Contextual Subspace Chemistry:
Practical Approaches to Quantum Computing for Electronic Structure

Timothy John Weaving

A thesis submitted for the degree of
Doctor of Philosophy

Primary Supervisor:	Prof Peter V. Coveney
Secondary Supervisors:	Prof Peter J. Love, Prof Alexandra Olaya-Castro

Department of Chemistry
University College London
United Kingdom

July 1, 2025

Declaration

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

Abstract

Quantum computers are expected to be a transformative technology for many domains of science. They promise to bring the most pressing questions of scientific discovery into focus, where the inherent limitations of conventional computing approaches fall short. One domain that should be an early beneficiary of quantum computation is quantum chemistry, which is concerned with resolving the electronic structure of molecular systems. The wavefunction describing a quantum many-body system lives within an exponentially vast vector space and direct treatment is typically intractable for classical platforms. By contrast, the state of a *quantum* computer describes a many-body system and is therefore a better representation of the problems we aim to solve. While the theoretical value of quantum computation is clear, there is still a staggering amount of progress to make before we see this novel computing paradigm providing solutions to problems of a sufficient scale to be considered scientifically relevant. Today, quantum hardware is very noisy, and the period of time over which we are able to sustain its fragile quantum state – the *coherence* time – is remarkably short. To extract useful information from current and near-term devices, the field of quantum error mitigation has seen rapid expansion with the goal of distilling utility from these noisy devices. Furthermore, to alleviate the burden on quantum hardware, hybrid algorithms leveraging the strengths of both quantum *and* classical resources have been developed which, alongside subspace techniques, extends the reach of current quantum computers. This thesis presents novel work at the intersection of these fields, with the development of improved hybrid algorithms, subspace techniques and error mitigation strategies, complemented by practical implementations that validate the methodology developed herein. We suggest that this approach reveals a path towards quantum advantage as hardware matures in the coming years, providing solutions to vital questions in chemistry.

Impact Statement

In this thesis we investigate algorithms for the near-term application of quantum computers. This involves the development of a flexible quantum resource management framework, allowing one to reduce the number of qubits required to represent a problem of interest and thus alleviate some burden on the quantum compute. While there is a tradeoff between the subspace size and an incurrence of error, the methodology presented herein facilitates the treatment of chemical systems at a scale that would otherwise lie outside the reach of current quantum capabilities. Moreover, our theoretical advances are complemented by practical validation on IBM Quantum superconducting hardware, necessitating the meticulous implementation of quantum error mitigation techniques. We perform a comprehensive benchmark of many contemporary techniques on the problem of ground state preparation for hydrogen chloride, HCl, to identify effective error mitigation strategies. The optimal strategy is then adopted in calculating the dissociation curve of molecular nitrogen, N_2, consisting of ten separate simulations on quantum hardware. The results are compared against a suite of off-the-shelf classical techniques and proves to be one of few instances of a quantum simulation remaining competitive or even outperforming classical methods of a similar computational cost. Our robust simulation methodology is also applied to the ground state preparation of an antiferromagnetic Heisenberg model on a single cell of the Kagome lattice. This problem was posed for the IBM Quantum Open Science Prize 2023 and entries were assessed on solution novelty and final energy error obtained from the provided quantum hardware. The qubit subspace approach developed in this thesis achieved an error ratio of 0.078% and achieved first place globally [1]. Finally, we simulate the time evolution of an Ising model over two/-four cells of the heavy-hex lattice, consisting of calculations up to 35-qubits in size. This tests the efficacy of a novel error mitigation strategy combining two contemporary techniques: Echo Verification and Clifford Data Regression. In all, the work presented in this thesis assesses the near-term outlook and practicality of quantum computation and provides a flexible qubit subspace framework that shall scale as quantum computers mature, so that we may address systems of increasing complexity with the endpoint of industrial and/or scientific relevance. With the realization of large-scale quantum computation, the societal and economic impact will be profound and shall have implications in many application domains such as drug discovery and the design of improved catalysis, batteries, solar cells and nanotechnology, driving advancements in next-generation technologies.

