modern Physics* **86**, 153–185 (2014), arXiv:1308.6253.
- [10] S. Lloyd, *Science* **273**, 1073–1078 (1996).
- [11] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, *Chemical Reviews* **119**, 10856–10915 (2019), arXiv:1812.09976.
- [12] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, *Quantum Science and Technology* **3**, 030503 (2018), arXiv:1710.01022.
- [13] Y. Cao, J. Romero, and A. Aspuru-Guzik, *IBM Journal of Research and Development* **62**, 6:1–6:20 (2018).
- [14] A. Aspuru-Guzik, R. Lindh, and M. Reiher, *ACS Central Science* **4**, 144–152 (2018).
- [15] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, *Physical Review X* **8**, 011021 (2018).
- [16] V. E. Elfving, B. W. Broer, M. Webber, J. Gavartin, M. D. Halls, K. P. Lorton, and A. Bochevarov, 1–20 (2020), arXiv:2009.12472.

- [17] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, *Reviews of Modern Physics* **92**, 015003 (2020), arXiv:1808.10402.
- [18] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st (MacMillan Publishing Co., New York, 2012).
- [19] D. B. Cook, *Handbook of Computational Quantum Chemistry* (Dover Publications, 2005).
- [20] M. Born and R. Oppenheimer, *Annalen der Physik* **389**, 457–484 (1927).
- [21] J. C. Slater, *Physical Review* **34**, 1293–1322 (1929).
- [22] L. Piela, in *Ideas of quantum chemistry* (Elsevier, 2014), e109–e120.
- [23] E. U. Condon, *Physical Review* **36**, 1121–1133 (1930).
- [24] J. L. Bao, L. Gagliardi, and D. G. Truhlar, *The Journal of Physical Chemistry Letters* **9**, 2353–2358 (2018).
- [25] G. Hall and J. E. Lennard-Jones, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **205**, 541–552 (1951).
- [26] C. C. J. Roothaan, *Reviews of Modern Physics* **23**, 69–89 (1951).
- [27] E. Cancès and C. Le Bris, *ESAIM: Mathematical Modelling and Numerical Analysis* **34**, 749–774 (2000).
- [28] M Cross, *Applied Mathematical Modelling* **9**, 226 (1985).
- [29] E. G. Lewars, *Computational Chemistry* (Springer International Publishing, Cham, 2016), pp. 1–728.
- [30] J. Frank, *Introduction to computational chemistry* (John Wiley and Sons Ltd, John wiley and sons, 2017).
- [31] T. Nakajima, M. Katouda, M. Kamiya, and Y. Nakatsuka, *International Journal of Quantum Chemistry* **115**, 349–359 (2015).
- [32] I. Kassal, J. D. Whitfield, A. Perdomo-Ortiz, M.-H. Yung, and A. Aspuru-Guzik, *Annual Review of Physical Chemistry* **62**, 185–207 (2011), arXiv:1007.2648.
- [33] Y. Su, D. W. Berry, N. Wiebe, N. Rubin, and R. Babbush, *PRX Quantum* **2**, 040332 (2021), arXiv:2105.12767.
- [34] J. T. Seeley, M. J. Richard, and P. J. Love, *The Journal of Chemical Physics* **137**, 224109 (2012), arXiv:1208.5986.
- [35] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, UK, 2000).

- [36] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, *Physical Review X* **8**, 031022 (2018), arXiv:1803.10238.
- [37] C. Møller and M. S. Plesset, *Physical Review* **46**, 618–622 (1934).
- [38] D. Cremer, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 509–530 (2011).
- [39] I. Shavitt, *Molecular Physics* **94**, 3–17 (1998).
- [40] R. J. Bartlett and M. Musiał, *Reviews of Modern Physics* **79**, 291–352 (2007).
- [41] A. Kitaev, A. Shen, and M. Vyalyi, *Classical and Quantum Computation*, Vol. 47, Graduate Studies in Mathematics (American Mathematical Society, Providence, Rhode Island, 2002).
- [42] T. E. O’Brien, B. Tarasinski, and B. M. Terhal, *New Journal of Physics* **21**, 023022 (2019), arXiv:1809.09697.
- [43] P. Hohenberg and W. Kohn, *Physical Review* **136**, B864–B871 (1964).
- [44] J. E. Harriman, *Physical Review A* **27**, 632–645 (1983).
- [45] R. Pino, O. Bokanowski, E. V. Ludeña, and R. L. Boada, *Theoretical Chemistry Accounts* **118**, 557–561 (2007).
- [46] E. H. Lieb, *International Journal of Quantum Chemistry* **24**, 243–277 (1983).
- [47] W. Koch and M. C. Holthausen, *A Chemist’s Guide to Density Functional Theory*, Vol. 3 (Wiley, 2001), p. 294.
- [48] M. Levy, *Proceedings of the National Academy of Sciences* **76**, 6062–6065 (1979).
- [49] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1995), pp. 1–352.
- [50] G. Zumbach and K. Maschke, *Physical Review A* **28**, 544–554 (1983).
- [51] A. Gonis and M. Däne, *Journal of Physics and Chemistry of Solids* **116**, 100–112 (2018).
- [52] J. Cioslowski, *Physical Review Letters* **60**, 2141–2143 (1988).
- [53] J. Cioslowski, *The Journal of Chemical Physics* **89**, 4871–4874 (1988).
- [54] J. Cioslowski, *International Journal of Quantum Chemistry* **36**, 255–262 (2009).
- [55] M. Levy, *International Journal of Quantum Chemistry* **110**, 3140–3144 (2010).
- [56] K. Capelle, C. A. Ullrich, and G. Vignale, *Physical Review A* **76**, 012508 (2007), arXiv:0610322 [cond-mat].

- [57] M. B. Einhorn and R. Blankenbecler, *Annals of Physics* **67**, 480–517 (1971).
- [58] R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke, *The Journal of Chemical Physics* **68**, 3801–3807 (1978).
- [59] E. J. Baerends, *Physical Chemistry Chemical Physics* **24**, 12745–12766 (2022), arXiv:2205.04851.
- [60] W. Kohn and L. J. Sham, *Physical Review* **140**, A1133–A1138 (1965).
- [61] T. L. Gilbert, *Physical Review B* **12**, 2111–2120 (1975).
- [62] J. E. Harriman, *Physical Review A* **24**, 680–682 (1981).
- [63] P. Jakobsen and F. Jensen, *Journal of Chemical Theory and Computation* **17**, 269–276 (2021).
- [64] *The Nobel Prize in Chemistry 1998*.
- [65] M. Y. Amusia, A. Z. Msezane, and V. R. Shaginyan, *Physica Scripta* **68**, C133–C140 (2003), arXiv:0204104 [cond-mat].
- [66] N. Mardirossian and M. Head-Gordon, *Molecular Physics* **115**, 2315–2372 (2017).
- [67] P. M. Gill, B. G. Johnson, and J. A. Pople, *Chemical Physics Letters* **209**, 506–512 (1993).
- [68] S. Dasgupta and J. M. Herbert, *Journal of Computational Chemistry* **38**, 869–882 (2017).
- [69] U von Barth and L. Hedin, *Journal of Physics C: Solid State Physics* **5**, 1629–1642 (1972).
- [70] C. R. Jacob and M. Reiher, *International Journal of Quantum Chemistry* **112**, 3661–3684 (2012), arXiv:1206.2234.
- [71] J. P. Perdew and A. Zunger, *Physical Review B* **23**, 5048–5079 (1981).
- [72] A. J. Cohen, P. Mori-Sanchez, and W. Yang, *Science* **321**, 792–794 (2008).
- [73] K. R. Bryenton, A. A. Adeleke, S. G. Dale, and E. R. Johnson, *WIREs Computational Molecular Science*, 29–32 (2022).
- [74] A. Ekert and R. Jozsa, *Reviews of Modern Physics* **68**, 733–753 (1996).
- [75] D. Coppersmith, **19642** (2002), arXiv:0201067 [quant-ph].
- [76] D. S. Abrams and S. Lloyd, *Physical Review Letters* **83**, 5162–5165 (1999), arXiv:9807070 [quant-ph].
- [77] A. Y. Kitaev, 1–22 (1995), arXiv:9511026 [quant-ph].
- [78] J. Preskill, *Quantum* **2**, 79 (2018), arXiv:1801.00862.

- [79] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, *Reviews of Modern Physics* **94**, 015004 (2022).
- [80] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, *Nature* **574**, 505–510 (2019).
- [81] Y. Li and S. C. Benjamin, *Physical Review X* **7**, 021050 (2017), arXiv:1611.09301.
- [82] E. Pednault, J. A. Gunnels, G. Nannicini, L. Horesh, and R. Wisnieff, 1–39 (2019), arXiv:1910.09534.
- [83] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, 1–27 (2022), arXiv:2211.03999.
- [84] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, *Science* **370**, 1460–1463 (2020), arXiv:2012.01625.
- [85] A. W. Harrow and J. C. Napp, *Physical Review Letters* **126**, 140502 (2021), arXiv:1901.05374.
- [86] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, Cambridge, 2012).
- [87] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, *Physics Reports* **986**, 1–128 (2022), arXiv:2111.05176.
- [88] J. R. McClean, R. Babbush, P. J. Love, and A. Aspuru-Guzik, *The Journal of Physical Chemistry Letters* **5**, 4368–4380 (2014), arXiv:1407.7863.

- [89] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, *Science* **309**, 1704–1707 (2005).
- [90] P. J. J. O’Malley, R Babbush, I. D. Kivlichan, J Romero, J. R. McClean, R Barends, J Kelly, P Roushan, A Tranter, N Ding, B Campbell, Y. Chen, Z Chen, B Chiaro, A Dunsworth, A. G. Fowler, E Jeffrey, E Lucero, A Megrant, J. Y. Mutus, M Neeley, C Neill, C Quintana, D Sank, A Vainsencher, J Wenner, T. C. White, P. V. Coveney, P. J. Love, H Neven, A. Aspuru-Guzik, and J. M. Martinis, *Physical Review X* **6**, 031007 (2016), arXiv:1512.06860.
- [91] J. H. van Vleck, *Physical Review* **49**, 232–240 (1936).
- [92] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, M. Kastoryano, R. Babbush, J. Preskill, D. R. Reichman, E. T. Campbell, E. F. Valeev, L. Lin, and G. K.-L. Chan, (2022), arXiv:2208.02199.
- [93] H. Mohammadbagherpoor, Y.-H. Oh, A. Singh, X. Yu, and A. J. Rindos, 1–6 (2019), arXiv:1903.07605.
- [94] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, *Nature Communications* **5**, 4213 (2014).
- [95] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, *New Journal of Physics* **18**, 023023 (2016), arXiv:1509.04279.
- [96] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, D. P. Tew, X. Zhou, J. L. O’Brien, and M. G. Thompson, *Science Advances* **4**, 1–12 (2018), arXiv:1611.03511.
- [97] O. Higgott, D. Wang, and S. Brierley, *Quantum* **3**, 156 (2019), arXiv:1805.08138.
- [98] E. K. U. Gross, L. N. Oliveira, and W. Kohn, *Physical Review A* **37**, 2805–2808 (1988).
- [99] F. M. Fernández, 1–7 (2022), arXiv:2206.05122.
- [100] B. Klahn and W. A. Bingel, *Theoretica Chimica Acta* **44**, 9–26 (1977).
- [101] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, *Physical Review A* **95**, 020501 (2017), arXiv:1506.00443.
- [102] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, D. P. Tew, X. Zhou, J. L. O’Brien, and M. G. Thompson, *Science Advances* **4**, 1–12 (2018), arXiv:1611.03511.

- [103] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Nature* **549**, 242–246 (2017), arXiv:1704.05018.
- [104] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, *Nature* **567**, 491–495 (2019).
- [105] I. G. Ryabinkin, T.-C. Yen, S. N. Genin, and A. F. Izmaylov, *Journal of Chemical Theory and Computation* **14**, 6317–6326 (2018), arXiv:1809.03827.
- [106] Y. Nam, J.-S. Chen, N. C. Pienti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, K. M. Beck, A. Blinov, V. Chaplin, M. Chmielewski, C. Collins, S. Debnath, K. M. Hudek, A. M. Ducore, M. Keesan, S. M. Kreikemeier, J. Mizrahi, P. Solomon, M. Williams, J. D. Wong-Campos, D. Moehring, C. Monroe, and J. Kim, *npj Quantum Information* **6**, 33 (2020), arXiv:1902.10171.
- [107] A. J. McCaskey, Z. P. Parks, J. Jakowski, S. V. Moore, T. D. Morris, T. S. Humble, and R. C. Pooser, *npj Quantum Information* **5**, 99 (2019), arXiv:1905.01534.
- [108] Q. Gao, H. Nakamura, T. P. Gujarati, G. O. Jones, J. E. Rice, S. P. Wood, M. Pistoia, J. M. Garcia, and N. Yamamoto, *The Journal of Physical Chemistry A* **125**, 1827–1836 (2021), arXiv:1906.10675.
- [109] S. E. Smart and D. A. Mazziotti, *Physical Review A* **100**, 022517 (2019), arXiv:2004.07739.
- [110] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, S. Habegger, M. P. Harrigan, A. Ho, S. Hong, T. Huang, W. J. Huggins, L. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, O. Martin, J. M. Martinis, J. R. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, H. Neven, M. Y. Niu, T. E. O’Brien, E. Ostby, A. Petukhov, H. Putterman, C. Quintana, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, D. Strain, K. J. Sung, M. Szalay, T. Y. Takeshita, A. Vainsencher, T. White, N. Wiebe, Z. J. Yao, P. Yeh, and A. Zalcman, *Science* **369**, 1084–1089 (2020), arXiv:2004.04174.
- [111] Q. Gao, G. O. Jones, M. Motta, M. Sugawara, H. C. Watanabe, T. Kobayashi, E. Watanabe, Y.-y. Ohnishi, H. Nakamura, and N. Yamamoto, *npj Computational Materials* **7**, 70 (2021), arXiv:2007.15795.

- [112] Y. Kawashima, E. Lloyd, M. P. Coons, Y. Nam, S. Matsuura, A. J. Garza, S. Johri, L. Huntington, V. Senicourt, A. O. Maksymov, J. H. V. Nguyen, J. Kim, N. Alidoust, A. Zaribafiyani, and T. Yamazaki, *Communications Physics* **4**, 245 (2021), arXiv:2102.07045.
- [113] J. E. Rice, T. P. Gujarati, M. Motta, T. Y. Takeshita, E. Lee, J. A. Latone, and J. M. Garcia, *The Journal of Chemical Physics* **154**, 134115 (2021).
- [114] A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, and S. Sheldon, *PRX Quantum* **3**, 010309 (2022), arXiv:2104.10220.
- [115] M. Motta, G. O. Jones, J. E. Rice, T. P. Gujarati, R. Sakuma, I. Liepuoniute, J. M. Garcia, and Y. Ohnishi, 1–13 (2022), arXiv:2208.02414.
- [116] K. Yamamoto, D. Z. Manrique, I. T. Khan, H. Sawada, and D. M. Ramo, *Physical Review Research* **4**, 033110 (2022), arXiv:2109.08401.
- [117] J. J. M. Kirsopp, C. Di Paola, D. Z. Manrique, M. Krompiec, G. Greene-Diniz, W. Guba, A. Meyder, D. Wolf, M. Strahm, and D. Muñoz Ramo, *International Journal of Quantum Chemistry* **122**, 1–16 (2022), arXiv:2110.08163.
- [118] I. T. Khan, M. Tudorovskaya, J. J. M. Kirsopp, D. M. Ramo, P. W. Warrier, D. K. Papanastasiou, and R. Singh, (2022), arXiv:2210.14834.
- [119] T. E. O'Brien, G. Anselmetti, F. Gkritis, V. E. Elfving, S. Polla, W. J. Huggins, O. Oumarou, K. Kechedzhi, D. Abanin, R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, K. Anderson, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, D. Bacon, J. C. Bardin, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, J. Campero, Y. Chen, Z. Chen, B. Chiaro, D. Chik, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, S. Demura, I. Drozdov, A. Dunsworth, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, V. S. Ferreira, L. F. Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, R. Gosula, A. G. Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. Hansen, M. P. Harrigan, S. D. Harrington, P. Heu, J. Hilton, M. R. Hoffmann, S. Hong, T. Huang, A. Huff, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, J. Kelly, T. Khattar, M. Khezri, M. Kieferová, S. Kim, P. V. Klimov, A. R. Klots, R. Kothari, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. T. Lill, W. Liu, W. P. Livingston, A. Locharla, E. Lucero, F. D. Malone, S. Mandra, O. Martin, S. Martin, J. R. McClean, T. McCourt, M. McEwen, A. Megrant, X. Mi, A. Mieszala, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, R. Movassagh, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. H. Ng,

- A. Nguyen, M. Nguyen, M. Y. Niu, S. Omonije, A. Opremcak, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, C. Rocque, P. Roushan, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, N. Shutty, V. Shvarts, J. Skruzny, V. Smelyanskiy, W. C. Smith, R. Somma, G. Sterling, D. Strain, M. Szalay, D. Thor, A. Torres, G. Vidal, B. Villalonga, C. V. Heidweiller, T. White, B. W. K. Woo, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, N. Zhu, N. Zobrist, C. Gogolin, R. Babbush, and N. C. Rubin, 1–23 (2022), arXiv:2210.10799.
- [120] L. Zhao, J. Goings, K. Wright, J. Nguyen, J. Kim, S. Johri, K. Shin, W. Kyoung, J. I. Fuks, J.-K. K. Rhee, and Y. M. Rhee, (2022), arXiv:2212.02482.
- [121] T. Weaving, A. Ralli, W. M. Kirby, P. J. Love, S. Succi, and P. V. Coveney, 1–18 (2023), arXiv:2303.00445.
- [122] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, *Physical Review X* **8**, 011044 (2018).
- [123] P. Jordan and E. Wigner, *Zeitschrift fur Physik* **47**, 631–651 (1928).
- [124] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, 1–15 (2017), arXiv:1701.08213.
- [125] S. B. Bravyi and A. Y. Kitaev, *Annals of Physics* **298**, 210–226 (2002), arXiv:0003137 [quant-ph].
- [126] Z. Jiang, A. Kalev, W. Mruzekiewicz, and H. Neven, *Quantum* **4**, 276 (2020), arXiv:1910.10746.
- [127] M. Steudtner and S. Wehner, *New Journal of Physics* **20**, 063010 (2018).
- [128] R. Babbush, D. W. Berry, Y. R. Sanders, I. D. Kivlichan, A. Scherer, A. Y. Wei, P. J. Love, and A. Aspuru-Guzik, *Quantum Science and Technology* **3**, 015006 (2018), arXiv:1506.01029.
- [129] W. Kirby, B. Fuller, C. Hadfield, and A. Mezzacapo, *PRX Quantum* **3**, 020351 (2022), arXiv:2109.14465.
- [130] N. Moll, A. Fuhrer, P. Staar, and I. Tavernelli, *Journal of Physics A: Mathematical and Theoretical* **49**, 295301 (2016), arXiv:1510.04048.
- [131] B. Harrison, D. Nelson, D. Adamiak, and J. Whitfield, 1–7 (2022), arXiv:2211.04501.
- [132] A. Tranter, P. J. Love, F. Mintert, and P. V. Coveney, *Journal of Chemical Theory and Computation* **14**, 5617–5630 (2018), arXiv:1812.02233.
- [133] K. Setia, R. Chen, J. E. Rice, A. Mezzacapo, M. Pistoia, and J. D. Whitfield, *Journal of Chemical Theory and Computation* **16**, 6091–6097 (2020), arXiv:1910.14644.