Research Paper Declaration Form

1. Weaving, T., Ralli, A., Kirby, W.M., Tranter, A., Love, P.J. and Coveney, P.V., 2023. A stabilizer framework for the contextual subspace variational quantum eigensolver and the noncontextual projection ansatz. Journal of Chemical Theory and Computation, 19(3), pp.808-821. [2]
    - Constitutes Sections 2.3 and 3.3.
2. Weaving, T., Ralli, A., Kirby, W.M., Love, P.J., Succi, S. and Coveney, P.V., 2023. Benchmarking noisy intermediate scale quantum error mitigation strategies for ground state preparation of the HCl molecule. Physical Review Research, 5(4), p.043054. [3], code repository [4].
    - Contributed to the Quantum Error Mitigation chapter, in particular to Sections 4.1, 4.2, 4.3, 4.6 and 4.7. The HCl application demonstration is given in Section 5.1.
3. Weaving, T., Ralli, A., Love, P.J., Succi, S. and Coveney, P.V., 2025. Contextual subspace variational quantum eigensolver calculation of the dissociation curve of molecular nitrogen on a superconducting quantum computer. npj Quantum Information, 11(1), p.25. [5], code repository [6].
    - Contributed the Hardware-Aware ADAPT-VQE algorithm presented in Section 1.4.5 and to the Quantum Error Mitigation chapter, in particular to Sections 4.4, 4.5 and 4.6. The N_2 application demonstration is given in Section 5.2.
4. Weaving, T., Ralli, A., Love, P.J., Succi, S. and Coveney, P.V., 2025. Accurately Simulating the Time Evolution of an Ising Model with Echo Verified Clifford Data Regression on a Superconducting Quantum Computer. Quantum, 9, p.1732. [7], code repository [8].
    - Constitutes the majority of Sections 4.7 and 4.8. The Ising model time evolution application demonstration is given in Section 5.4.
5. Weaving, T., Ralli, A., Wimalaweera, V., Love, P.J. and Coveney, P.V., 2025. Simulating the Antiferromagnetic Heisenberg Model on a Spin-Frustrated Kagome Lattice with the Contextual Subspace Variational Quantum Eigensolver. arXiv preprint arXiv:2506.12391. [9]
    - The above preprint elaborates and improves upon Section 5.3, whereas the thesis version preserves the information as it was submitted for the IBM Quantum Open Science prize 2023 [1].

Novel Contributions

This thesis presents a number of novel contributions to the field of quantum algorithms and error mitigation, particularly related to qubit subspace methods, the variational quantum eigensolver (VQE) and simulations of quantum systems on quantum hardware. In addition to showcasing several hardware demonstrations of successful quantum simulations in the domains of molecular electronic structure and condensed matter, the main contributions are as follows:

- **Hardware-Aware Adaptive Circuit Construction** A hardware-aware modification to the Adaptive Derivative-Assembled Pseudo-Trotter VQE (ADAPT-VQE) algorithm is introduced. While ADAPT-VQE dynamically constructs compact and problem-specific ansatz circuits, the modification developed herein incorporates hardware-awareness to minimize the cost of transpilation for quantum hardware. This is achieved by incorporating qubit-connectivity constraints from the target quantum device into the circuit construction, thereby reducing circuit depths post-transpilation. Such an approach is essential for maximizing the performance and reliability of simulations on noisy intermediate-scale quantum (NISQ) hardware. A detailed account of this development is provided in Section 1.4.5.

- **The Stabilizer Framework for Qubit Subspace Methods** This thesis develops a stabilizer-based framework for the efficient implementation of qubit subspace techniques, including Qubit Tapering and the Contextual Subspace method. By leveraging the symplectic representation of Pauli operators, many core operations reduce to standard binary logic, enabling scalable and efficient computation. Situating the Contextual Subspace method within this stabilizer formalism facilitates its application to large-scale Hamiltonian systems, thereby overcoming a key limitation of earlier approaches.

  The relevant aspects of the symplectic formalism are introduced in Section 2.1 and the stabilizer subspace projection is detailed in Section 2.3. In Section 3.1 a new algorithm for detecting contextuality is proposed, which substantially outperforms the existing method and achieves a speedup of several orders of magnitude. Section 3.3 demonstrates the integration of the Contextual Subspace method into the stabilizer framework and includes the first simulations using the Contextual Subspace Variational Quantum Eigensolver (CS-VQE). These noiseless simulations, presented in Section 3.3.2, leverage the ADAPT-VQE algorithm to solve subspaces, in contrast to prior work which relied on direct diagonalization.

  Moreover, this thesis presents the first demonstrations of CS-VQE deployed on real quantum hardware for a range of molecular and many-body systems.

These include the ground state preparation of hydrogen chloride (HCl, Section 5.1), the computation of the potential energy curve for molecular nitrogen (N_2, Section 5.2) and simulation of the field-free Heisenberg model on a Kagome lattice (Section 5.3).

The `symmer` Python package [10] was developed to support the work presented in this thesis and provides an easy-to-use symplectic implementation of the Contextual Subspace method. The project secured Unitary Fund support and has received recognition from the broader quantum computing community.

- **Quantum Error Mitigation Benchmarking** This thesis presents a thorough benchmarking of many error mitigation strategies in Section 5.1, composed of different combinations of Measurement-Error Mitigation, Symmetry Verification, Zero-Noise Extrapolation and Echo Verification. With HCl ground state preparation taken as the testbed system, the goal was to identify optimal strategies to get the best performance from the available quantum hardware.