- [134] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, *Advanced Quantum Technologies* **2**, 1900070 (2019), arXiv:1905.10876.
- [135] T. Volkoff and P. J. Coles, *Quantum Science and Technology* **6**, 025008 (2021), arXiv:2005.12200.
- [136] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, *Nature Communications* **9**, 4812 (2018), arXiv:1803.11173.
- [137] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, *Physical Review A* **98**, 1–14 (2018), arXiv:1805.04340.
- [138] M. Ganzhorn, D. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp, *Physical Review Applied* **11**, 044092 (2019), arXiv:1809.05057.
- [139] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, *Quantum* **3**, 214 (2019), arXiv:1903.05076.
- [140] O. R. Meitei, B. T. Gard, G. S. Barron, D. P. Pappas, S. E. Economou, E. Barnes, and N. J. Mayhall, *npj Quantum Information* **7**, 155 (2021).
- [141] M. Kühn, S. Zanker, P. Deglmann, M. Marthaler, and H. Weiß, *Journal of Chemical Theory and Computation* **15**, 4764–4780 (2019), arXiv:1812.06814.
- [142] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, *Quantum Science and Technology* **4**, 014008 (2018), arXiv:1701.02691.
- [143] D. A. Fedorov, B. Peng, N. Govind, and Y. Alexeev, *Materials Theory* **6**, 1–23 (2021), arXiv:2103.08505.
- [144] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, *Journal of Chemical Theory and Computation* **15**, 311–324 (2019), arXiv:1810.02327.
- [145] W. Mizukami, K. Mitarai, Y. O. Nakagawa, T. Yamamoto, T. Yan, and Y.-y. Ohnishi, *Physical Review Research* **2**, 033421 (2020), arXiv:1910.11526.
- [146] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, *Nature Communications* **10**, 3007 (2019), arXiv:1812.11173.
- [147] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, *PRX Quantum* **2**, 020310 (2021).
- [148] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, *Physical Review A* **98**, 022322 (2018), arXiv:1805.04340.
- [149] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, *npj Quantum Information* **6**, 10 (2020), arXiv:1904.10910.

- [150] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, *PRX Quantum* **3**, 010313 (2022), arXiv:2101.02138.
- [151] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, *Physical Review A* **98**, 1–3 (2018), arXiv:1803.00745.
- [152] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, *Physical Review A* **99**, 032331 (2019), arXiv:1811.11184.
- [153] T. E. O’Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, *npj Quantum Information* **5**, 10.1038/s41534-019-0213-4 (2019), arXiv:1905.03742.
- [154] M. Cerezo and P. J. Coles, *Quantum Science and Technology* **6**, 035006 (2021), arXiv:2008.07454.
- [155] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, *Quantum* **5**, 558 (2021), arXiv:2011.12245.
- [156] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, *Physical Review Research* **3**, 033090 (2021), arXiv:2012.12658.
- [157] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, *Nature Communications* **12**, 6961 (2021), arXiv:2007.14384.
- [158] H. R. Grimsley, G. S. Barron, E. Barnes, S. E. Economou, and N. J. Mayhall, (2022), arXiv:2204.07179.
- [159] T. Haug and M. S. Kim, 1–15 (2021), arXiv:2104.14543.
- [160] A. Pellow-Jarman, I. Sinayskiy, A. Pillay, and F. Petruccione, *Quantum Information Processing* **20**, 202 (2021), arXiv:2106.08682.
- [161] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, *Nature Reviews Physics* **3**, 625–644 (2021), arXiv:2012.09265.
- [162] B. Bauer, S. Bravyi, M. Motta, and G. K.-l. Chan, *Chemical Reviews* **120**, 12685–12717 (2020), arXiv:2001.03685.
- [163] J. F. Gonthier, M. D. Radin, C. Buda, E. J. Duskocil, C. M. Abuan, and J. Romero, *Physical Review Research* **4**, 033154 (2022).
- [164] D. Wang, O. Higgott, and S. Brierley, *Physical Review Letters* **122**, 140504 (2019), arXiv:1802.00171.
- [165] L. Bittel and M. Kliesch, *Physical Review Letters* **127**, 120502 (2021), arXiv:2101.07267.
- [166] M. W. Krentel, *Journal of Computer and System Sciences* **36**, 490–509 (1988).

- [167] W. Saib, P. Wallden, and I. Akhalwaya, in 2021 IEEE International Conference on Quantum Computing and Engineering (QCE) (2021), pp. 42–53, arXiv:2108.12388.
- [168] K. Dalton, C. K. Long, Y. S. Yordanov, C. G. Smith, C. H. W. Barnes, N. Mertig, and D. R. M. Arvidsson-Shukur, 1–16 (2022), arXiv:2211.04505.
- [169] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, 1–41 (2022), arXiv:2210.11505.
- [170] G. Tóth and D. Petz, *Physical Review A* **87**, 032324 (2013), arXiv:1109.2831.
- [171] P. Gibilisco and T. Isola, *Journal of Mathematical Analysis and Applications* **384**, 670–676 (2011).
- [172] N. C. Rubin, R. Babbush, and J. McClean, *New Journal of Physics* **20**, 053020 (2018).
- [173] D. Wecker, M. B. Hastings, and M. Troyer, *Physical Review A* **92**, 042303 (2015).
- [174] A. Arrasmith, L. Cincio, R. D. Somma, and P. J. Coles, 1–11 (2020), arXiv:2004.06252.
- [175] P. D. Johnson, A. A. Kunitsa, J. F. Gonthier, M. D. Radin, C. Buda, E. J. Duskocil, C. M. Abuan, and J. Romero, 1–15 (2022), arXiv:2203.07275.
- [176] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, 1–20 (2022), arXiv:2203.11374.
- [177] J. Cotler and F. Wilczek, *Physical Review Letters* **124**, 100401 (2020), arXiv:1908.02754.
- [178] X. Bonet-Monroig, R. Babbush, and T. E. O’Brien, *Physical Review X* **10**, 031064 (2020), arXiv:1908.05628.
- [179] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, *npj Quantum Information* **7**, 23 (2021), arXiv:1907.13117.
- [180] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. K.-L. Chan, *npj Quantum Information* **7**, 83 (2021), arXiv:1808.02625.
- [181] D. W. Berry, C. Gidney, M. Motta, J. R. McClean, and R. Babbush, *Quantum* **3**, 208 (2019), arXiv:1902.02134.
- [182] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, *Physical Review Letters* **120**, 110501 (2018), arXiv:1711.04789.
- [183] J. F. Gonthier, M. D. Radin, C. Buda, E. J. Duskocil, C. M. Abuan, and J. Romero, 1–27 (2020), arXiv:2012.04001.

- [184] S. Aaronson, *SIAM Journal on Computing* **49**, STOC18–368–STOC18–394 (2020).
- [185] H.-Y. Huang, R. Kueng, and J. Preskill, *Nature Physics* **16**, 1050–1057 (2020), arXiv:2002.08953.
- [186] S. Aaronson and D. Gottesman, *Physical Review A* **70**, 052328 (2004), arXiv:0406196 [quant-ph].
- [187] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, *Communications in Mathematical Physics* **391**, 951–967 (2022), arXiv:2006.15788.
- [188] H. Y. Huang, R. Kueng, and J. Preskill, *Physical Review Letters* **127**, 30503 (2021), arXiv:2103.07510.
- [189] D. Gottesman, *Stabilizer codes and quantum error correction* (California Institute of Technology, 1997).
- [190] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, *IEEE Transactions on Quantum Engineering* **1**, 1–24 (2019), arXiv:1907.13623.
- [191] J. Lawrence, Č. Brukner, and A. Zeilinger, *Physical Review A* **65**, 032320 (2002), arXiv:0104012 [quant-ph].
- [192] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, *IEEE Transactions on Quantum Engineering* **1**, 1–24 (2021), arXiv:1908.11857.
- [193] A. Jena, S. Genin, and M. Mosca, (2019), arXiv:1907.07859.
- [194] T.-C. Yen, V. Verteletskyi, and A. F. Izmaylov, *Journal of Chemical Theory and Computation* **16**, 2400–2409 (2020), arXiv:1907.09386.
- [195] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, *IEEE Transactions on Quantum Engineering* **1**, 1–24 (2020), arXiv:1908.11857.
- [196] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, *The Journal of Chemical Physics* **152**, 124114 (2020), arXiv:1907.03358.
- [197] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, *Journal of Chemical Theory and Computation* **16**, 190–195 (2020), arXiv:1907.09040.
- [198] A. Zhao, A. Tranter, W. M. Kirby, S. F. Ung, A. Miyake, and P. J. Love, *Physical Review A* **101**, 062322 (2020), arXiv:1908.08067.
- [199] A. Zhao, A. Tranter, W. M. Kirby, S. F. Ung, A. Miyake, and P. J. Love, *Physical Review A* **101**, 062322 (2020), arXiv:1908.08067.

- [200] R. M. Karp, in *Complexity of computer computations* (Springer US, Boston, MA, 1972), pp. 85–103.
- [201] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, *Quantum* **5**, 385 (2021), arXiv:1908.06942.
- [202] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 4th Editio (MIT Press, 2022), p. 1332.
- [203] T.-C. Yen and A. F. Izmaylov, *PRX Quantum* **2**, 040320 (2021), arXiv:2007.01234.
- [204] Y.-Y. Shi, *Quantum Information and Computation* **3**, 84–92 (2003), arXiv:0205115 [quant-ph].
- [205] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, *Molecular Physics* **109**, 735–750 (2011), arXiv:1001.3855.
- [206] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurel, *Physical Review A* **101**, 012350 (2020), arXiv:1905.05374.
- [207] A. Cowtan, S. Dilkes, R. Duncan, W. Simmons, and S. Sivarajah, *Electronic Proceedings in Theoretical Computer Science* **318**, 213–228 (2019), arXiv:1906.01734.
- [208] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, *Quantum Information and Computation* **15**, 1–21 (2015), arXiv:1403.1539.
- [209] A. M. Childs and N. Wiebe, *Physical Review A* **52**, 3457–3467 (2012), arXiv:1202.5822.
- [210] G. H. Low and I. L. Chuang, *Quantum* **3**, 163 (2019), arXiv:1610.06546.
- [211] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, in *Proceedings of the 51st annual acm sigact symposium on theory of computing* (2019), pp. 193–204, arXiv:1806.01838.
- [212] D. Camps and R. Van Beeumen, *Physical Review A* **102**, 052411 (2020), arXiv:2007.01417.
- [213] D. Camps and R. Van Beeumen, in *2022 ieee international conference on quantum computing and engineering (qce)* (2022), pp. 104–113, arXiv:2205.00081.
- [214] G.-L. Long and Y. Sun, *Physical Review A* **64**, 014303 (2001), arXiv:0104030 [quant-ph].
- [215] M. Mottonen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, *Quantum Information and Computation* **5**, 467–473 (2004), arXiv:0407010 [quant-ph].
- [216] V. Shende, S. Bullock, and I. Markov, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **25**, 1000–1010 (2006).

- [217] I. F. Araujo, D. K. Park, F. Petruccione, and A. J. da Silva, *Scientific Reports* **11**, 6329 (2021), arXiv:2008.01511.
- [218] S. S. Bullock and I. L. Markov, 1–12 (2003), arXiv:0303039 [quant-ph].
- [219] M. Möttönen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, *Physical Review Letters* **93**, 130502 (2004), arXiv:0404089 [quant-ph].
- [220] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, *Physical Review A* **52**, 3457–3467 (1995).
- [221] A. Ralli, P. J. Love, A. Tranter, and P. V. Coveney, *Physical Review Research* **3**, 033195 (2021), arXiv:2012.02765.
- [222] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. Chan, *WIREs Computational Molecular Science* **8**, 10.1002/wcms.1340 (2018).
- [223] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, *The Journal of Chemical Physics* **153**, 024109 (2020), arXiv:2002.12531.
- [224] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, 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. P. D. Sawaya, B. Senjean, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, *Quantum Science and Technology* **5**, 034014 (2020), arXiv:1710.07629.
- [225] A. A. Hagberg, D. A. Schult, and P. J. Swart, in *Proceedings of the 7th python in science conference*, edited by G. Varoquaux, T. Vaught, and J. Millman (2008), pp. 11–15.
- [226] D. J. A. Welsh, *The Computer Journal* **10**, 85–86 (1967).
- [227] F. Leymann and J. Barzen, *Quantum Science and Technology* **5**, 044007 (2020).

- [228] M. Treinish, J. Gambetta, P. Nation, Qiskit-bot, P. Kassebaum, D. M. Rodríguez, S. d. l. P. González, S. Hu, K. Krsulich, J. Lishman, J. Garrison, L. Zdanski, J. Yu, M. Marques, J. Gacon, D. McKay, J. Gomez, L. Capelluto, Travis-S-IBM, A. Panigrahi, Lerongil, R. I. Rahman, S. Wood, L. Bello, T. Itoko, C. J. Wood, D. Singh, Drew, E. Arbel, and Glen, 10.5281/ZENODO.6924865 (2022).
- [229] F. J. Massey, *Journal of the American Statistical Association* **46**, 68–78 (1951).
- [230] S. S. Shapiro and M. B. Wilk, *Biometrika* **52**, 591 (1965).
- [231] B. Efron, in *Annals of statistics*, Vol. 4, 6 (1992), pp. 569–593.
- [232] S. S. Tannu and M. K. Qureshi, in *Proceedings of the 52nd annual ieee/acm international symposium on microarchitecture* (2019), pp. 279–290.
- [233] Y. Chen, M. Farahzad, S. Yoo, and T.-C. Wei, *Physical Review A* **100**, 052315 (2019).
- [234] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, *PRX Quantum* **2**, 1 (2021), arXiv:2105.02859.
- [235] S. Resch and U. R. Karpuzcu, *ACM Computing Surveys* **54**, 1–35 (2022), arXiv:1912.00546.
- [236] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, in *Proceedings of the forty-sixth annual acm symposium on theory of computing* (2014), pp. 283–292, arXiv:1312.1414.
- [237] L. K. Grover, *Physical Review Letters* **79**, 325–328 (1997).
- [238] G. G. Guerreschi, *Physical Review A* **99**, 022306 (2019), arXiv:1808.02900.
- [239] M. Boyer, G. Brassard, P. Høyer, and A. Tapp, *Fortschritte der Physik* **46**, 493–505 (1998), arXiv:9605034 [quant-ph].
- [240] E. Koridon, S. Yalouz, B. Senjean, F. Buda, T. E. O’Brien, and L. Visscher, *Physical Review Research* **3**, 033127 (2021), arXiv:2103.14753.
- [241] E. Farhi, J. Goldstone, and S. Gutmann, 1–16 (2014), arXiv:1411.4028.
- [242] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, 1–21 (2019), arXiv:1909.05820.
- [243] W. J. Huggins, B. A. O’Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, *Nature* **603**, 416–420 (2022), arXiv:2106.16235.
- [244] W. M. Kirby, A. Tranter, and P. J. Love, *Quantum* **5**, 456 (2021), arXiv:2011.10027.
- [245] A. F. Izmaylov, T.-C. Yen, and I. G. Ryabinkin, *Chemical Science* **10**, 3746–3755 (2019), arXiv:1810.11602.

- [246] S. Kochen and E. Specker, *Indiana University Mathematics Journal* **17**, 59–87 (1967).
- [247] C. Budroni, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **377**, 20190141 (2019).
- [248] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J. Å. Larsson, *Reviews of Modern Physics* **94**, 10.1103/RevModPhys.94.045007 (2022).
- [249] N. de Silva, *Physical Review A* **95**, 032108 (2017).
- [250] A. Cabello, *Physical Review Letters* **127**, 070401 (2021), arXiv:2011.13790.
- [251] M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Nature* **510**, 351–355 (2014), arXiv:1401.4174.
- [252] N. D. Mermin, *Physical Review Letters* **65**, 3373–3376 (1990).
- [253] A. Peres, *Physics Letters A* **151**, 107–108 (1990).
- [254] W. M. Kirby and P. J. Love, *Physical Review Letters* **123**, 200501 (2019), arXiv:1904.02260.
- [255] D. Wagner and K. Weihe, *Combinatorica* **15**, 135–150 (1995).
- [256] W. M. Kirby and P. J. Love, *Physical Review A* **102**, 032418 (2020), arXiv:2002.05693.
- [257] R. W. Spekkens, *Fundamental Theories of Physics* **181**, 83–135 (2014), arXiv:1409.5041.
- [258] G. Chiribella, R. W. S. Editors, P. Blanchard, C. Fuchs, G. Jaeger, and N. P. Landsman, *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens, 1st ed., Vol. 181, *Fundamental Theories of Physics* (Springer, Dordrecht, 2016), pp. 85–135.
- [259] D. Poulin, M. B. Hastings, D. Wecker, N. Wiebe, A. C. Doherty, and M. Troyer, *Quantum Information and Computation* **15**, 361–384 (2015), arXiv:1406.4920.
- [260] R. A. Lang, I. G. Ryabinkin, and A. F. Izmaylov, *Journal of Chemical Theory and Computation* **17**, 66–78 (2021), arXiv:2002.05701.
- [261] J. Dehaene and B. De Moor, *Physical Review A* **68**, 042318 (2003), arXiv:0304125 [quant-ph].
- [262] A. Ralli, T. Weaving, A. Tranter, W. M. Kirby, P. J. Love, and P. V. Coveney, *Physical Review Research* **5**, 013095 (2023).
- [263] W. Kirby and A. Tranter, *ContextualSubspaceVQE*, 2022.
- [264] A. Ralli and T. Weaving, *Symmer*, 2022.

- [265] L. O. Jones, M. A. Mosquera, G. C. Schatz, and M. A. Ratner, *Journal of the American Chemical Society* **142**, 3281–3295 (2020).
- [266] *The Nobel Prize in Chemistry 2013*, 2013.
- [267] Q. Sun and G. K. L. Chan, *Accounts of Chemical Research* **49**, 2705–2712 (2016), arXiv:1612.02576.
- [268] J. E. Inglesfield, *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* **334**, 527–538 (1991).
- [269] A. Georges and G. Kotliar, *Physical Review B* **45**, 6479–6483 (1992).
- [270] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, *Reviews of Modern Physics* **78**, 865–951 (2006).
- [271] D. Zgid and G. K.-L. Chan, *The Journal of Chemical Physics* **134**, 094115 (2011), arXiv:1012.3609.
- [272] W. Chibani, X. Ren, M. Scheffler, and P. Rinke, *Physical Review B* **93**, 165106 (2016), arXiv:1506.03680.
- [273] G. Knizia and G. K.-L. Chan, *Physical Review Letters* **109**, 186404 (2012), arXiv:1204.5783.
- [274] G. Knizia and G. K.-L. Chan, *Journal of Chemical Theory and Computation* **9**, 1428–1432 (2013), arXiv:1212.2679.
- [275] S. Wouters, C. A. Jiménez-Hoyos, Q. Sun, and G. K. Chan, *Journal of Chemical Theory and Computation* **12**, 2706–2719 (2016), arXiv:1603.08443.
- [276] K. Pernal, *Physical Chemistry Chemical Physics* **18**, 21111–21121 (2016).
- [277] P. Cortona, *Physical Review B* **44**, 8454–8458 (1991).
- [278] T. A. Wesolowski and A. Warshel, *The Journal of Physical Chemistry* **97**, 8050–8053 (1993).
- [279] T. A. Wesolowski, S. Shedge, and X. Zhou, *Chemical Reviews* **115**, 5891–5928 (2015).
- [280] N. Govind, Y. A. Wang, and E. A. Carter, *The Journal of Chemical Physics* **110**, 7677–7688 (1999).
- [281] T. Kluner, N. Govind, Y. A. Wang, and E. A. Carter, *The Journal of Chemical Physics* **116**, 42 (2002).
- [282] F. Libisch, C. Huang, and E. A. Carter, *Accounts of Chemical Research* **47**, 2768–2775 (2014).
- [283] P. Elliott, M. H. Cohen, A. Wasserman, and K. Burke, *Journal of Chemical Theory and Computation* **5**, 827–833 (2009).

- [284] P. Elliott, K. Burke, M. H. Cohen, and A. Wasserman, *Physical Review A* **82**, 024501 (2010), arXiv:0901.0942.
- [285] C. Huang and E. A. Carter, *The Journal of Chemical Physics* **135**, 194104 (2011).
- [286] F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller, *Journal of Chemical Theory and Computation* **8**, 2564–2568 (2012).
- [287] J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller, *The Journal of Chemical Physics* **137**, 224113 (2012), arXiv:1211.6052.
- [288] M. A. Mosquera, M. A. Ratner, and G. C. Schatz, *The Journal of Chemical Physics* **149**, 034105 (2018).
- [289] M. A. Mosquera, L. O. Jones, M. A. Ratner, and G. C. Schatz, *Physical Review A* **98**, 062505 (2018).
- [290] M. A. Mosquera, L. O. Jones, C. H. Borca, M. A. Ratner, and G. C. Schatz, *The Journal of Physical Chemistry A* **123**, 4785–4795 (2019).
- [291] M. Bursch, J. Mewes, A. Hansen, and S. Grimme, *Angewandte Chemie International Edition* **61**, 10.1002/anie.202205735 (2022).
- [292] R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke, *The Journal of Chemical Physics* **68**, 3801–3807 (1978).
- [293] T. A. Wesolowski, *Physical Review A* **77**, 012504 (2008).
- [294] C. R. Jacob and J. Neugebauer, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 325–362 (2014).
- [295] C. R. Jacob and J. Neugebauer, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 325–362 (2014).
- [296] C. Huang, M. Pavone, and E. A. Carter, *The Journal of Chemical Physics* **134**, 154110 (2011).
- [297] M. H. Cohen and A. Wasserman, *Journal of Statistical Physics* **125**, 1121–1139 (2006).
- [298] O. Roncero, M. P. de Lara-Castells, P. Villarreal, F. Flores, J. Ortega, M. Paniagua, and A. Aguado, *The Journal of Chemical Physics* **129**, 184104 (2008).
- [299] S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, *Journal of Chemical Physics* **132**, 10.1063/1.3376251 (2010).
- [300] J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller, *Journal of Chemical Physics* **133**, 10.1063/1.3474575 (2010).
- [301] Q. Wu and W. Yang, *Journal of Chemical Physics* **118**, 2498–2509 (2003).

- [302] Q. Zhao, R. C. Morrison, and R. G. Parr, *Physical Review A* **50**, 2138–2142 (1994).
- [303] S. Nam, R. J. McCarty, H. Park, and E. Sim, *The Journal of Chemical Physics* **154**, 124122 (2021), arXiv:2008.08783.
- [304] T. Heaton-Burgess, F. A. Bulat, and W. Yang, *Physical Review Letters* **98**, 256401 (2007), arXiv:0610884 [cond-mat].
- [305] F. A. Bulat, T. Heaton-Burgess, A. J. Cohen, and W. Yang, *The Journal of Chemical Physics* **127**, 174101 (2007).
- [306] T. A. Wesolowski, P.-Y. Morgantini, and J. Weber, *The Journal of Chemical Physics* **116**, 6411–6421 (2002).
- [307] S. Lehtola and H. Jónsson, *Journal of Chemical Theory and Computation* **9**, 5365–5372 (2013).
- [308] D. Claudino and N. J. Mayhall, *Journal of Chemical Theory and Computation* **15**, 6085–6096 (2019).
- [309] S. J. R. Lee, M. Welborn, F. R. Manby, and T. F. Miller, *Accounts of Chemical Research* **52**, 1359–1368 (2019).
- [310] T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller, *The Journal of Chemical Physics* **139**, 024103 (2013).
- [311] B. Hégyely, P. R. Nagy, G. G. Ferenczy, and M. Kállay, *The Journal of Chemical Physics* **145**, 064107 (2016).
- [312] S. Huzinaga and A. A. Cantu, *The Journal of Chemical Physics* **55**, 5543–5549 (1971).
- [313] E. Francisco, A. Martín Pendás, and W. H. Adams, *The Journal of Chemical Physics* **97**, 6504–6508 (1992).
- [314] T. Shimazaki, K. Kitaura, D. G. Fedorov, and T. Nakajima, *The Journal of Chemical Physics* **146**, 084109 (2017).
- [315] D. V. Chulhai and J. D. Goodpaster, *Journal of Chemical Theory and Computation* **14**, 1928–1942 (2018).
- [316] D. V. Chulhai and J. D. Goodpaster, *Journal of Chemical Theory and Computation* **13**, 1503–1508 (2017).
- [317] J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller, *The Journal of Chemical Physics* **140**, 18A507 (2014).
- [318] D. S. Graham, X. Wen, D. V. Chulhai, and J. D. Goodpaster, *The Journal of Chemical Physics* **156**, 054112 (2022).

- [319] J. M. Foster and S. F. Boys, *Reviews of Modern Physics* **32**, 300–302 (1960).
- [320] C. Edmiston and K. Ruedenberg, *Reviews of Modern Physics* **35**, 457–464 (1963).
- [321] J. Pipek and P. G. Mezey, *The Journal of Chemical Physics* **90**, 4916–4926 (1989).
- [322] G. Knizia, *Journal of Chemical Theory and Computation* **9**, 4834–4843 (2013), arXiv:1306.6884.
- [323] R. C. Raffenetti, K. Ruedenberg, C. L. Janssen, and H. F. Schaefer, *Theoretica Chimica Acta* **86**, 149–165 (1993).
- [324] T. E. Abrudan, J. Eriksson, and V. Koivunen, *IEEE Transactions on Signal Processing* **56**, 1134–1147 (2008).
- [325] T. Abrudan, J. Eriksson, and V. Koivunen, *Signal Processing* **89**, 1704–1714 (2009).
- [326] I.-M. Høyvik, B. Jansik, and P. Jørgensen, *Journal of Chemical Theory and Computation* **8**, 3137–3146 (2012).
- [327] S. Lehtola and H. Jónsson, *Journal of Chemical Theory and Computation* **9**, 5365–5372 (2013).
- [328] J. E. Subotnik, Y. Shao, W. Liang, and M. Head-Gordon, *The Journal of Chemical Physics* **121**, 9220–9229 (2004).
- [329] S. Barison, D. E. Galli, and M. Motta, *Physical Review A* **106**, 022404 (2022), arXiv:2011.08137.
- [330] P. Löwdin, *The Journal of Chemical Physics* **18**, 365–375 (1950).
- [331] B. C. Carlson and J. M. Keller, *Physical Review* **105**, 102–103 (1957).
- [332] I. Mayer, *International Journal of Quantum Chemistry* **90**, 63–65 (2002).
- [333] A. E. Reed, R. B. Weinstock, and F. Weinhold, *The Journal of Chemical Physics* **83**, 735–746 (1985).
- [334] T. Giovannini and H. Koch, *Journal of Chemical Theory and Computation* **17**, 139–150 (2021), arXiv:2009.05330.
- [335] A. Ralli and M. I. Williams de la Bastida, *Nbed*, 2022.
- [336] A. Ralli, M. I. Williams, and P. V. Coveney, 1–24 (2022), arXiv:2203.01135.
- [337] H. Bergwerf, *MolView*, 2015.
- [338] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, *Nucleic Acids Research* **47**, D1102–D1109 (2019).

BIBLIOGRAPHY

- [339] A. D. Bochevarov and R. A. Friesner, *The Journal of Chemical Physics* **128**, 034102 (2008).
- [340] X. Yuan, L. Visscher, and A. S. P. Gomes, *The Journal of Chemical Physics* **156**, 224108 (2022), arXiv:2202.01146.
- [341] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, *The Journal of Chemical Physics* **152**, 124107 (2020), arXiv:1911.10864.
- [342] A. J. Cohen, P. Mori-Ssnchez, and W. Yang, *Chemical Reviews* **112**, 289–320 (2012).
- [343] J. P. T. Stenger, D. Gunlycke, and C. S. Hellberg, *Physical Review A* **105**, 022438 (2022), arXiv:2112.05063.
- [344] M. Rossmannek, P. K. Barkoutsos, P. J. Ollitrault, and I. Tavernelli, (2020), arXiv:2009.01872.

Appendix A

CS-VQE Appendix

A.1 Mapping Pauli operators to single-qubit Pauli Z operators

In Appendix A of [244] a proof is given on how to map a completely commuting set of Pauli operators to a single qubit Pauli Z operator. We summarise their approach here. Take a Pauli operator on n qubits to be: $P = \bigotimes_{j=0}^{n-1} \sigma_j^P$, where σ_j are single qubit Pauli operators. There are two cases we need to consider (diagonal and non-diagonal), with the goal to reduce the operators in $\mathcal{W}'_{all} \equiv \{q_0 G_0, q_1 G_1, \dots, q_{|\mathcal{G}|-1} G_{|\mathcal{G}|-1}, \xi P_0^{(k)}\}$ (Equation 5.21) to single-qubit Z Pauli operators.

For a non-diagonal Pauli operators $P_a \in \mathcal{W}'_{all}$, there must be at least one single qubit Pauli operator indexed by qubit k such that: $\sigma_k^{P_a} \in \{X, Y\}$. We can use this to define operator P_b that must anticommute with P_a :

$$\begin{aligned} P_a &= \left(\bigotimes_{j=0}^{k-1} \sigma_j^{P_a} \right) \otimes \sigma_k^{P_a} \otimes \left(\bigotimes_{j=k+1}^{n-1} \sigma_j^{P_a} \right) \\ P_b &= \left(\bigotimes_{j=0}^{k-1} \sigma_j^{P_a} \right) \otimes \sigma'_k \otimes \left(\bigotimes_{j=k+1}^{n-1} \sigma_j^{P_a} \right) \end{aligned} \quad \text{where } \{P_a, P_b\} = 0 \Leftrightarrow \sigma'_k = \begin{cases} X, & \text{if } \sigma_k^{P_a} = Y \\ Y, & \text{if } \sigma_k^{P_a} = X \end{cases}. \quad (\text{A.1})$$

These two Pauli operators differ by exactly one Pauli operator on qubit index k . We can define the rotation:

$$B = \exp\left(i\frac{\pi}{4}P_b\right) = \frac{1}{\sqrt{2}}(1 + iP_b). \quad (\text{A.2})$$

Conjugating P_a with this operator results in:

$$BP_a B^\dagger = \pm 1 \left(\bigotimes_{j=0}^{k-1} I_j \right) \otimes Z_k \otimes \left(\bigotimes_{j=k+1}^{n-1} I_j \right) = P'_a, \quad (\text{A.3})$$

and P_a has been mapped to a single qubit Pauli Z operator.

For diagonal operators $P_c \in \mathcal{W}'_{all}$, all the n -fold tensor products of single qubit Pauli operators must be either Z or I : $P_c = \bigotimes_{j=0}^{n-1} \sigma_j^{P_c}$ where $\sigma_j^{P_c} \in \{I, Z\} \forall i$. Since

\mathcal{W}'_{all} (equation 5.21) is an independent set, for all the rotated P'_a there must be at least one index l such that $\sigma_l^{P'_a} = I$ and $\sigma_l^{P_c} = Z$. We denote this operator P_c . We also define a new operator P_d from this, which only acts non-trivially on the l -th qubit with a single qubit Y . To summarise:

$$\begin{aligned}
P_c &= \left(\bigotimes_{j=0}^{l-1} \sigma_j^{P_c} \right) \otimes Z_l \otimes \left(\bigotimes_{j=l+1}^{n-1} \sigma_j^{P_c} \right) \text{ where } \sigma_j^{P_c} \in \{I, Z\} \forall i \\
P_d &= \left(\bigotimes_{j=0}^{l-1} I_j \right) \otimes Y_l \otimes \left(\bigotimes_{j=l+1}^{n-1} I_j \right) \\
P'_a &= \left(\bigotimes_{j=0}^{l-1} \sigma_j^{P'_a} \right) \otimes I_l \otimes \left(\bigotimes_{j=l+1}^{n-1} \sigma_j^{P'_a} \right) \text{ where } [P_d, P'_a] = 0 \text{ and } \sigma_j^{P'_a} \in \{I, Z\} \forall j
\end{aligned} \tag{A.4}$$

We can define the rotation:

$$D = \exp\left(i\frac{\pi}{4}P_d\right) = \frac{1}{\sqrt{2}}(1 + iP_d). \tag{A.5}$$

Conjugating P_c with this operator results in:

$$DP_cD^\dagger = \pm 1 \left(\bigotimes_{j=0}^{l-1} \sigma_j^{P_c} \right) \otimes X_l \otimes \left(\bigotimes_{j=l+1}^{n-1} \sigma_j^{P_c} \right) = P'_c. \tag{A.6}$$

P'_c is now a non-diagonal Pauli operator (contains a single qubit X acting on qubit l). This operator P'_c can now be mapped to a single qubit Z operator using a further $\frac{\pi}{2}$ -rotation following the previously given procedure for non-diagonal Pauli operators.