- **Circuit Tiling and Batching** To address variability in qubit noise and improve statistical efficiency on NISQ hardware, this thesis introduces a circuit tiling/batching technique. This approach improves shot efficiency and passively averages over non-uniformity in qubit performance by distributing logically equivalent circuits across different regions of a quantum chip. Introduced in Section 4.5, the technique has been successfully applied to several simulations on hardware, including those of the N_2 molecule (Section 5.2), the Kagome lattice Heisenberg model (Section 5.3), and time evolution under the Ising model (Section 5.4). An implementation of this methodology is provided in the `evtools` Python package [8], which was developed to support the Echo Verification protocols also presented in this thesis.

- **Novel Echo Verification Error Mitigation Techniques** This thesis provides the first work conducting comprehensive noise modelling of the Echo Verification method under both depolarizing and more general Pauli noise, as discussed in Sections 4.7.1 and 4.7.2, respectively. A novel approach is introduced for estimating depolarization rate based on the postselection probability and purity of an ancilla qubit, enabling more precise characterization of quantum noise and could also find use as a quantum benchmarking tool (Section 4.7.1).

  In Section 4.7.4 we also explore various strategies for multi-ancilla Echo Verification, allowing the simultaneous estimation of several observables. Motivated by the theoretical insights gained from Pauli noise modelling of Section 4.7.2, a new hybrid error mitigation strategy is proposed, combining Echo Verification with Clifford Data Regression. This combined technique, presented in Section 4.8, offers improved mitigation of coherent and stochastic errors. The practical utility of this technique is demonstrated through a benchmarking study in Section 5.4, where we simulate time evolution of the Ising model on heavy-hex lattices up to 35-qubits in size.

Acknowledgements

First of all, I would like to express my profound gratitude to the primary and secondary supervisors who supported this work, Prof Peter V. Coveney, Prof Peter J. Love and Prof Alexandra Olaya-Castro. Your unique perspectives, insights and guidance helped shape the direction of this research and revealed new avenues for me to pursue. I also wish to extend this thanks to Prof Sauro Succi and the Italian Institute of Technology for their support.

I am also grateful to the UCL Centre for Doctoral Training in Delivering Quantum Technologies and the brilliant team: Prof Paul Warburton, Dr Abbie Bray, Lopa Murgai and Prof Dan Browne as my cohort mentor.

Throughout my PhD research I had the privilege of collaborating extensively with Dr Alexis Ralli. Together, we developed the `symmer` codebase that provides implementations of many of the techniques presented within this thesis; we are both thankful for Unitary Fund support on the project.

My appreciation extends also to Dr William M. Kirby and Dr Andrew Tranter, from whom I learnt a great deal about strong measurement contextuality, qubit subspace techniques and related topics through many discussions in the earlier years of my research.

I am grateful to fellow PhD students Michael Williams de la Bastida, Angus Mingare and Tom Bickley for countless thought-provoking discussions, as well as the wider group in the UCL Centre for Computational Science; thank you to the colleagues, both old and new, that I have been fortunate to work alongside.

The final thanks goes to my family and wife Becca, for their love and support.

Contents

List of Figures

List of Tables

Chapter 1

Near-Term Quantum Computation for Quantum Chemistry

> Now I go to the question of how we can simulate with a computer the quantum mechanical effects. Can you do it with a new kind of computer – a *quantum* computer? . . . It's not a Turing machine, but a machine of a different kind.
>
> ---
>
> *Richard Feynman* [11]

1.1 The Qubit

Digital computation operates on the basis of binary information, whereby 'bits' – the logical unit of information – may be concatenated to store and manipulate more complex expressions. Each bit assumes one of two distinct values, 0 or 1, which can be realized physically by any system exhibiting two distinct states, such as an electrical signal that might be *on* or *off*. Alternatively, one can search for more exotic two-level systems in which to embed binary information, such as energy levels in a quantum mechanical system; examples of this are provided in Table 1.1. Engineering bits into such small-scale systems renders the information stored within susceptible to quantum mechanical effects, revealing a wealth of powerful computational possibilities. For example, while we still possess two orthogonal states $|0\rangle$ and $|1\rangle$, the principle of quantum superposition allows us to form linear combinations of these discrete values $\alpha\,|0\rangle + \beta\,|1\rangle$ where $|\alpha|^2 + |\beta|^2 = 1$. In part, this encodes the probabilistic nature of quantum physics, the *Born rule*; if one takes a measurement of their quantum bit - or *qubit* - then a value of 0 or 1 will be observed with probability $|\alpha|^2$ or $|\beta|^2$, respectively. This is made richer by the fact that the amplitudes $\alpha, \beta \in \mathbb{C}$ may in fact be complex quantities, allowing us to harness quantum interference and further diversify our computational toolkit.