The V_i operators in the main text (equation 5.22) are defined by these $\frac{\pi}{2}$ -rotations, such that each $q_i G_i$ and $P_0^{(k)}$ is mapped to a single qubit Pauli Z term. At worst, two $\frac{\pi}{2}$ -rotations are needed for every operator in \mathcal{W}'_{all} (Equation 5.21), which occurs when all operators in \mathcal{W}'_{all} are diagonal.

A.2 CS-VQE via a sequence of rotations

In this subsection, we show how the sequence of rotations step in CS-VQE can cause an exponential increasing in the number of Pauli operators of a given Hamiltonian when rotating into the contextual subspace [198, 221, 262]. We use the following shorthand notation, where Pauli operators with multiple indices represent the multiplication of Pauli operators: $P_a P_b P_c = P_{abc}$.

Given the set of anticommuting operators $A(\vec{r})$ (Equation 5.11), we can define the

following self-inverse operators:

$$\{\mathcal{X}_{kj} = iP_0^{(k)}P_0^{(j)} \quad \forall P_0^{(j)} \in \mathcal{A} \text{ where } j \neq k\}, \quad (\text{A.7})$$

where $P_0^{(k)} \in \mathcal{A}$. To simplify the notation we drop the subscript 0 (denoting the first operator in a clique) and write each $P_0^{(k)}$, $P_0^{(j)}$ as P_k and P_j respectively.

The adjoint rotation generated by one of these \mathcal{X}_{kj} operators will be:

$$\begin{aligned} e^{(-i\frac{\theta_{kj}}{2}\mathcal{X}_{kj})}A(\vec{r})e^{(+i\frac{\theta_{kj}}{2}\mathcal{X}_{kj})} &= R_{S_{kj}}(\theta_{kj})A(\vec{r})R_{S_{kj}}^\dagger(\theta_{kj}) \\ &= (r_j \cos \theta_{kj} - r_k \sin \theta_{kj})P_j + (\beta_j \sin \theta_{kj} + r_k \cos \theta_{kj})P_k + \sum_{\substack{P_l \in \mathcal{A} \\ \forall l \neq k, j}} \beta_l P_l. \end{aligned} \quad (\text{A.8})$$

The coefficient of P_j can be made to go to 0, by setting $r_j \cos \theta_{kj} = r_k \sin \theta_{kj}$. This approach removes the term with index j and increases the coefficient of P_k from $r_k \mapsto \sqrt{r_k^2 + r_j^2}$ [198]. This process is repeated over all indices excluding $j = k$ until only the P_k term remains. This procedure can be concisely written using the following operator [198]:

$$R_S = \prod_{\substack{j=0 \\ \forall j \neq k}}^{|\mathcal{A}|-1} e^{(-i\frac{\theta_{kj}}{2}\mathcal{X}_{kj})} = \prod_{\substack{j=0 \\ \forall j \neq k}}^{|\mathcal{A}|-1} R_{S_{kj}}(\theta_{kj}) = \prod_{\substack{j=0 \\ \forall j \neq k}}^{|\mathcal{A}|-1} \left[\cos\left(\frac{\theta_{kj}}{2}\right)\mathcal{I} - i \sin\left(\frac{\theta_{kj}}{2}\right)\mathcal{X}_{kj} \right], \quad (\text{A.9})$$

which is simply a sequence of rotations. The angle θ_{kj} is defined recursively at each step of the removal process, as the coefficient of P_k increases at each step and thus must be taken into account. The correct solution for θ_{kj} must be chosen given the signs of r_k and r_k [198]. The overall action of this sequence of rotations is:

$$R_S A(\vec{r}) R_S^\dagger = P_k. \quad (\text{A.10})$$

Looking at Equation A.9, expanding the product of rotations results in R_S containing $\mathcal{O}(2^{|\mathcal{A}|-1})$ Pauli operators. We write this operator as:

$$R_S = \sum_b^{\mathcal{O}(2^{|\mathcal{A}|-1})} \delta_b P_b. \quad (\text{A.11})$$

The adjoint rotation of R_S on a general Hamiltonian $H_q = \sum_a^{|H_q|} c_a P_a$ is:

$$\begin{aligned} R_S H_q R_S^\dagger &= \left(\sum_b^{\mathcal{O}(2^{|\mathcal{A}|-1})} \delta_b P_b \right) \sum_a^{|H_q|} c_a P_a \left(\sum_c^{\mathcal{O}(2^{|\mathcal{A}|-1})} \delta_c^* P_c \right) \\ &= \sum_b^{\mathcal{O}(2^{|\mathcal{A}|-1})} \sum_a^{|H_q|} \sum_c^{\mathcal{O}(2^{|\mathcal{A}|-1})} (\delta_b c_a \delta_c^*) P_b P_a P_c. \end{aligned} \quad (\text{A.12})$$

We see that the number of terms increases as $\mathcal{O}(2^{|\mathcal{A}|}|H_q|)$ which was previously shown in [256]. What we show next is the additional structure in R_S - due to the X_{kj} operators - mean that the base of the exponent can be slightly lower; however, it still remains exponential in $|\mathcal{A}|$.

Consider the adjoint rotation of a particular X_{kj} in R_S (Equation A.9):

$$\begin{aligned} R_{S_{kj}} &= \cos\left(\frac{\theta_{kj}}{2}\right) \mathcal{I} + \sin\left(\frac{\theta_{kj}}{2}\right) P_{kj}, \\ R_{S_{kj}}^\dagger &= \cos\left(\frac{\theta_{kj}}{2}\right) \mathcal{I} + \sin\left(\frac{\theta_{kj}}{2}\right) P_{jk}. \end{aligned} \quad (\text{A.13})$$

Performing the adjoint rotation on H_q results in the following:

$$\begin{aligned} R_{S_{kj}} H_q R_{S_{kj}}^\dagger &= \left[\alpha_{kj} \mathcal{I} + \beta_{kj} P_{kj} \right] \sum_a c_a P_a \left[\alpha_{kj} \mathcal{I} + \beta_{kj} P_{jk} \right] \\ &= \sum_a c_a (\alpha_{kj} P_a + \beta_{kj} P_{kj} P_a) \left[\alpha_{kj} \mathcal{I} + \beta_{kj} P_{jk} \right] \\ &= \sum_a c_a (\alpha_{kj}^2 P_a + \alpha_{kj} \beta_{kj} P_a P_{jk} + \alpha_{kj} \beta_{kj} P_{kj} P_a + \beta_{kj}^2 P_{kj} P_a P_{jk}) \\ &= \sum_a c_a (\alpha_{kj}^2 P_a + \alpha_{kj} \beta_{kj} \underline{P_a P_{jk}} - \alpha_{kj} \beta_{kj} \underline{P_{jk} P_a} + \beta_{kj}^2 P_{kj} P_a P_{jk}) \\ &= \sum_a c_a (\alpha_{kj}^2 P_a + \alpha_{kj} \beta_{kj} [P_a, P_{jk}] + \beta_{kj}^2 P_{kj} P_a P_{jk}) \\ &= \sum_a c_a \begin{cases} (\alpha_{kj}^2 P_a + \beta_{kj}^2 P_{kj} P_a P_{jk}), & \text{if } [P_a, P_{jk}] = 0 \\ (\alpha_{kj}^2 P_a + 2\alpha_{kj} \beta_{kj} P_a P_{jk} + \beta_{kj}^2 P_{kj} P_a P_{jk}), & \text{else } \{P_a, P_{jk}\} = 0. \end{cases} \end{aligned} \quad (\text{A.14})$$

When $[P_a, P_{jk}] = 0$, we get:

$$\begin{aligned}
\sum_a c_a (\alpha_{kj}^2 P_a + \beta_{kj}^2 P_{kj} \underline{P_a P_{jk}}) &= \sum_a c_a (\alpha_{kj}^2 P_a + \beta_{kj}^2 \underline{P_{kj} P_{jk}} P_a) \\
&= \sum_a c_a (\alpha_{kj}^2 P_a + \beta_{kj}^2 P_a) \\
&= \sum_a c_a (\alpha_{kj}^2 + \beta_{kj}^2) P_a \\
&= \sum_a c_a P_a.
\end{aligned} \tag{A.15}$$

When $\{P_a, P_{jk}\} = 0$, we find:

$$\begin{aligned}
\sum_a c_a (\alpha_{kj}^2 P_a + 2\alpha_{kj}\beta_{kj} P_a P_{jk} + \beta_{kj}^2 P_{kj} \underline{P_a P_{jk}}) &= \sum_a c_a (\alpha_{kj}^2 P_a + 2\alpha_{kj}\beta_{kj} P_a P_{jk} - \beta_{kj}^2 \underline{P_{kj} P_{jk}} P_a) \\
&= \sum_a c_a (\alpha_{kj}^2 P_a + 2\alpha_{kj}\beta_{kj} P_a P_{jk} - \beta_{kj}^2 P_a) \\
&= \sum_a c_a (\alpha_{kj}^2 P_a + \sin(\theta_{kj}) P_a P_{jk} - \beta_{kj}^2 P_a) \\
&= \sum_a c_a ((\alpha_{kj}^2 - \beta_{kj}^2) P_a + \sin(\theta_{kj}) P_a P_{jk}) \\
&= \sum_a c_a (\cos(\theta_{kj}) P_a + \sin(\theta_{kj}) P_a P_{jk}).
\end{aligned} \tag{A.16}$$

Both cases use the following identities:

$$\begin{aligned}
\alpha_{kj}^2 - \beta_{kj}^2 &= \cos^2\left(\frac{\theta_{kj}}{2}\right) - \sin^2\left(\frac{\theta_{kj}}{2}\right) = \cos(\theta_{kj}), \\
\alpha_{kj}^2 + \beta_{kj}^2 &= 1, \\
2\alpha_{kj}\beta_{kj} &= 2\cos\left(\frac{\theta_{kj}}{2}\right)\sin\left(\frac{\theta_{kj}}{2}\right) = \sin(\theta_{kj}),
\end{aligned} \tag{A.17}$$

where $\alpha_{kj} = \cos\left(\frac{\theta_{kj}}{2}\right)$ and $\beta_{kj} = \sin\left(\frac{\theta_{kj}}{2}\right)$. Using these results Equation A.14 reduces to:

$$\begin{aligned}
R_{S_{kj}} H_q R_{S_{kj}}^\dagger &= \sum_{\forall [P_a, P_{jk}] = 0} c_a P_a + \sum_{\forall \{P_a, P_{jk}\} = 0} c_a \left(\cos(\theta_{kj}) P_a + \sin(\theta_{kj}) P_a P_{jk} \right) \\
&= \sum_a \eta_a P_a + \sum_{\forall \{P_a, P_{jk}\} = 0} \eta_a (P_{jk} P_a),
\end{aligned} \tag{A.18}$$

where η_a represent the new real coefficients.

Consider the application of the next rotation operator R_{kl} in R_S (note k index

represents the same Pauli operator P_k):

$$R_{S_{kl}} R_{S_{kj}} H_q R_{S_{kj}}^\dagger R_{S_{kl}}^\dagger = R_{S_{kl}} \left(\sum_a \eta_a P_a \right) R_{S_{kl}}^\dagger + \underbrace{R_{S_{kl}} \left(\sum_{\forall \{P_a, P_{jk}\}=0} \eta_a (P_{jk} P_a) \right) R_{S_{kl}}^\dagger}_{\text{(A.19)}} .$$

Focusing on the last term in Equation A.19:

$$\begin{aligned} R_{S_{kl}} \left(\sum_{\forall \{P_a, P_{jk}\}=0} \eta_a (P_{jk} P_a) \right) R_{S_{kl}}^\dagger &= \left[\gamma_{kl} \mathcal{I} + \delta_{kl} P_{kl} \right] \sum_{\forall \{P_a, P_{jk}\}=0} \eta_a (P_{jk} P_a) \left[\gamma_{kl} \mathcal{I} + \delta_{kl} P_{lk} \right] \\ &= \sum_{\forall \{P_a, P_{jk}\}=0} \eta_a \left(\gamma_{kl} P_{jk} P_a + \delta_{kl} P_{kl} P_{jk} P_a \right) \left[\gamma_{kl} \mathcal{I} + \delta_{kl} P_{lk} \right] \\ &= \sum_{\forall \{P_a, P_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + \gamma_{kl} \delta_{kl} P_{jk} P_a P_{lk} + \gamma_{kl} \delta_{kl} P_{kl} P_{jk} P_a + \delta_{kl}^2 P_{kl} P_{jk} P_a P_{lk} \right) \\ &= \sum_{\forall \{P_a, P_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + \gamma_{kl} \delta_{kl} P_{jk} P_a P_{lk} - \gamma_{kl} \delta_{kl} P_{kj} P_{lk} P_a + \delta_{kl}^2 P_{kl} P_{jk} P_a P_{lk} \right) \\ &= \sum_{\forall \{P_a, P_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + \gamma_{kl} \delta_{kl} P_{jk} P_a P_{lk} + \gamma_{kl} \delta_{kl} P_{jk} P_{lk} P_a + \delta_{kl}^2 P_{kl} P_{jk} P_a P_{lk} \right) \\ &= \sum_{\forall \{P_a, P_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + \gamma_{kl} \delta_{kl} P_{jk} \{P_a, P_{lk}\} + \delta_{kl}^2 P_{kl} P_{jk} P_a P_{lk} \right) \\ &= \sum_{\forall \{P_a, P_{jk}\}=0} \eta_a \begin{cases} \left(\gamma_{kl}^2 P_{jk} P_a + 2\gamma_{kl} \delta_{kl} P_{jk} P_a P_{lk} + \delta_{kl}^2 P_{kl} P_{jk} P_a P_{lk} \right), & \text{if } [P_a, P_{lk}] = 0 \\ \left(\gamma_{kl}^2 P_{jk} P_a + \delta_{kl}^2 P_{kl} P_{jk} P_a P_{lk} \right), & \text{if } \{P_a, P_{lk}\} = 0 \end{cases} \end{aligned} \quad \text{(A.20)}$$

For the case $\{P_a, P_{lk}\} = 0$:

$$\begin{aligned}
\sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + \delta_{kl}^2 P_{kl} P_{jk} \underline{P_a P_{lk}} \right) &= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a - \delta_{kl}^2 \underline{P_{kl} P_{jk}} P_{lk} P_a \right) \\
&= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + \delta_{kl}^2 P_{jk} P_{kl} P_{lk} P_a \right) \\
&= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + \delta_{kl}^2 P_{jk} P_a \right) \\
&= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(P_{jk} P_a \right).
\end{aligned} \tag{A.21}$$

We observe that there is no increase in the number of terms and the weight of each Pauli operator changes.

For the case $[P_a, P_{lk}] = 0$:

$$\begin{aligned}
\sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + 2\gamma_{kl}\delta_{kl} P_{jk} P_a P_{lk} + \delta_{kl}^2 P_{kl} P_{jk} \underline{P_a P_{lk}} \right) \\
&= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + 2\gamma_{kl}\delta_{kl} P_{jk} P_a P_{lk} + \delta_{kl}^2 \underline{P_{kl} P_{jk}} P_{lk} P_a \right) \\
&= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + 2\gamma_{kl}\delta_{kl} P_{jk} P_a P_{lk} - \delta_{kl}^2 P_{jk} P_{kl} P_{lk} P_a \right) \\
&= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\gamma_{kl}^2 P_{jk} P_a + 2\gamma_{kl}\delta_{kl} P_{jk} P_a P_{lk} - \delta_{kl}^2 P_{jk} P_a \right) \\
&= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left((\gamma_{kl}^2 - \delta_{kl}^2) P_{jk} P_a + 2\gamma_{kl}\delta_{kl} P_{jk} P_a P_{lk} \right) \\
&= \sum_{\forall\{P_a, \overset{a}{P}_{jk}\}=0} \eta_a \left(\cos(\theta_{kl}) P_{jk} P_a + \sin(\theta_{kl}) P_{jk} P_a P_{lk} \right).
\end{aligned} \tag{A.22}$$

The number of terms in the resulting operator has increased for each case where

$[P_a, P_{lk}] = 0$. The action of two rotations of R_S on the whole Hamiltonian results in:

$$\begin{aligned}
R_{S_{kl}} R_{S_{kj}} H R_{S_{kj}}^\dagger R_{S_{kl}}^\dagger &= R_{S_{kl}} \left(\sum_a \eta_a P_a \right) R_{S_{kl}}^\dagger + R_{S_{kl}} \left(\sum_{\substack{a \\ \forall \{P_a, P_{jk}\}=0}} \eta_a (P_{jk} P_a) \right) R_{S_{kl}}^\dagger \\
&= R_{S_{kl}} \left(\sum_a \eta_a P_a \right) R_{S_{kl}}^\dagger + \sum_{\substack{a \\ \forall \{P_a, P_{jk}\}=0 \\ \forall \{P_a, P_{lk}\}=0}} \eta_a P_{jk} P_a + \\
&\quad \sum_{\substack{a \\ \forall \{P_a, P_{jk}\}=0 \\ [P_a, P_{lk}]=0}} \eta_a (\cos(\theta_{kl}) P_{jk} P_a + \sin(\theta_{kl}) P_{jk} P_a P_{lk}) \\
&= R_{S_{kl}} \left(\sum_a \eta_a P_a \right) R_{S_{kl}}^\dagger + \sum_{\substack{a \\ \forall \{P_a, P_{jk}\}=0}} \mu_a P_{jk} P_a + \sum_{\substack{a \\ \forall \{P_a, P_{jk}\}=0 \\ [P_a, P_{lk}]=0}} \mu_a P_{jk} P_a P_{lk} \\
&= \sum_a \nu_a P_a + \sum_{\substack{a \\ \forall \{P_a, P_{lk}\}=0}} \nu_a (P_{lk} P_a) + \sum_{\substack{a \\ \forall \{P_a, P_{jk}\}=0}} \mu_a P_{jk} P_a + \sum_{\substack{a \\ \forall \{P_a, P_{jk}\}=0 \\ [P_a, P_{lk}]=0}} \mu_a P_{jk} P_a P_{lk},
\end{aligned} \tag{A.23}$$

where Greek letters are new coefficients according to the expansion. We use the results of equations A.21 and A.22 to determine what occurs to the second term of Equation A.23. We have applied the result in Equation A.18 to the first term ($R_{S_{kl}} \left(\sum_i \eta_i P_i \right) R_{S_{kl}}^\dagger$) in Equation A.23.

From these results, we can infer how the terms in H_q will scale for a general sequence of rotations of size $|R_S|$ (Equation A.9), which in general change as:

$$|H_q| \sum_{g=0}^{|R_S|} \binom{|R_S|}{g} = 2^{|R_S|} |H_q|. \tag{A.24}$$

This operation increases the number of terms in H_q to $\mathcal{O}(2^{(|A|-1)} |H_q|)$. However, the structure of the sequence of rotation operator actually requires 2^g commuting/anticommuting conditions to be met for new Pauli operators to be generated by subsequent rotations. We therefore need to consider the probability that a given Pauli operator will either commute or anticommute with another. For the case of single qubit Pauli matrices $\sigma_a, \sigma_b \in \{I, X, Y, Z\}$ by a simple counting argument $P([\sigma_a, \sigma_b] = 0) = \frac{5}{8}$ and $P(\{\sigma_a, \sigma_b\} = 0) = \frac{3}{8}$, for Pauli matrices selected uniformly at random. Generalising this to tensor products of Pauli matrices on n qubits, for a Pauli operator to anticommute with another there needs to be an odd number of anticommuting tensor factors. First, consider the binomial distribution:

$$P(x) = \binom{n}{x} p^x q^{n-x}, \tag{A.25}$$

where n is the number of trials (repeated experiments), p is the probability of success - here the probability a single Pauli matrix anticommutes with another ($p = \frac{3}{8}$) - and q is the probability of failure - here the probability a single Pauli matrix commutes with another ($q = \frac{5}{8}$). Under these conditions, $P(x)$ gives the probability that two n -fold Pauli operators, selected uniformly at random, anticommute in x -many tensor factors. Therefore, the probability of two uniformly random Pauli operators anticommute (commute) is given as a sum over odd (even) values of $x \leq n$:

$$P(\{P_a, P_b\} = 0) = \sum_{c=1}^{\lceil n/2 \rceil} P(2c-1). \quad (\text{A.26})$$

Now, the binomial theorem states

$$(p+q)^n = \sum_{c=0}^n \binom{n}{c} p^c q^{n-c}, \quad (\text{A.27})$$

for any $p, q \in \mathbb{R}$. We can define the following difference:

$$\begin{aligned} (p+q)^n - (-p+q)^n &= \sum_{c=0}^n \binom{n}{c} \underbrace{[1 - (-1)^c]}_{\begin{cases} 2, & \text{if } c \text{ odd} \\ 0, & \text{if } c \text{ even} \end{cases}} p^c q^{n-c} = 2 \sum_{c=1}^{\lceil n/2 \rceil} \binom{n}{2c-1} p^{2c-1} q^{n-(2c-1)}. \end{aligned} \quad (\text{A.28})$$

Overall we find the probability that two n -fold Pauli operators anticommute to be:

$$\begin{aligned} P(\{P_a, P_b\} = 0) &= \sum_{c=1}^{\lceil n/2 \rceil} P(2c-1) \\ &= \sum_{c=1}^{\lceil n/2 \rceil} \binom{n}{2c-1} \cdot \left(\frac{3}{8}\right)^{2c-1} \cdot \left(\frac{5}{8}\right)^{n-(2c-1)} \\ &= \frac{1}{2} \left[\left(\frac{3}{8} + \frac{5}{8}\right)^n - \left(-\frac{3}{8} + \frac{5}{8}\right)^n \right] \\ &= \frac{1}{2} \left[1 - \left(\frac{1}{4}\right)^n \right], \end{aligned} \quad (\text{A.29})$$

when each operator P_a, P_b is chosen uniformly at random. The n choose $2c-1$ term in equation A.29 counts all the possible ways an odd number of single qubit pairs of Pauli tensor factors can differ on n qubits, the first fraction gives the probability that there are $2c-1$ anticommuting terms on each pair of qubits and the final fraction gives the probability that the remaining $n-(2c-1)$ qubit positions pairwise commute on each qubit. The penultimate line of equation A.29 uses the definition in A.28, with the factor of two taken into account. Through equation A.29, it can be seen that the

probability of two n -fold Pauli operators anticommuting quickly converges to 0.5 as the number of qubits n increases. The motivation for A.28 arises from observing that the quantity we subtract, $(1/4)^n$, is the probability of obtaining an n -fold identity operator, which has the unique property of commuting universally. The complement $1 - (1/4)^n$ therefore corresponds with the probability of selecting uniformly at random a Pauli operator with at least one non-trivial tensor factor. After discounting identity operators from consideration, the probabilities of anticommuting or commuting coincide, hence each occurs half of the time, explaining the $1/2$ factor in A.29; the probability bias towards commutation is a consequence of the identity operator commuting universally, whereas there is no such operator that can anticommute universally.

If we consider how the number of terms in H_q changes upon the sequence of rotations transformation: $H_q \mapsto R_S H_q R_S^\dagger$ where terms either commute or anticommute with a probability of 0.5, then the scaling is as follows:

$$\sum_{g=0}^{|R_S|} \frac{|H_q|}{2^g} \binom{|R_S|}{g} = \left(\frac{3}{2}\right)^{|R_S|} |H_q|. \quad (\text{A.30})$$

Equation A.24 is modified to have a constant factor of 2^{-g} , where g represents the number of commuting or anticommuting conditions required for operators in H_q to obey in order to increase the number of terms upon a rotation of R_S . Here each condition is assumed to occur with a probability of 0.5. This operation increases the number of terms in H_q to $\mathcal{O}(1.5^{(|\mathcal{A}|-1)} |H_q|)$. Note $|R_S| = |\mathcal{A}| - 1$. In general, the scaling will be $\mathcal{O}(x^{(|\mathcal{A}|-1)} |H_q|)$ where $1 \leq x \leq 2$, depending on how each rotation in the sequence of rotations commutes with terms in H_q . The $x = 1$ case occurs if each rotation in R_S commutes with the whole Hamiltonian. Apart from this special case, the number of terms in H_q will increase exponentially with the size of \mathcal{A} or equivalently with the number of qubits n (as $|\mathcal{A}| \leq 2n + 1$ [256]) when R is defined by a sequence of rotations.

A.3 CS-VQE via a linear combination of unitaries

Here, we analyse the LCU approach of CS-VQE when mapping a Hamiltonian into the contextual subspace. We consider the set of anticommuting Pauli operators making up $A(\vec{r})$ (Equation 5.11). We can rewrite this Equation, with the term we are reducing to $(r_k P_0^{(k)})$ outside the sum:

$$A(\vec{r}) = r_k P_0^{(k)} + \sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} r_j P_0^{(j)}. \quad (\text{A.31})$$

To simplify the notation we drop the subscript 0 (denoting the first operator in a clique) and write each $P_0^{(k)}$, $P_0^{(j)}$ as P_k and P_j respectively.

A re-normalization can be performed on the remaining sum yielding:

$$\begin{aligned} A(\vec{r}) &= r_k P_k + \Omega \sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} \delta_j P_j \\ &= r_k P_k + \Omega H_{\mathcal{A} \setminus \{r_k P_k\}}, \end{aligned} \tag{A.32}$$

where:

$$\sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} |\delta_j|^2 = 1, \tag{A.33a}$$

$$r_j = \Omega \delta_j, \tag{A.33b}$$

$$H_{\mathcal{A} \setminus \{r_k P_k\}} = \sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} \delta_j P_j. \tag{A.33c}$$

Using the Pythagorean trigonometric identity: $\sin^2(x) + \cos^2(x) = 1$, $A(\vec{r})$ can be re-written as:

$$\begin{aligned} A(\vec{r}) &= \cos(\phi_k) P_k + \sin(\phi_k) \sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} \delta_j P_j \\ &= \cos(\phi_k) P_k + \sin(\phi_k) H_{\mathcal{A} \setminus \{r_k P_k\}}. \end{aligned} \tag{A.34}$$

Comparing Equations A.32 and A.34, it is clear that $\cos(\phi_k) = r_k$ and $\sin(\phi_k) = \Omega$.

It was shown in [198] that one can consider rotations of $A(\vec{r})$ around an axis that is Hilbert-Schmidt orthogonal to both $H_{\mathcal{A} \setminus \{r_k P_k\}}$ and P_k :

$$\mathcal{X} = \frac{i}{2} [H_{\mathcal{A} \setminus \{r_k P_k\}}, P_k] = i \sum_{\substack{j=0 \\ \forall j \neq k}}^{|H_{\mathcal{A} \setminus \{r_k P_k\}}|^{-1}} \delta_j P_j P_k. \tag{A.35}$$

\mathcal{X} anticommutes with \mathcal{A} and is self-inverse [198] - see equation 4.44. We can therefore define the rotation [198, 221]:

$$R_{LCU} = e^{(-i\frac{\alpha}{2}\mathcal{X})} = \cos\left(\frac{\alpha}{2}\right)\mathcal{I} - i\sin\left(\frac{\alpha}{2}\right)\mathcal{X} \quad (\text{A.36a})$$

$$= \cos\left(\frac{\alpha}{2}\right)\mathcal{I} - i\sin\left(\frac{\alpha}{2}\right)\left(i\sum_{\substack{j=0 \\ \forall j \neq k}}^{|H_{\mathcal{A} \setminus \{r_k P_k\}}|-1} \delta_j P_j P_k\right) \quad (\text{A.36b})$$

$$= \cos\left(\frac{\alpha}{2}\right)\mathcal{I} + \sin\left(\frac{\alpha}{2}\right)\sum_{\substack{j=0 \\ \forall j \neq k}}^{|H_{\mathcal{A} \setminus \{r_k P_k\}}|-1} \delta_j P_{jk}, \quad (\text{A.36c})$$

$$= \delta_{\mathcal{I}}\mathcal{I} + \sum_{\substack{j=0 \\ \forall j \neq k}}^{|H_{\mathcal{A} \setminus \{r_k P_k\}}|-1} \delta_j P_{jk}. \quad (\text{A.36d})$$

The conjugate rotation will be:

$$R_{LCU}^\dagger = \cos\left(\frac{\alpha}{2}\right)\mathcal{I} + i\sin\left(\frac{\alpha}{2}\right)i\sum_{\substack{j=0 \\ \forall j \neq k}}^{|H_{\mathcal{A} \setminus \{r_k P_k\}}|-1} \delta_j P_j P_k, \quad (\text{A.37a})$$

$$= \cos\left(\frac{\alpha}{2}\right)\mathcal{I} + \sin\left(\frac{\alpha}{2}\right)\sum_{\substack{j=0 \\ \forall j \neq k}}^{|H_{\mathcal{A} \setminus \{r_k P_k\}}|-1} \delta_j \underbrace{P_k P_j}_{\text{order change}}, \quad (\text{A.37b})$$

$$= \delta_{\mathcal{I}}\mathcal{I} + \sum_{\substack{j=0 \\ \forall j \neq k}}^{|H_{\mathcal{A} \setminus \{r_k P_k\}}|-1} \delta_j P_{kj}. \quad (\text{A.37c})$$

Note the different order of j and k for R_{LCU} and R_{LCU}^\dagger . The adjoint action of R_{LCU} on $A(\vec{r})$ is:

$$R_{LCU}A(\vec{r})R_{LCU}^\dagger = \cos(\phi_k - \alpha)P_k + \sin(\phi_k - \alpha)H_{\mathcal{A} \setminus \{r_k P_k\}}. \quad (\text{A.38})$$

By choosing $\alpha = \phi_k$, the following transformation occurs $R_{LCU}A(\vec{r})R_{LCU}^\dagger = P_k$ [198, 221]. This fully defines the R_{LCU} operator required by unitary partitioning. Next, we need to consider the use of this operator in CS-VQE.

The adjoint action of R_{LCU} on a general Hamiltonian $H_q = \sum_i^{|H_q|} c_i P_i$ is:

$$R_{LCU}H_qR_{LCU}^\dagger = \left(\delta_{\mathcal{I}}\mathcal{I} + \sum_j^{|R_{LCU}|-1} \delta_j P_{jk} \right) \sum_i^{|H_q|} c_i P_i \left(\delta_{\mathcal{I}}\mathcal{I} + \sum_l^{|R_{LCU}|-1} \delta_l P_{kl} \right) \quad (\text{A.39a})$$

$$= \left(\delta_{\mathcal{I}} \sum_i^{|H_q|} c_i P_i + \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_j c_i P_{jk} P_i \right) \left(\delta_{\mathcal{I}}\mathcal{I} + \sum_l^{|R_{LCU}|-1} \delta_l P_{kl} \right) \quad (\text{A.39b})$$

$$= \delta_{\mathcal{I}}^2 \sum_i^{|H_q|} c_i P_i \quad (\text{A.39c})$$

$$+ \sum_l^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} c_i \delta_l P_i P_{kl} + \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} \delta_j c_i P_{jk} P_i \quad (\text{A.39d})$$

$$+ \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_l^{|R_{LCU}|-1} \delta_j c_i \delta_l P_{jk} P_i P_{kl}. \quad (\text{A.39e})$$

We can rewrite the final term (Equation A.39e) as:

$$\sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_l^{|R_{LCU}|-1} \delta_j c_i \delta_l P_{jk} P_i P_{kl} = \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{\substack{l=j \\ [P_{jk}, P_i]=0}}^{|R_{LCU}|-1} (c_i \delta_j \delta_j^*) P_i \quad (\text{A.40a})$$

$$+ \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{\substack{l=j \\ \{P_{jk}, P_i\}=0}}^{|R_{LCU}|-1} (-c_i \delta_j \delta_j^*) P_i \quad (\text{A.40b})$$

$$+ \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{l \neq j}^{|R_{LCU}|-1} (\delta_j c_i \delta_l) P_{jk} P_i P_{kl}. \quad (\text{A.40c})$$

Here we have applied the identity of conjugating a Pauli operator P_u with another Pauli operator P_v resulting in two cases:

$$P_v P_u P_v = \begin{cases} P_u, & \text{if } [P_v, P_u] = 0 \\ -P_u, & \text{otherwise } \{P_v, P_u\} = 0 \end{cases}. \quad (\text{A.41})$$

Focusing on the last term of Equation A.40, we can simplify A.40c as j and l run over the same indices we can re-write each $l \neq j$ sum as $l > j$ and expand into two terms:

$$\sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{l \neq j}^{|R_{LCU}|-1} (\delta_j c_i \delta_l) P_{jk} P_i P_{kl} = \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{l > j}^{|R_{LCU}|-1} (\delta_j c_i \delta_l) (P_{jk} P_i P_{kl} + P_{lk} P_i P_{kj}). \quad (\text{A.42})$$

We can expand then expand this equation into the four cases for when:

1. $[P_{jk}, P_i] = 0$ and $[P_{lk}, P_i] = 0$
2. $[P_{jk}, P_i] = 0$ and $\{P_{lk}, P_i\} = 0$
3. $\{P_{jk}, P_i\} = 0$ and $[P_{lk}, P_i] = 0$
4. $\{P_{jk}, P_i\} = 0$ and $\{P_{lk}, P_i\} = 0$

For the first case and last case:

$$\begin{aligned}
& \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{l>j}^{|R_{LCU}|-1} (\delta_j c_i \delta_l) \left(\underline{P_{jk} P_i P_{kl}} + \underline{P_{lk} P_i P_{kj}} \right) \\
&= \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{l>j}^{|H_q|} (\delta_j c_i \delta_l) \left(\pm P_i \underline{P_{jk} P_{kl}} \pm P_i \underline{P_{lk} P_{kj}} \right) \\
&= \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{l>j}^{|H_q|} (\delta_j c_i \delta_l) \left(\pm P_i P_j P_l \pm P_i P_l P_j \right) \\
&= \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{l>j}^{|H_q|} (\delta_j c_i \delta_l) \pm P_i \{P_j, P_l\} \\
&= 0.
\end{aligned} \tag{A.43}$$

Whereas, for the second and third cases:

$$\begin{aligned}
& \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{l>j}^{|R_{LCU}|-1} (\delta_j c_i \delta_l) \left(\underline{P_{jk} P_i P_{kl}} + \underline{P_{lk} P_i P_{kj}} \right) \\
&= \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{l>j}^{|H_q|} (\delta_j c_i \delta_l) \left(\pm P_i \underline{P_{jk} P_{kl}} \mp P_i \underline{P_{lk} P_{kj}} \right) \\
&= \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{l>j}^{|H_q|} (\delta_j c_i \delta_l) \left(\pm P_i P_j P_l \mp P_i \underline{P_l P_j} \right) \\
&= \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{l>j}^{|H_q|} (\delta_j c_i \delta_l) \left(\pm P_i P_j P_l \pm P_i P_j P_l \right) \\
&= \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{l>j}^{|H_q|} (\delta_j c_i \delta_l) \pm 2P_i P_j P_l.
\end{aligned} \tag{A.44}$$

We can rewrite Equation A.40 using this result:

$$\begin{aligned}
& \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_l^{|R_{LCU}|-1} (\delta_j c_i \delta_l) P_{jk} P_i P_{kl} \\
&= \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{\substack{l=j \\ [P_{jk}, P_i]=0}} (c_i \delta_j \delta_j^*) P_i + \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{\substack{l=j \\ \{P_{jk}, P_i\}=0}} (-c_i \delta_j \delta_j^*) P_i + \\
& \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{\substack{l>j \\ \forall [P_{jk}, P_i]=0 \\ \{P_{lk}, P_i\}=0}} (\delta_j c_i \delta_l) 2P_i P_j P_l - \sum_j^{|H_q|} \sum_i^{|H_q|} \sum_{\substack{l>j \\ \forall \{P_{jk}, P_i\}=0 \\ [P_{lk}, P_i]=0}} (\delta_j c_i \delta_l) 2P_i P_j P_l \\
&= \sum_i^{|H_q|} \nu_i P_i + \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{\substack{l>j \\ \forall \{P_i, P_j P_l\}=0}}^{|R_{LCU}|-1} \nu_{ijl} P_i P_j P_l,
\end{aligned} \tag{A.45}$$

where we have combined the second and third conditions into a single condition of $\{P_a, P_{jk} P_{kl}\} = \{P_a, P_j P_l\} = 0$ and combined the new coefficients into one coefficient denoted ν .