The normalization criterion above implies that $\alpha = e^{i\phi_0}\cos(\theta), \beta = e^{i\phi_1}\sin(\theta)$

for some parameters $\theta, \phi_0, \phi_1 \in \mathbb{R}$ and therefore

$$\begin{aligned} |\psi\rangle &:= \alpha\,|0\rangle + \beta\,|1\rangle \\ &= e^{i\phi_0}\cos(\theta)\,|0\rangle + e^{i\phi_1}\sin(\theta)\,|1\rangle \\ &= e^{i\phi_0}\Big(\cos(\theta)\,|0\rangle + e^{i(\phi_1-\phi_0)}\sin(\theta)\,|1\rangle\Big). \end{aligned} \tag{1.1}$$

Looking at the final line of Equation (1.1) above, one notes that by factoring out one of the phases associated with α or β, we can make it global and therefore not measurable. For this reason, global phases are unphysical and can be disregarded; it is only the relative phase between basis states, in this case $\phi_1 - \phi_0$, that we are able to probe via quantum measurement. Relabeling $\phi \leftarrow \phi_1 - \phi_0$, our quantum state can be expressed without loss of generality as $|\psi\rangle = \cos(\theta)\,|0\rangle + e^{i\phi}\sin(\theta)\,|1\rangle$. We find that just two parameters $\theta, \phi \in \mathbb{R}$ are sufficient to fully characterize the physical state of our qubit, which evokes the popular geometrical representation of a qubit as a point of the unit 2-sphere as per Figure 1.1.

Qubit Platform	**Physical Realization of Two-Level System**
Superconducting	Qubits are represented by two energy levels in a superconducting circuit containing a Josephson junction. They are controlled using microwave pulses and operate at the millikelvin temperature range to maintain coherence.
Ion-Trap	Qubits are embedded in two stable hyperfine levels within the electronic ground state of an ion. The ion is held in place using electromagnetic fields (Paul traps) and control is achieved through laser pulses or microwave radiation.
Neutral Atom	Neutral atoms (i.e. not ionized) are cooled and trapped using lasers (optical tweezers), while qubits are stored in two stable hyperfine energy levels at the ground states of the atom. To enable coupling between qubits, neutral atoms may also be excited to high-energy Rydberg states where they more readily interact.
Photonic	Qubits are encoded in properties of individual photons, such as their polarization (horizontal vs. vertical) or the path they travel. These qubits are especially useful for transmitting quantum information over long distances, such as in quantum communication.
NV-Centre	A nitrogen-vacancy (NV) centre in diamond is a type of defect in its crystal structure; qubits are stored in the electronic states associated with this defect. These systems can be controlled with microwaves and measured via fluorescence; they remain stable even at room temperature.

Table 1.1: Different example two-level systems for realizing a qubit

The orthonormality of $|0\rangle, |1\rangle$ allows us identify them with unit vectors in $\mathbb{C}^2$, canonically

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.2}$$

so that

$$|\psi\rangle = \begin{pmatrix} \cos(\theta) \\ e^{i\phi}\sin(\theta) \end{pmatrix}. \tag{1.3}$$

Vector states of this form live on the surface of the so-called *Bloch* sphere, named after 20th century physicist Felix Bloch. However, this is far from being the most general notion of a qubit state. In fact, the full interior of the Bloch sphere encodes quantum information, but in order to describe it we need to formalize our language surrounding qubits. In the above, we viewed single-qubit quantum states as unit vectors in $\mathbb{C}^2$; more generally, quantum systems live in so-called *Hilbert* spaces. Speaking abstractly, a Hilbert space is defined to be a complex vector space with a notion of angle/overlap between vectors, which could be thought of as a generalisation of the dot-product in standard Euclidean space. This is called an inner-product and is commonly denoted by angled brackets $\langle \cdot, \cdot \rangle$, so that for vectors ψ, ϕ the quantity $\langle \psi, \phi \rangle \in \mathbb{C}$ provides some sort of comparative metric that relates vectors to each other in the Hilbert space. If $\langle \psi, \phi \rangle = 0$, then the vectors are orthogonal with respect to the chosen inner product. Inner products also induce what are called *norms* – a notion of vector magnitude and are denoted $|| \cdot ||$. Specifically, the norm of a vector is $||\psi|| = \sqrt{|\langle \psi, \psi \rangle|} \in \mathbb{R}$ and the distance between vectors is evaluated as $||\psi - \phi||$.