Next consider the A.39d term of equation A.39. One can use the fact that j and l run over the same indices:

$$\begin{aligned}
& \underbrace{\sum_l^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} c_i \delta_l P_i P_{kl}}_{\text{re-write using } l=j} + \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} \delta_j c_i P_{jk} P_i \\
&= \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} c_i \delta_j P_i P_{kj} + \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} \delta_j c_i P_{jk} P_i \\
&= \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} -\delta_{\mathcal{I}} c_i \delta_j P_i P_{jk} + \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} \delta_j c_i P_{jk} P_i \\
&= \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} c_i \delta_j (P_{jk} P_i - P_i P_{jk}) \\
&= \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \delta_{\mathcal{I}} c_i \delta_j [P_{jk}, P_i] \\
&= \sum_j^{|R_{LCU}|-1} \sum_{\substack{i \\ \forall \{P_{jk}, P_i\}=0}}^{|H_q|} 2\delta_{\mathcal{I}} c_i \delta_j P_{jk} P_i.
\end{aligned} \tag{A.46}$$

Overall we can re-write equation A.39 using these results, yielding:

$$\begin{aligned}
R_{LCU}H_qR_{LCU}^\dagger &= \underbrace{\delta_{\mathcal{I}}^2 \sum_i^{|H_q|} c_i P_i}_{\text{A.39c}} + \underbrace{\sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \mathbb{1}_{\forall\{P_{jk}, P_i\}=0} 2\delta_{\mathcal{I}c_i} \delta_j P_{jk} P_i}_{\text{A.39d using A.46}} + \\
&\quad \underbrace{\sum_i^{|H_q|} \nu_i P_i + \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{\substack{l>j \\ \forall\{P_i, P_j, P_l\}=0}}^{|R_{LCU}|-1} \nu_{ijl} P_i P_j P_l}_{\text{A.39e using A.45}} \\
&= \sum_i^{|H_q|} (\delta_{\mathcal{I}}^2 c_i + \nu_i) P_i + \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \mathbb{1}_{\forall\{P_{jk}, P_i\}=0} 2\delta_{\mathcal{I}c_i} \delta_j P_{jk} P_i + \\
&\quad \sum_j^{|R_{LCU}|-1} \sum_i^{|H_q|} \sum_{\substack{l>j \\ \forall\{P_i, P_j, P_l\}=0}}^{|R_{LCU}|-1} \nu_{ijl} P_i P_j P_l.
\end{aligned} \tag{A.47}$$

We observe that the number of terms in $R_{LCU}H_qR_{LCU}^\dagger$ at worst scales as $|H_q| + |H_q| \cdot (|R_{LCU}|-1) + |H_q| \binom{|R_{LCU}|-1}{2} \binom{|R_{LCU}|-2}{2}$ or $\mathcal{O}(|H_q| \cdot |\mathcal{A}|^2)$. The total number of qubits n bounds the size of $|\mathcal{A}| \leq 2n + 1$ [256], and thus the number of terms in H_q will increase quadratically with the size of \mathcal{A} or number of qubits n when R is defined by a linear combination of unitaries.

A.4 Numerical details of the toy example

This section provides all the details for the Toy problem described in Section 5.3.2. The full noncontextual ground state is:

$$(\underbrace{-1, +1, -1}_{\vec{q}_0}, \underbrace{0.25318483, -0.65828059, -0.70891756}_{\vec{r}_0}). \tag{A.48}$$

This defines the $A(\vec{r}_0)$:

$$A(\vec{r}_0) = 0.25318483 YXYI - 0.65828059 XYXI - 0.70891756 XZ XI. \tag{A.49}$$

The operators to map $A(\vec{r}_0)$ to a single Pauli operator are:

$$R_S = e^{+1i \cdot -0.7879622757719398 \cdot ZYZI} \cdot e^{+1i \cdot 1.2036225088338255 \cdot ZZZI}, \tag{A.50}$$

and

$$R_{LCU} = 0.79157591 IIII + 0.41580383i ZZZI - 0.44778874i ZYZI. \quad (\text{A.51})$$

Their action results in: $R_S A(\vec{r}_0) R_S^\dagger = R_{LCU} A(\vec{r}_0) R_{LCU}^\dagger = YXYI$.

We then defined U depending on which generators we wished to fix. We found the optimal ordering of stabilizers to fix via a brute force search over all $\sum_{i=1}^{|\mathcal{W}_{all}|} \binom{|\mathcal{W}_{all}|}{i} = 2^4 - 1 = 15$ possibilities for \mathcal{W} . The following optimal ordering was obtained:

1. $\{-1 IIIZ\}$
2. $\{+1 IXYI, -1 IIIZ\}$
3. $\{+1 IXYI, -1 IIIZ, +1 \mathcal{A}(\vec{r}_0)\}$
4. $\{-1 YIYI, +1 IXYI, -1 IIIZ, +1 \mathcal{A}(\vec{r}_0)\}$.

This defines all the information required to implement CS-VQE. Table A.1 summarises the stabilizers fixed, the rotation $U_{\mathcal{W}}$, required projection $Q_{\mathcal{W}}$ and final projected Hamiltonian $Q_{\mathcal{W}}^\dagger U_{\mathcal{W}}^\dagger H U_{\mathcal{W}} Q_{\mathcal{W}}$ for this ordering.

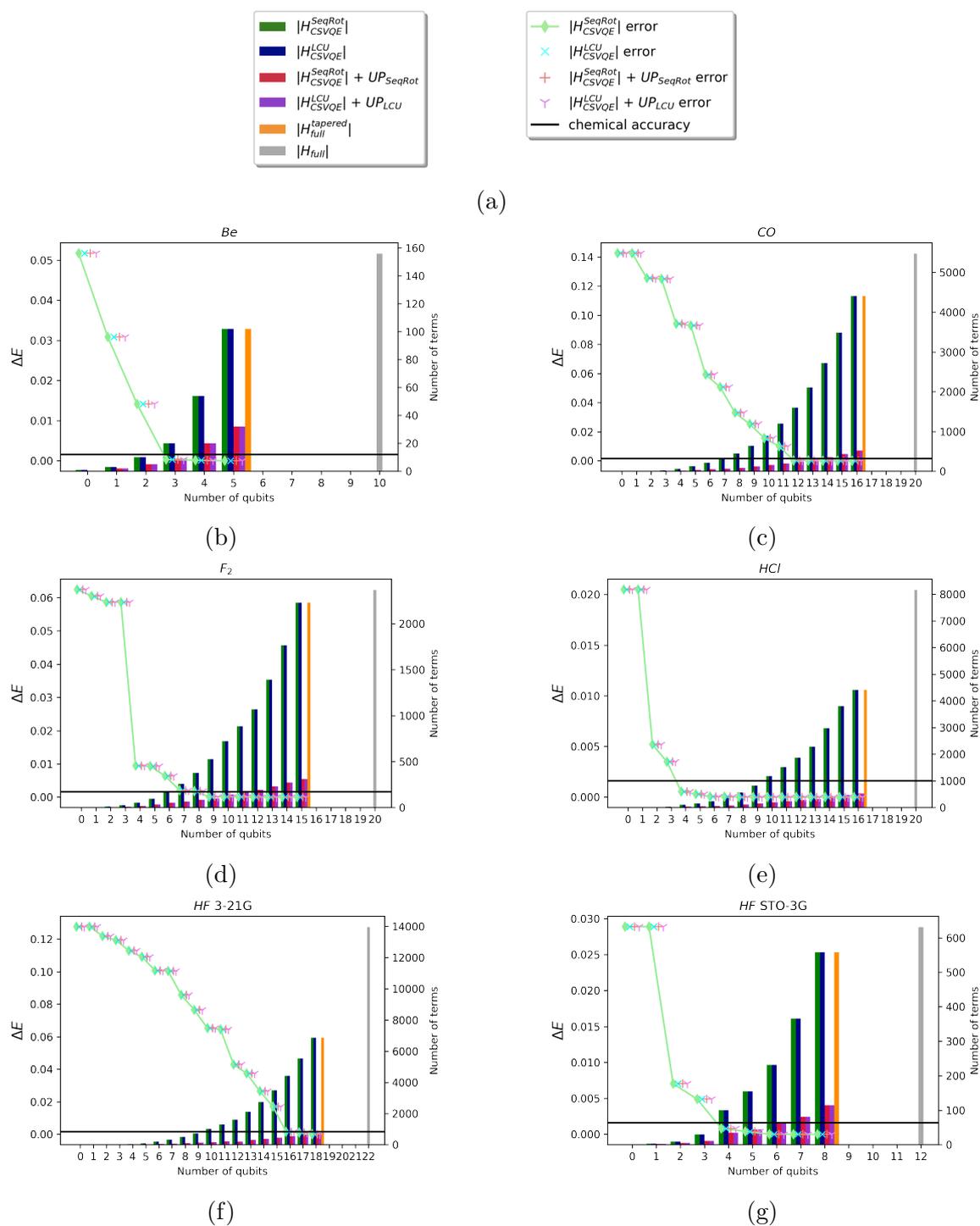
The old approach of applying $U_{\mathcal{W}_{all}}^\dagger H U_{\mathcal{W}_{all}}$ and then fixing certain stabilizer eigenvalues are summarised in Table A.2. It can be seen from these results, that always implementing the unitary partitioning rotation R can unnecessarily increase the number of terms in the Hamiltonian and thus should only be applied if the eigenvalue for $\langle A(\vec{r}) \rangle$ is fixed.

A.4. NUMERICAL DETAILS OF THE TOY EXAMPLE

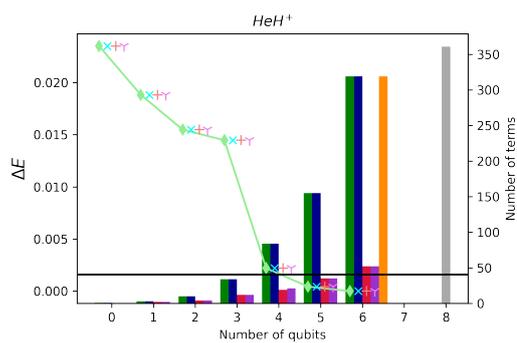
\mathcal{W}	$U_{\mathcal{W}}^\dagger$	$\mathcal{W}^Z = U_{\mathcal{W}}^\dagger \mathcal{W} U_{\mathcal{W}}$	$Q_{\mathcal{W}}$	$Q_{\mathcal{W}}^\dagger U_{\mathcal{W}}^\dagger H U_{\mathcal{W}} Q_{\mathcal{W}}$																																																																																	
-1 YIYI +1 IXYI -1 IIIZ $\mathcal{A}(\vec{r}_0)$	$e^{i\frac{\pi}{4}XYI} e^{i\frac{\pi}{4}IYYI} R_S/LCU$	- ZIII + IZII - IIIZ + IIZI	1⟩ 1⟩ ⊗ 0⟩ ⟨0 ⊗ 1⟩ ⟨1 ⊗ 0⟩ ⟨0	-2.475+0.000j																																																																																	
+1 IXYI -1 IIIZ $\mathcal{A}(\vec{r}_0)$	$e^{i\frac{\pi}{4}IYYI} R_S/LCU$	+ IZII - IIIZ + IIZI	I ⊗ 0⟩ ⟨0 ⊗ 0⟩ ⟨0 ⊗ 1⟩ ⟨1	<table style="width: 100%; border-collapse: collapse; border: none;"> <tr> <td style="text-align: center; border-right: 1px solid black; padding: 2px 5px;">SeqRot</td> <td style="padding: 2px 5px;"></td> <td style="text-align: center; padding: 2px 5px;">LCU</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">-1.827+0.000j I +</td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">-1.827+0.000j I +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">-0.198+0.000j X +</td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">-0.414+0.000j X +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">-0.467+0.000j Z +</td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">-0.292+0.000j Z +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">0.648+0.000j Y</td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.648+0.000j Y</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">-0.500+0.000j II +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.500+0.000j XI +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.700+0.000j XX +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.100+0.000j YI +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">-0.100+0.000j YX +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">1.300+0.000j XZ +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.600+0.000j IY +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.700+0.000j ZZ</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">-0.500+0.000j III +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.100+0.000j XXX +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.200+0.000j YXX +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.700+0.000j XZX +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.700+0.000j YXX +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.100+0.000j YZX +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.200+0.000j XXZ +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.600+0.000j IIY +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.500+0.000j XXY +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.100+0.000j YXY +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.600+0.000j XZZ +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.700+0.000j ZZZ +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.200+0.000j YYZ +</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;"></td> <td style="padding: 2px 5px;">+</td> <td style="padding: 2px 5px;">0.100+0.000j ZYY</td> </tr> </table>	SeqRot		LCU	-1.827+0.000j I +	+	-1.827+0.000j I +	-0.198+0.000j X +	+	-0.414+0.000j X +	-0.467+0.000j Z +	+	-0.292+0.000j Z +	0.648+0.000j Y	+	0.648+0.000j Y		+	-0.500+0.000j II +		+	0.500+0.000j XI +		+	0.700+0.000j XX +		+	0.100+0.000j YI +		+	-0.100+0.000j YX +		+	1.300+0.000j XZ +		+	0.600+0.000j IY +		+	0.700+0.000j ZZ		+	-0.500+0.000j III +		+	0.100+0.000j XXX +		+	0.200+0.000j YXX +		+	0.700+0.000j XZX +		+	0.700+0.000j YXX +		+	0.100+0.000j YZX +		+	0.200+0.000j XXZ +		+	0.600+0.000j IIY +		+	0.500+0.000j XXY +		+	0.100+0.000j YXY +		+	0.600+0.000j XZZ +		+	0.700+0.000j ZZZ +		+	0.200+0.000j YYZ +		+	0.100+0.000j ZYY
SeqRot		LCU																																																																																			
-1.827+0.000j I +	+	-1.827+0.000j I +																																																																																			
-0.198+0.000j X +	+	-0.414+0.000j X +																																																																																			
-0.467+0.000j Z +	+	-0.292+0.000j Z +																																																																																			
0.648+0.000j Y	+	0.648+0.000j Y																																																																																			
	+	-0.500+0.000j II +																																																																																			
	+	0.500+0.000j XI +																																																																																			
	+	0.700+0.000j XX +																																																																																			
	+	0.100+0.000j YI +																																																																																			
	+	-0.100+0.000j YX +																																																																																			
	+	1.300+0.000j XZ +																																																																																			
	+	0.600+0.000j IY +																																																																																			
	+	0.700+0.000j ZZ																																																																																			
	+	-0.500+0.000j III +																																																																																			
	+	0.100+0.000j XXX +																																																																																			
	+	0.200+0.000j YXX +																																																																																			
	+	0.700+0.000j XZX +																																																																																			
	+	0.700+0.000j YXX +																																																																																			
	+	0.100+0.000j YZX +																																																																																			
	+	0.200+0.000j XXZ +																																																																																			
	+	0.600+0.000j IIY +																																																																																			
	+	0.500+0.000j XXY +																																																																																			
	+	0.100+0.000j YXY +																																																																																			
	+	0.600+0.000j XZZ +																																																																																			
	+	0.700+0.000j ZZZ +																																																																																			
	+	0.200+0.000j YYZ +																																																																																			
	+	0.100+0.000j ZYY																																																																																			
+1 IXYI -1 IIIZ	$e^{i\frac{\pi}{4}IYYI}$	+ IZII - IIIZ	I ⊗ 0⟩ ⟨0 ⊗ I ⊗ 1⟩ ⟨1																																																																																		
-1 IIIZ	IIII	-IIIZ	I ⊗ I ⊗ I ⊗ 1⟩ ⟨1																																																																																		

Table A.1: Different contextual subspace Hamiltonians defined from H (Equation 5.27). R_S and R_{LCU} are defined in Equations A.50 and A.51.

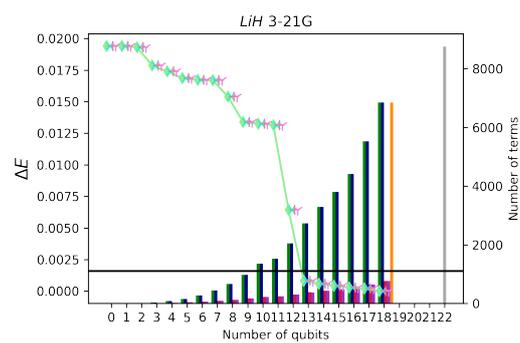
A.5 Graphical results for CS-VQE simulation of each molecular Hamiltonian



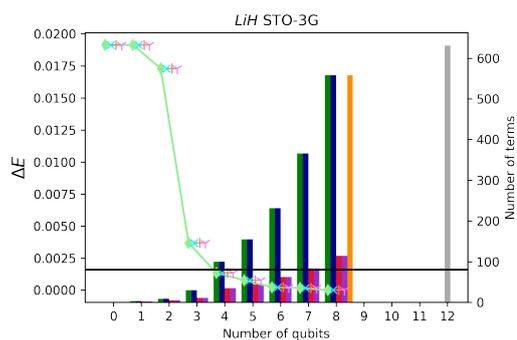
A.5. GRAPHICAL RESULTS FOR CS-VQE SIMULATION OF EACH MOLECULAR HAMILTONIAN



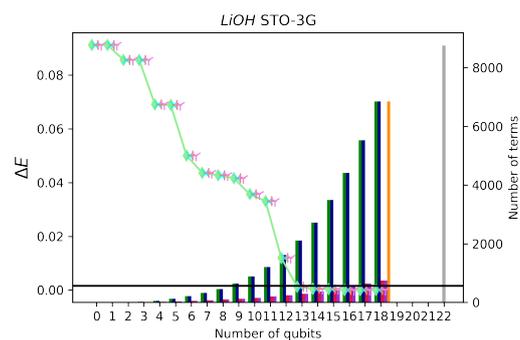
(h)



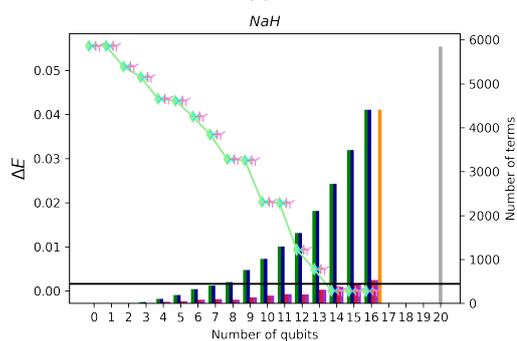
(i)



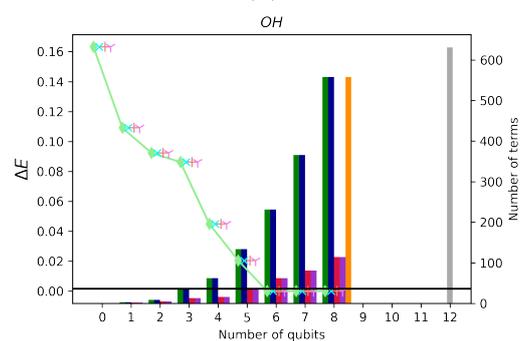
(j)



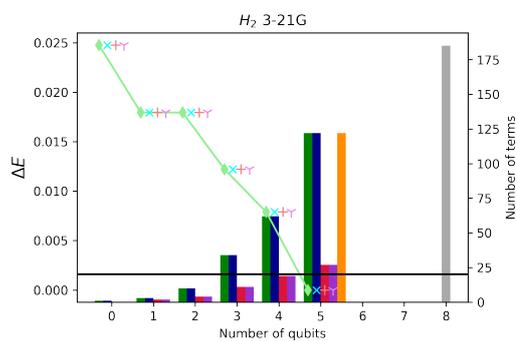
(k)



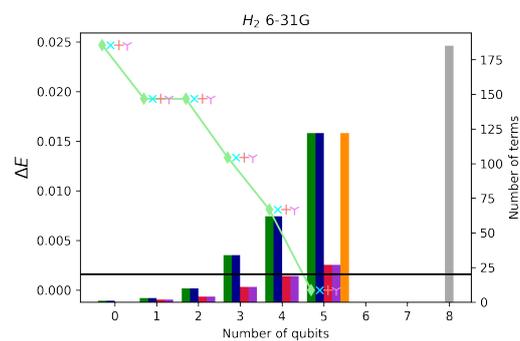
(l)



(m)

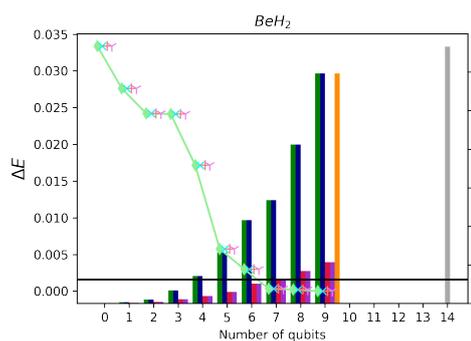


(n)

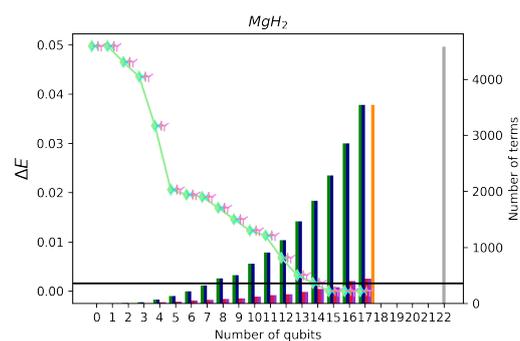


(o)

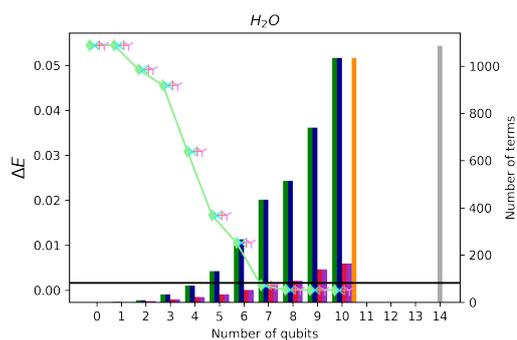
A.5. GRAPHICAL RESULTS FOR CS-VQE SIMULATION OF EACH MOLECULAR HAMILTONIAN



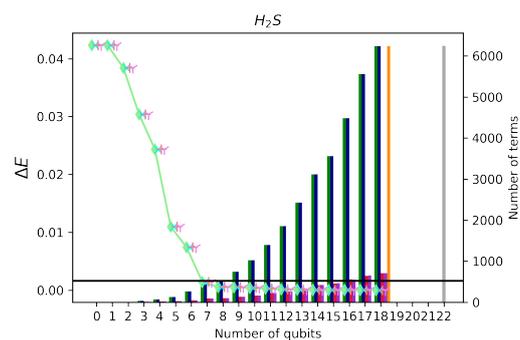
(p)



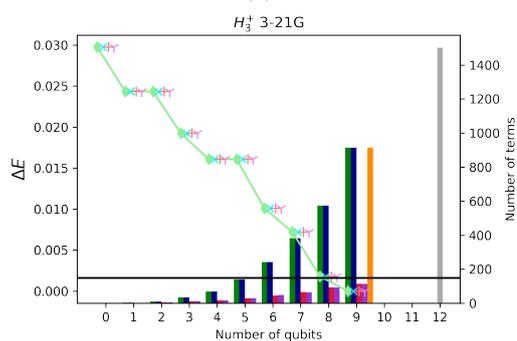
(q)



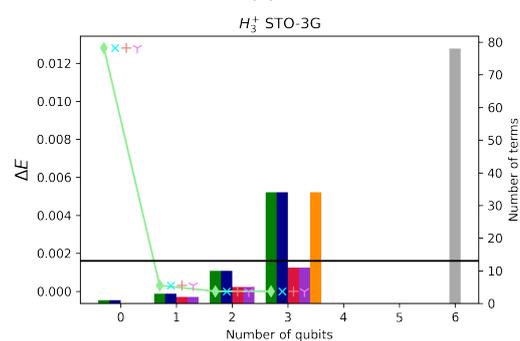
(r)



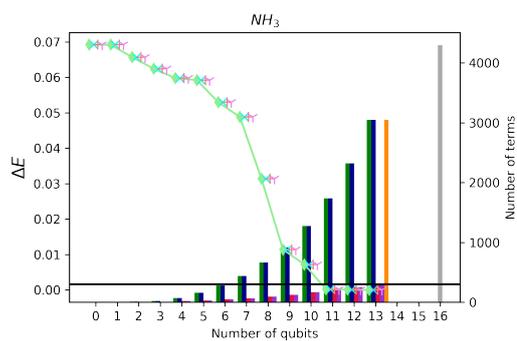
(s)



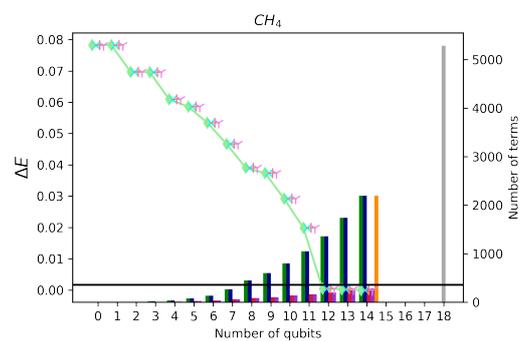
(t)



(u)



(v)



(w)

A.5. GRAPHICAL RESULTS FOR CS-VQE SIMULATION OF EACH MOLECULAR HAMILTONIAN

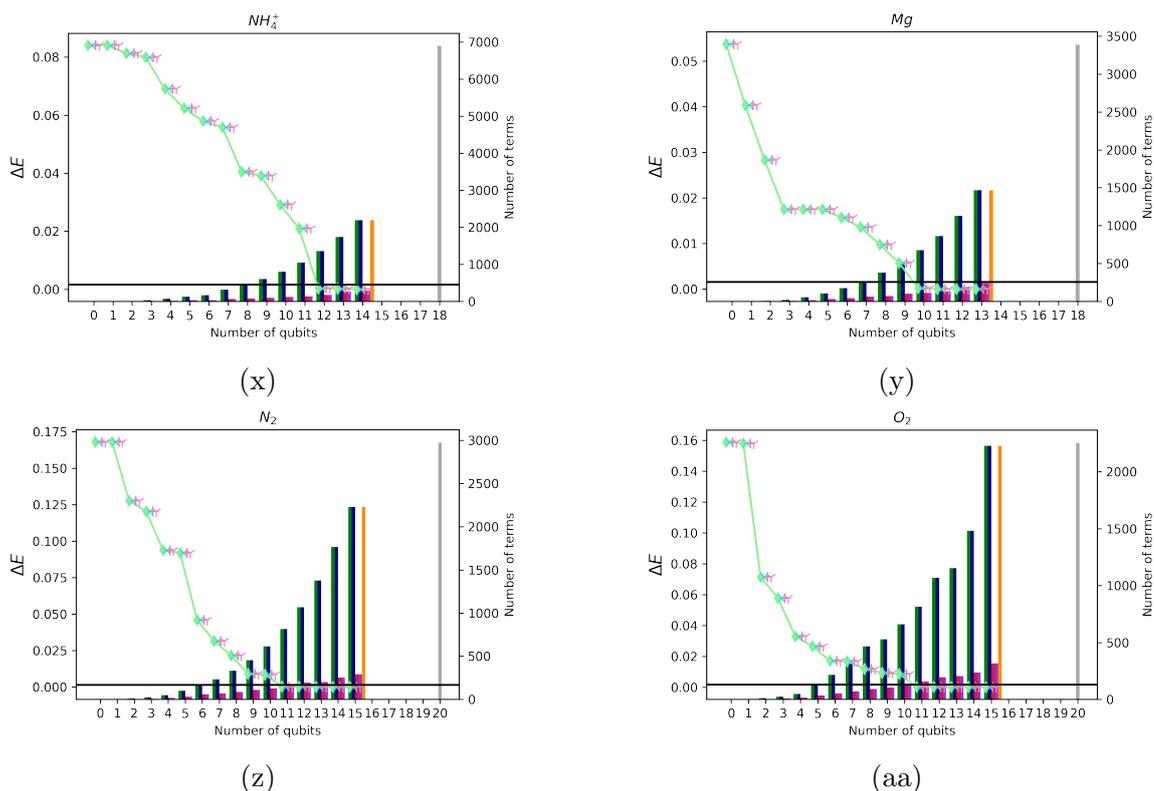


Figure A.1: CS-VQE approximation errors ΔE versus number of qubits used on the quantum computer (scatter plot). The horizontal solid black lines indicate chemical accuracy. The number of terms in each molecular Hamiltonian is given by the bar chart.

All the subplots in Figure A.1 give the simulation results of each molecular Hamiltonian at different levels of noncontextual approximations. This is equivalent to how many contextual stabilizers \mathcal{W} eigenvalues are fixed. In each plot, the leftmost data represents the case when all the noncontextual stabilizer eigenvalues are fixed and is the case for the full noncontextual approximation to a given problem [256]. Moving right, we remove a single stabilizer from \mathcal{W} and thus don't fix the eigenvalue of that stabilizer. This reintroduces a qubit's worth degree of freedom into the problem. At the limit that no stabilizer eigenvalues are fixed ($\mathcal{W} = \{\}$) we return to standard VQE over the full problem and no noncontextual approximation is made. In each plot this scenario is represented by the far right data point (excluding the data for the full non-tapered Hamiltonian that is supplied for reference only). The raw data for these results is supplied in the Supplemental Material of [262]. We include data beyond Hamiltonians achieving chemical accuracy, to show the different possible approximations, rather than stopping once chemical accuracy was achieved.

A.6. TABULATED RESULTS OF SIMULATION

molecule	basis	$H_{\text{CS-VQE}}$	$H_{\text{CS-VQE}} + UP^{(LCU)}$	$H_{\text{CS-VQE}} + UP^{(SeqRot)}$	H_{tapered}	$RH_{\text{tapered}}R^\dagger$	H_{full}
BeH ₂	STO-3G	(7, 268)	(7, 61)	(7, 61)	(9, 596)	(9, 614)	(14, 666)
Mg	STO-3G	(10, 675)	(10, 114)	(10, 114)	(13, 1465)	(13, 1465)	(18, 3388)
H ₃ ⁺	3-21G	(9, 914)	(9, 115)	(9, 115)	(9, 914)	(9, 786)	(12, 1501)
O ₂	STO-3G	(11, 815)	(11, 157)	(11, 157)	(15, 2229)	(15, 2374)	(20, 2255)
OH	STO-3G	(6, 231)	(6, 62)	(6, 62)	(8, 558)	(8, 558)	(12, 631)
CH ₄	STO-3G	(12, 1359)	(12, 203)	(12, 203)	(14, 2194)	(14, 2194)	(18, 5288)
Be	STO-3G	(3, 20)	(3, 9)	(3, 9)	(5, 102)	(5, 108)	(10, 156)
NH ₃	STO-3G	(11, 1733)	(11, 200)	(11, 200)	(13, 3048)	(13, 2738)	(16, 4293)
H ₂ S	STO-3G	(7, 435)	(7, 92)	(7, 92)	(18, 6237)	(18, 6237)	(22, 6246)
H ₂	3-21G	(5, 122)	(5, 27)	(5, 27)	(5, 122)	(5, 124)	(8, 185)
HF	3-21G	(17, 5530)	(17, 648)	(17, 648)	(18, 6852)	(18, 6852)	(22, 13958)
F ₂	STO-3G	(9, 527)	(9, 99)	(9, 99)	(15, 2229)	(15, 2229)	(20, 2367)
HCl	STO-3G	(4, 100)	(4, 35)	(4, 35)	(16, 4409)	(16, 4409)	(20, 8159)
HeH ⁺	3-21G	(5, 155)	(5, 35)	(5, 35)	(6, 319)	(6, 319)	(8, 361)
MgH ₂	STO-3G	(15, 2285)	(15, 289)	(15, 289)	(17, 3540)	(17, 3540)	(22, 4582)
CO	STO-3G	(12, 1599)	(12, 241)	(12, 241)	(16, 4409)	(16, 4409)	(20, 5475)
LiH	STO-3G	(4, 100)	(4, 35)	(4, 35)	(8, 558)	(8, 586)	(12, 631)
N ₂	STO-3G	(11, 815)	(11, 153)	(11, 153)	(15, 2229)	(15, 2229)	(20, 2975)
NaH	STO-3G	(14, 2722)	(14, 375)	(14, 375)	(16, 4409)	(16, 4409)	(20, 5851)
H ₂ O	STO-3G	(7, 435)	(7, 73)	(7, 73)	(10, 1035)	(10, 1035)	(14, 1086)
H ₃ ⁺	STO-3G	(1, 3)	(1, 2)	(1, 2)	(3, 34)	(3, 35)	(6, 78)
LiOH	STO-3G	(13, 2104)	(13, 296)	(13, 296)	(18, 6852)	(18, 6852)	(22, 8758)
LiH	3-21G	(13, 2732)	(13, 375)	(13, 383)	(18, 6852)	(18, 6852)	(22, 8758)
H ₂	6-31G	(5, 122)	(5, 27)	(5, 27)	(5, 122)	(5, 124)	(8, 185)
NH ₄ ⁺	STO-3G	(12, 1359)	(12, 176)	(12, 176)	(14, 2194)	(14, 2194)	(18, 6892)
HF	STO-3G	(4, 100)	(4, 35)	(4, 35)	(8, 558)	(8, 558)	(12, 631)

Table A.3: Different resource requirements to study different electronic structure Hamiltonians required to achieve chemical accuracy. Each round bracket tuple reports $(n, |H|)$ and gives the number of qubits and terms for each Hamiltonian considered. $RH_{\text{tapered}}R^\dagger$ describes the effect of the CS-VQE unitary partitioning rotation on the problem Hamiltonian and $H_{\text{CS-VQE}} = Q_{\mathcal{W}}U_{\mathcal{W}}^\dagger H_{\text{full}}U_{\mathcal{W}}Q_{\mathcal{W}}^\dagger$. The size of the Hamiltonian for LiH (3-21G singlet) with measurement reduction applied is different for the sequence of rotations and LCU unitary partitioning methods. This is an artefact of the graph colouring heuristic when finding different anticommuting cliques in the CS-VQE Hamiltonian.