In this work we refer to Hilbert spaces with the script face $\mathscr{H}$. This provides a very general framework in which to describe complex vectors and may even accommodate infinite-dimensional vector spaces. In quantum computing, one may typically make the restriction to finite dimensions for practical uses, but bear in mind that the flexibility of Hilbert space results in broad applicability. For example, when describing open quantum systems – a useful tool in quantum error

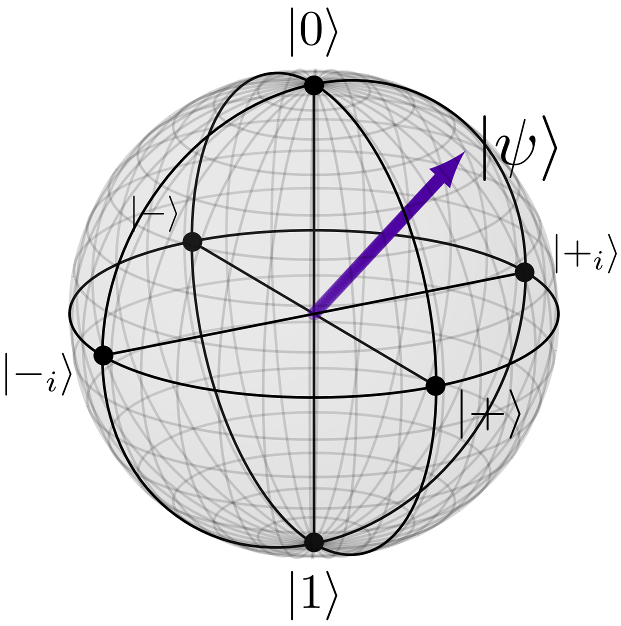


Figure 1.1: The Bloch sphere representation of single-qubit quantum states.

mitigation – one might need to describe the coupling of a quantum computer with its environment to understand how to mitigate against the resulting undesirable effects. Dealing with infinite dimensional space comes with its own difficulties, but this is thankfully circumvented for much of quantum computing research.

With this abstract Hilbert space framework in place, we now need a way of talking about mappings between vectors $|\psi\rangle \to |\phi\rangle$, which one might also view as a transformation of the global space $\mathscr{H} \to \mathscr{H}$, such as a change of basis. The mathematical terminology describing this is the linear operator, namely a map $O : \mathscr{H} \to \mathscr{H}$ between vectors. In particular, it is the *bounded* linear operators that are relevant here, ensuring that transformations are "well-behaved" in some sense, which we denote again with script face as $\mathscr{B}(\mathscr{H})$. We shall call an operator satisfying $O = O^\dagger$ *Hermitian*, which describe the "observable" objects in our space that we will later want to measure, subsequently yielding properties such as energy. It turns out that the space of bounded linear operators on a Hilbert space is again a Hilbert space of a higher dimension, meaning it possesses its own inner-product structure, induced norm and so on. In the finite dimensional case, all linear operators are bounded, so this is an occupational hazard for infinite dimensions that we need not concern ourselves with. As such, we may concretely identify linear operators with complex matrices, with the familiar idea of matrices mapping vectors onto other vectors.

We will remain on the abstract level for a moment longer, before settling into the more grounded setting of quantum computing. So far, we have been thinking of quantum states as unit vectors in a Hilbert space, visualized in the single-qubit case as the surface of the Bloch sphere in Figure 1.1 above. However, consider how we as an observer could possibly interface with the state of a quantum system, which will typically be hidden from us. Instead, we will have an observable quantity in mind and a goal of evaluating it with respect to some carefully prepared, but unknown, quantum state. As such, one way of viewing a quantum state is as a mapping from observables onto real numbers, or mathematically speaking a linear functional $\omega : \mathscr{B}(\mathscr{H}) \to \mathbb{R}$. The vector states that we have encountered thus far can be viewed in this way via the inner product $O \to \langle \psi, O\psi \rangle$. More convenient is the Dirac notation, where we write $\langle \psi | O | \psi \rangle = \langle \psi, O\psi \rangle$ and will be the convention going forward. However, another way to write this is $\langle \psi | O | \psi \rangle = \mathrm{Tr}(O\rho)$, where $\rho = |\psi\rangle \langle \psi| \in \mathscr{B}(\mathscr{H})$ is the outer-product of our vector state $|\psi\rangle$. In this way, one may also view the state at the level of the linear operators as a linear functional mapping $O \to \mathrm{Tr}(O\rho)$. For a finite two-level system such as a qubit, we can take $\mathscr{H} = \mathbb{C}^2$ and $\mathscr{B}(\mathscr{H}) = \mathbb{C}^{2\times 2}$, although the framework accommodates more general units of quantum information, including higher-level systems such as qutrits (3-level) and qudits (arbitrary d-level, $d \in \mathbb{N}$).