A.6 Tabulated results of simulation

Table A.3 summarises the numerical results of Figures 5.3 and 5.4.

Appendix B

PBE Appendix

B.1 Molecular ground state energy

In addition to the result displayed in Figure 6.2, which were calculated using the SPADE projection method, we present results for the same molecules using the Intrinsic Bonding Orbitals localisation method in Figure B.1. Numerical values for these results are given in Table B.1 for reference values, and Tables B.2 and B.3 for our calculated results.

Molecule	ϵ_{DFT}	Q	$ H $
N-methylmethanamine	0.5733	44	338971
acetaldehyde	0.569	38	182702
acetonitrile	0.485	36	136067
ethanamine	0.573	44	329283
ethanol	0.609	42	283020
fluoroethane	0.637	40	217385
formamide	0.619	36	138235

Table B.1: Full-system reference values for embedding calculations of small molecules, as shown in Figures 6.2 and B.1. ϵ_{DFT} gives the difference between full-system RKS DFT, using the B3LYP functional, and CCSD(T). Q and $|H|$ give respectively the number of qubits and terms in the Jordan-Wigner encoded qubit Hamiltonian of the full system. All energies are reported in Hartree (Ha).

B.2 Strong correlation

We provide the numerical details of our strongly correlated H₂O study in this section, where SPADE localization has been used. These results form Figure 6.3 in the main text. We also include results for IBO localization, with the localization threshold lowered.

For the H₂O projection-based embedding calculations, at different molecular geometries, we considered two different active regions. One had the atoms in the fixed OH bond set active and the other had the atoms in the changing OH bond set active. The structure for H₂O with an OH bond length of 0.4 Å is given in Table B.4. The

B.2. STRONG CORRELATION

Molecule	ϵ_{huz}	ϵ_{μ}	Q_{huz}	Q_{μ}	$ H_{emb}^{huz} $	$ H_{emb}^{\mu} $
N-methylmethanamine	0.491	0.494	24	24	29701	29737
acetaldehyde	0.099	0.178	30	30	70118	69554
acetonitrile	0.462	0.462	18	18	9232	9016
ethanamine	0.394	0.399	26	26	38716	39924
ethanol	0.340	0.350	26	26	40948	41060
fluoroethane	0.299	0.312	26	26	35020	38180
formamide	0.043	0.109	30	30	62434	61914

Table B.2: Numerical values of the calculations shown in Figure B.1. For each calculation the energy difference between CCSD(T)-in-DFT embedding and full system CCSD(T), ϵ , is given. The number of qubits Q and the number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. Results for IBO localisation are shown, with results for SPADE in Table B.3 reference values using the full system given in Table B.1. All energies are reported in Hartree (Ha).

Molecule	ϵ_{huz}	ϵ_{μ}	Q_{huz}	Q_{μ}	$ H_{emb}^{huz} $	$ H_{emb}^{\mu} $
N-methylmethanamine	0.135	0.169	36	36	152223	152415
acetaldehyde	0.098	0.176	30	30	70474	69982
acetonitrile	0.403	0.409	20	20	13439	13111
ethanamine	0.136	0.169	36	36	145067	149819
ethanol	0.132	0.169	34	34	120928	121036
fluoroethane	0.136	0.170	32	32	85781	89377
formamide	0.045	0.110	30	30	62366	61958

Table B.3: Numerical values of the calculations shown in Figure 6.2. For each calculation the energy difference between CCSD(T)-in-DFT embedding and full system CCSD(T), ϵ , is given. The number of qubits Q and the number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. Results for SPADE localisation are shown, with results for IBO in Table B.2 reference values using the full system given in Table B.1. All energies have been reported in Hartree (Ha).

B.2. STRONG CORRELATION

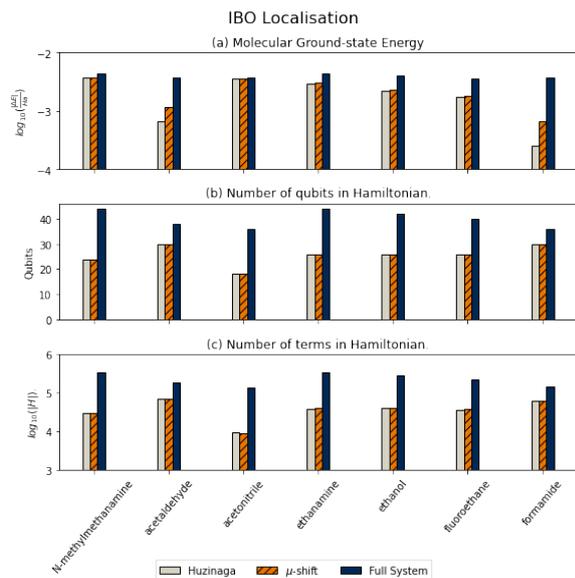


Figure B.1: Results for embedding of small molecules using the IBO localisation method. (a) Ground state energies error for small molecules, with full-system DFT energy error as a reference, μ -shift CCSD-in-DFT embedding energy in orange and Huzinaga CCSD-in-DFT embedding in grey. All error values are calculated with respect to whole system CCSD(T) energies. (b) The number of qubits required to describe the embedded FCI-in-DFT Hamiltonians, with the reference showing the number required for the full system second quantized Hamiltonian. (c) The number of terms in the Jordan-Wigner encoded FCI-in-DFT qubit Hamiltonian for each molecule. Again the reference gives the number needed for the full system second quantized Hamiltonian.

other geometries can be generated from this structure. Tables B.5 and B.6 summarise the numerical results for the different active systems where SPADE localization has been used. Tables B.7 and B.8 give the numerical results for the different active systems when IBO localization was used. Finally, Figure B.2 provides a summary of the IBO results when the minimum localization threshold was reduced to 90%.

B.2. STRONG CORRELATION

atom	x	y	z
H	0.3751747	0.0000000	0.1387225
O	0.0000000	0.0000000	0.0000000
H	-0.7493682	0.0000000	0.2770822

Table B.4: Cartesian coordinates of atoms in H₂O for the structure with an OH bond length of 0.4 Å defined from the first H and O atoms in this Table. The other structures (different OH bond lengths) were generated from this file by changing the position of the first H atom. Note the H-O-H angle remained fixed.

OH length	E_{FCI}^{global}	E_{DFT}^{global}	$E_{FCI-in-DFT}^{hu\bar{z}}$	$E_{FCI-in-DFT}^{\mu}$	$ H_{full} $	$ H_{FCI-in-DFT}^{hu\bar{z}} $	$ H_{FCI-in-DFT}^{\mu} $	Q_{FCI}^{global}	$Q_{FCI-in-DFT}^{hu\bar{z}}$	$Q_{FCI-in-DFT}^{\mu}$	# active MOs
0.400000	-72.981056	-73.259936	-72.988008	-72.988009	2110	1079	1079	14	12	12	4
0.600000	-74.499220	-74.773419	-74.508545	-74.508545	2110	1079	1079	14	12	12	4
0.798954	-74.851089	-75.122053	-74.864002	-74.864002	1086	1079	1079	14	12	12	4
1.000000	-74.900658	-75.170068	-74.918225	-74.918226	2110	1079	1079	14	12	12	4
1.200000	-74.867498	-75.134418	-74.890912	-74.890913	2110	1079	1079	14	12	12	4
1.500000	-74.807539	-75.057384	-74.840739	-74.840739	2110	1079	1079	14	12	12	4
2.000000	-74.776263	-74.962535	-74.816902	-74.816902	2110	1079	1079	14	12	12	4
3.000000	-74.771687	-74.890577	-74.820223	-74.820224	2530	1327	1327	14	12	12	4
4.000000	-74.771719	-74.876129	-74.820261	-74.820262	3346	1819	1819	14	12	12	4
5.000000	-74.771718	-74.872116	-74.820410	-74.820411	3054	1799	1799	14	12	12	4

Table B.5: Numerical values of the calculations shown in Figure 6.3a for the case when the changing OH bond is set as the active region and SPADE localization has been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits Q and the number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies are reported in Hartree (Ha).

OH length	E_{FCI}^{global}	E_{DFT}^{global}	$E_{FCI-in-DFT}^{hu\bar{z}}$	$E_{FCI-in-DFT}^{\mu}$	$ H_{full} $	$ H_{FCI-in-DFT}^{hu\bar{z}} $	$ H_{FCI-in-DFT}^{\mu} $	Q_{FCI}^{global}	$Q_{FCI-in-DFT}^{hu\bar{z}}$	$Q_{FCI-in-DFT}^{\mu}$	# active MOs
0.400000	-72.981056	-73.259936	-72.887822	-72.887827	2110	1079	1079	14	12	12	4
0.600000	-74.499220	-74.773419	-74.473689	-74.473691	2110	1079	1079	14	12	12	4
0.798954	-74.851089	-75.122053	-74.864002	-74.864002	1086	1079	1079	14	12	12	4
1.000000	-74.900658	-75.170068	-74.936101	-74.936101	2110	1079	1079	14	12	12	4
1.200000	-74.867498	-75.134418	-74.914336	-74.914336	2110	1079	1079	14	12	12	4
1.500000	-74.807539	-75.057384	-74.847328	-74.847328	2110	1079	1079	14	12	12	4
2.000000	-74.776263	-74.962535	-74.755506	-74.755506	2110	1079	1079	14	12	12	4
3.000000	-74.771692	-74.890579	-74.683921	-74.683921	2406	1203	1327	14	12	12	4
4.000000	-74.771719	-74.876127	-74.669427	-74.669427	2598	1083	1119	14	12	12	4
5.000000	-74.771718	-74.872381	-74.665626	-74.665626	3046	1759	1799	14	12	12	4

Table B.6: Numerical values of the calculations shown in Figure 6.3a for the case when the fixed OH bond is set as the active region and SPADE localization has been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits Q and the number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies are reported in Hartree (Ha).

B.2. STRONG CORRELATION

OH length	E_{FCI}^{global}	E_{DFT}^{global}	$E_{\text{FCI-in-DFT}}^{\text{huz}}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$ H_{\text{full}} $	$ H_{\text{FCI-in-DFT}}^{\text{huz}} $	$ H_{\text{FCI-in-DFT}}^{\mu} $	Q_{FCI}^{global}	$Q_{\text{FCI-in-DFT}}^{\text{huz}}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	# active MOs
0.400000	-72.981056	-73.259936	-72.928542	-72.923051	2110	492	492	14	10	10	3
0.600000	-74.499220	-74.773419	-74.484545	-74.484551	2110	492	492	14	10	10	3
0.798954	-74.851089	-75.122053	-74.852614	-74.852616	1086	1079	1079	14	12	12	4
1.000000	-74.900658	-75.170068	-74.899540	-74.899543	2110	1079	1079	14	12	12	4
1.200000	-74.867498	-75.134418	-74.868096	-74.868099	2110	1079	1079	14	12	12	4
1.500000	-74.807539	-75.057384	-74.819776	-74.819778	2110	1079	1079	14	12	12	4
2.000000	-74.776263	-74.962535	-74.798952	-74.798954	2110	1079	1079	14	12	12	4
3.000000	-74.771692	-74.890578	-74.804924	-74.804927	2238	1379	1383	14	12	12	4
4.000000	-74.771719	-74.873986	-74.470749	-74.470821	2342	876	876	14	10	10	3
5.000000	-74.771718	-74.872112	-74.805202	-74.805204	3210	1779	1775	14	12	12	4

Table B.7: Numerical values of the calculations shown in Figure 6.3b for the case when the changing OH bond is set as the active region and IBO localization (95% threshold) is been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits Q and the number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies are reported in Hartree (Ha).

OH length	E_{FCI}^{global}	E_{DFT}^{global}	$E_{\text{FCI-in-DFT}}^{\text{huz}}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$ H_{\text{full}} $	$ H_{\text{FCI-in-DFT}}^{\text{huz}} $	$ H_{\text{FCI-in-DFT}}^{\mu} $	Q_{FCI}^{global}	$Q_{\text{FCI-in-DFT}}^{\text{huz}}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	# active MOs
0.400000	-72.981056	-73.259936	-72.928542	-72.923051	2110	492	492	14	10	10	3
0.600000	-74.499220	-74.773419	-74.484545	-74.484551	2110	492	492	14	10	10	3
0.798954	-74.851089	-75.122053	-74.852614	-74.852616	1086	1079	1079	14	12	12	4
1.000000	-74.900658	-75.170068	-74.899540	-74.899543	2110	1079	1079	14	12	12	4
1.200000	-74.867498	-75.134418	-74.868096	-74.868099	2110	1079	1079	14	12	12	4
1.500000	-74.807539	-75.057384	-74.819776	-74.819778	2110	1079	1079	14	12	12	4
2.000000	-74.776263	-74.962535	-74.798952	-74.798954	2110	1079	1079	14	12	12	4
3.000000	-74.771692	-74.890578	-74.804924	-74.804927	2238	1379	1383	14	12	12	4
4.000000	-74.771719	-74.873986	-74.470749	-74.470821	2342	876	876	14	10	10	3
5.000000	-74.771718	-74.872112	-74.805202	-74.805204	3210	1779	1775	14	12	12	4

Table B.8: Numerical values of the calculations shown in Figure 6.3b when the fixed OH bond is set as the active region and IBO localization (95% threshold) is been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits Q and the number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies are reported in Hartree (Ha).

B.2. STRONG CORRELATION

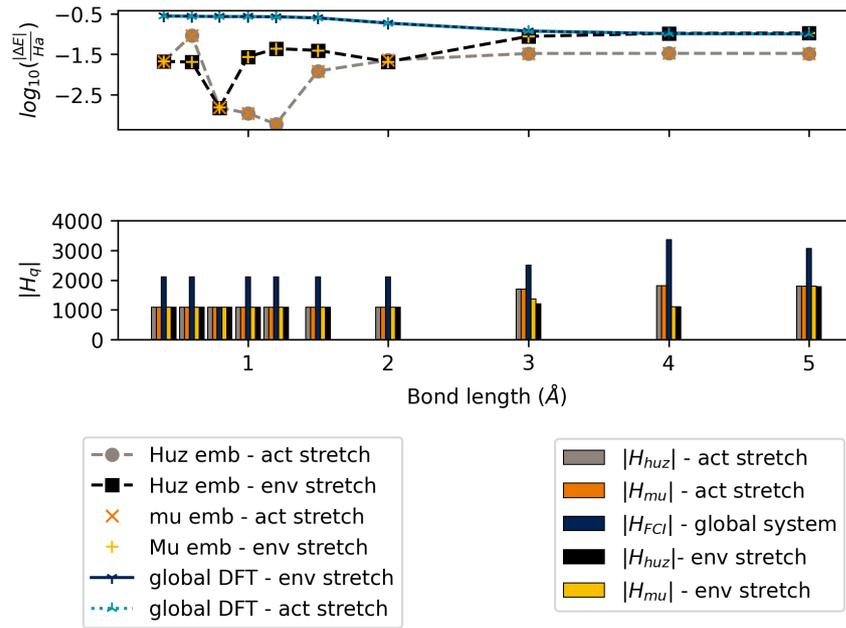


Figure B.2: Potential energy curve for H_2O , with changing OH bond length. *Active stretch* result has the changing OH bond as the active region and *environment stretch* result has the fixed OH bond selected as the active region. results use IBO localization (see raw data [336] for threshold values - lowest setting was 90%). For each data set the full problem is reduced from 14 to 12 qubits, with the number of active MOs being four in all cases. The top plot reports the \log base 10 error with respect to the exact FCI ground state energy (E_{FCI}) of the whole system, where $|\Delta E| = |E_{exp} - E_{FCI}|$. Here E_{exp} is obtained from an FCI-in-DFT calculation. The bottom plot reports the number of terms in each Jordan-Wigner encoded qubit Hamiltonian. The blue result gives the size of the full system Hamiltonian, the orange and yellow results are for μ -shifted embedded Hamiltonians while the grey and black results are for the Huzinaga embedded Hamiltonians.