The operator ρ has a few properties, for example the sum of its diagonal elements is $\mathrm{Tr}(\rho) = 1$, again revealing a probability distribution of measurement outcomes. Furthermore, it is Hermitian ($\rho = \rho^\dagger$) and positive semi-definite ($\rho \geq 0$), meaning all its eigenvalues are non-negative. These properties encompass a broader class of operator ρ, namely the *density* operators $\mathscr{D}(\mathscr{H}) \subset \mathscr{B}(\mathscr{H})$ that need not possess the outer-product form above. In fact, the vector states are a special case of density operator, distinguished by the additional property $\gamma(\rho) \coloneqq \mathrm{Tr}(\rho^2) = 1$, a quantity we shall henceforth refer to as the purity of a qubit state. States for which $\gamma(\rho) = 1$ are called *pure* and define one-dimensional projectors, while in general we will have $\gamma(\rho) \leq 1$. We introduce the Pauli 'spin' operators $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3) = (X, Y, Z)$ where

$$\sigma_1 = X \coloneqq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = Y \coloneqq \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = Z \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.4}$$

and i is the imaginary unit which, together with the identity operator $\sigma_0 = I$, form a basis of $\mathbb{C}^{2\times 2}$. Note the Pauli operators satisfy $XYZ = iI$. The general expression for a single-qubit density operator over the Pauli basis is

$$\rho = \frac{1}{2}\big(I + \boldsymbol{r} \cdot \boldsymbol{\sigma}\big) = \frac{1}{2}\big(I + xX + yY + zZ\big) = \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} \tag{1.5}$$

for real numbers $\boldsymbol{r} = (x, y, z)$, that give its cartesian coordinate in the Bloch sphere. Note that the Pauli operators are Hermitian $(X = X^\dagger, Y = Y^\dagger, Z = Z^\dagger)$ and traceless, $(\mathrm{Tr}(X) = \mathrm{Tr}(Y) = \mathrm{Tr}(Z) = 0)$, so $\rho = \rho^\dagger$ and $\mathrm{Tr}(\rho) = 1$ is satisfied by the linearity of the trace operation.

The purity of this general state is given as

$$\gamma(\rho) = \frac{1}{4}\big[(1+z)^2 + 2(x-iy)(x+iy) + (1-z)^2\big] = \frac{1}{2}\big(1 + x^2 + y^2 + z^2\big) \tag{1.6}$$

and therefore the pure states are those with $x^2 + y^2 + z^2 = 1$. States for which $\mathrm{Tr}(\rho^2) < 1$ are called *mixed*, with the extreme case $x = y = z = 0$ found at the centre of the sphere and is referred to as the maximally mixed state $\frac{1}{2}I$ with purity $\gamma(\frac{1}{2}I) = \frac{1}{2}$.

Realizing a qubit is one thing, but for quantum computation to be possible we need to be able to manipulate its state by moving around the Bloch sphere; to do so, we need a formal definition for what it means to transform a quantum state into another. This motivates the notion of a quantum channel: a mapping between spaces of density operators that describes the valid operations on our quantum system. An appropriate definition is a map $\Phi : \mathscr{D}(\mathscr{H}) \to \mathscr{D}(\mathscr{H})$ that is completely-positive and trace-preserving (CPTP), i.e. $\rho \geq 0 \Rightarrow \Phi(\rho) \geq 0$ and $\mathrm{Tr}\,(\rho) = 1 \Rightarrow \mathrm{Tr}\,(\Phi(\rho)) = 1$, thus ensuring we remain within the space of density operators.

Choi's theorem on completely positive maps [12] gives us a convenient formulation of a quantum channel: there exist operators $\mathcal{K} = \{K_i\}_{1\leq i\leq d^2}$, where $K_i \in \mathscr{B}(\mathscr{H})$ and d is the dimension of $\mathscr{H}$, satisfying $\sum_{K\in\mathcal{K}} K^\dagger K = I$ (referred to as *Kraus* operators) such that

$$\Phi(\rho) = \sum_{K\in\mathcal{K}} K\rho K^\dagger \quad \forall\, \rho \in \mathscr{D}(\mathscr{H}). \tag{1.7}$$

The number of Kraus terms $|\mathcal{K}|$ defines the rank of this channel, which provides a rudimentary metric for the purity of Φ.

The rank-1 channel is a special case, whereby $\mathcal{K}$ contains just a single element U and thus $U^\dagger U = UU^\dagger = I$. This is the definition of a unitary map, and due to the cyclicity of trace it preserves both the norm and purity of a quantum state: $\mathrm{Tr}(U\rho U^\dagger) = \mathrm{Tr}(\rho U^\dagger U) = \mathrm{Tr}(\rho)$ and similarly $\mathrm{Tr}(U\rho^2 U^\dagger) = \mathrm{Tr}(\rho^2)$. We can derive a general matrix form of a single-qubit unitary operator

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{1.8}$$

with complex entries $a, b, c, d \in \mathbb{C}$ by asserting

$$\begin{aligned} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} &= \quad U^\dagger U = \begin{pmatrix} |a|^2 + |c|^2 & \overline{a}b + \overline{c}d \\ \overline{b}a + \overline{d}c & |b|^2 + |d|^2 \end{pmatrix} \\ &= \quad UU^\dagger = \begin{pmatrix} |a|^2 + |b|^2 & a\overline{c} + b\overline{d} \\ c\overline{a} + d\overline{b} & |c|^2 + |d|^2 \end{pmatrix}. \end{aligned} \tag{1.9}$$

Inspecting the diagonal elements and subtracting pairwise across each row of Equation (1.8), one obtains $|c|^2 - |b|^2 = 0 \implies |b|^2 = |c|^2$, while subtracting cross-terms (i.e. top left of line one minus bottom right of line two and vice versa) yields $|a|^2 - |d|^2 = 0 \implies |a|^2 = |d|^2$. Combined with the fact that $|a|^2 + |c|^2 = 1 = |b|^2 + |d|^2$ and using the property $|e^{i\theta}| = 1$ for all angles $\theta \in \mathbb{R}$, we may in general express the complex entries of the single-qubit unitary U as $a = \cos(\theta)e^{i\alpha}, b = \sin(\theta)e^{i\beta}, c = \sin(\theta)e^{i\gamma}, d = \cos(\theta)e^{i\delta}$. Furthermore, following the argument of Equation (1.1), we may choose to impose $\alpha = 0$ without loss of generality, as it can be removed as a global phase. To find a relation between the remaining parameters β, γ, δ, we look to the off-diagonals in Equation (1.9); for example, $0 = \overline{a}b + \overline{c}d = \cos(\theta)\sin(\theta)[e^{i\beta} + e^{i(\delta-\gamma)}]$. Given that this must be identically zero for all parametrizations, it must be the case that $e^{i\beta} + e^{i(\delta-\gamma)} = 0$ and so, rearranging, $e^{i\delta} = -e^{i(\beta+\gamma)}$. Finally, all single-qubit unitaries may be parametrized with three quantities θ, β, γ, with matrix representation

$$U(\theta, \beta, \gamma) = \begin{pmatrix} \cos(\theta) & \sin(\theta)e^{i\beta} \\ \sin(\theta)e^{i\gamma} & -\cos(\theta)e^{i(\beta+\gamma)} \end{pmatrix}. \tag{1.10}$$

1.2 Gate-Based Computation and Circuits

In the previous Section 1.1 we introduced the density operator framework for representing general quantum states. However, in quantum computation we are more often than not interested in preparing a pure state on our quantum device. The density matrix formulation is particularly useful when trying to deal with quantum noise, since this typically affects the purity of our state and can no longer be represented as a vector; this framework will be used extensively in Section 4.7 where we introduce an error mitigation technique that deals directly with the purity. Until then, it will be convenient to restrict ourselves to the surface of the Bloch sphere and return to the vector representation of pure quantum states $|\psi\rangle$, which may be viewed as unit vectors in a Hilbert space.

In Equation (1.2) we made a choice of basis in which to define our pure quantum states. One might note that the canonical, 'computational' basis we chose is the eigenbasis of the Pauli Z operator introduced in Equation (1.4), but we may just as well have chosen the eigenbases of the X or Y operator, which are

$$\begin{aligned} |+\rangle &= \frac{1}{\sqrt{2}}\big(\,|0\rangle + |1\rangle\,\big),\; |-\rangle = \frac{1}{\sqrt{2}}\big(\,|0\rangle - |1\rangle\,\big) \qquad (X \text{ basis}) \\ |+_i\rangle &= \frac{1}{\sqrt{2}}\big(\,|0\rangle + i\,|1\rangle\,\big),\; |-_i\rangle = \frac{1}{\sqrt{2}}\big(\,|0\rangle - i\,|1\rangle\,\big) \quad (Y \text{ basis}). \end{aligned} \tag{1.11}$$

In Table 1.2 we show how each of the Pauli operators acts on these basis states, noting how they act as logic gates on our quantum unit of information; for example, with respect to the Z-basis, the X operator is equivalent to the classical

State	X	Y	Z
$\|0\rangle$	$\|1\rangle$	$i\|1\rangle$	$\|0\rangle$
$\|1\rangle$	$\|0\rangle$	$-i\|0\rangle$	$-\|1\rangle$
$\|+\rangle$	$\|+\rangle$	$-i\|-\rangle$	$\|-\rangle$
$\|-\rangle$	$-\|-\rangle$	$i\|+\rangle$	$\|+\rangle$
$\|+_i\rangle$	$i\|-_i\rangle$	$\|+_i\rangle$	$\|-_i\rangle$
$\|-_i\rangle$	$-i\|+_i\rangle$	$-\|-_i\rangle$	$\|+_i\rangle$

Table 1.2: The action of Pauli operators on the basis states in Eqs. (1.2) and (1.11).

NOT operation $0 \leftrightarrows 1$. Geometrically, the Pauli X, Y, Z operations are reflections through the axes of the Bloch sphere, where it is seen in Figure 1.1 that each x, y, z-axis is spanned by the eigenbasis of the respective Pauli. It is also possible to perform partial rotations around each axis, rather than the full 180°; the rotation gates

$$\begin{aligned} R_x(\theta) &:= \exp\Big(-i\frac{\theta}{2}X\Big) = \cos\big(\theta/2\big)I - i\sin\big(\theta/2\big)X, \\ R_y(\theta) &:= \exp\Big(-i\frac{\theta}{2}Y\Big) = \cos\big(\theta/2\big)I - i\sin\big(\theta/2\big)Y, \\ R_z(\theta) &:= \exp\Big(-i\frac{\theta}{2}Z\Big) = \cos\big(\theta/2\big)I - i\sin\big(\theta/2\big)Z \end{aligned} \tag{1.12}$$

allow us to traverse the the surface of the Bloch sphere continuously.

It is useful to be able to transform between the different bases in Equations (1.2) and (1.11) to probe different parts of the Bloch sphere in Figure 1.1. For example, a unitary operator U satisfying $U|0\rangle = |+\rangle$ would require

$$U_{0,0} = \langle 0| U |0\rangle = \langle 0|+\rangle = \frac{1}{\sqrt{2}}, \quad U_{1,0} = \langle 1| U |0\rangle = \langle 1|+\rangle = \frac{1}{\sqrt{2}}, \tag{1.13}$$

where we have used the definition of $|+\rangle$ in Equation (1.11). Similarly, to satisfy $U|1\rangle = |-\rangle$ we need

$$U_{0,1} = \langle 0| U |1\rangle = \langle 0|-\rangle = \frac{1}{\sqrt{2}}, \quad U_{1,1} = \langle 1| U |1\rangle = \langle 1|-\rangle = -\frac{1}{\sqrt{2}} \tag{1.14}$$

and therefore yields the unitary $U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. This is the *Hadamard* gate that we shall henceforth refer to as "Had" which takes a qubit in a discrete level $|0\rangle$ or $|1\rangle$ and rotates into an equal superposition state.

We can follow a similar procedure to construct the matrix representation of a unitary U mapping $|+\rangle \mapsto |+_i\rangle$ and $|-\rangle \mapsto |-_i\rangle$. However, we shall first derive the matrix form of the product $U\mathrm{Had}$ and use the definition of the Hadamard gate to extract U, together with the fact that $\mathrm{Had}^2 = I$. This time, we need

$$\begin{aligned} \langle 0| U |+\rangle &= \langle 0| U\mathrm{Had} |0\rangle = \frac{1}{\sqrt{2}}, \quad \langle 1| U |+\rangle = \langle 1| U\mathrm{Had} |0\rangle = \frac{i}{\sqrt{2}}, \\ \langle 0| U |-\rangle &= \langle 0| U\mathrm{Had} |1\rangle = \frac{1}{\sqrt{2}}, \quad \langle 1| U |-\rangle = \langle 1| U\mathrm{Had} |1\rangle = -\frac{i}{\sqrt{2}} \end{aligned} \tag{1.15}$$

so that $U\text{Had} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$ and using the fact that $\text{Had}^2 = I$ we retrieve the matrix representation $U = (U\text{Had})\text{Had} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$.

This is the phase gate S that induces a complex phase on the state it is applied to. As an aside, a gate that we will not use within this work, but is hugely important in fault-tolerant quantum computation [1], is the T gate, which can be obtained as the square root of the phase gate $T = \sqrt{S}$. A pursuit of longer-term quantum computation is to engineer error-corrected T gates as an element of a standardized gate set going forward.

So far, our logical operators include many examples of single-qubit quantum gates: $X, Y, Z, R_x, R_y, R_z, \text{Had}, S, T$, all of which correspond with rotations around the Bloch sphere in Figure 1.1. However, single-qubit gates alone are insufficient to exceed the capabilities of classical computation. For that, we need to look past the Bloch sphere representation to be able to perform multi-qubit operations – entangling gates that, when combined with the single-qubit operations, unleash the full potential of quantum computation.

First of all, we need a robust language to talk about collections of qubits. A convenient mathematical framework to describe the composition of many subsystems is the tensor product space. Consider the general bipartite case in which we have two Hilbert spaces $\mathscr{H}_A$ and $\mathscr{H}_B$, with bases $\{|a\rangle_A\}_{a=1}^{N_A}, \{|b\rangle_B\}_{b=1}^{N_B}$ where N_A, N_B are the respective dimensions. A tensor product is a bilinear map $\otimes : \mathscr{H}_A \times \mathscr{H}_B \to \mathscr{H}_A \otimes \mathscr{H}_B$ whose composite elements take the form

$$|\psi\rangle_{AB} = \sum_{a,b} \alpha_{a,b} \, |a\rangle_A \otimes |b\rangle_B \tag{1.16}$$

where $\alpha_{a,b} \in \mathbb{C}$. The dimension of the resulting composite system $\mathscr{H} = \mathscr{H}_A \otimes \mathscr{H}_B$ is $N_A N_B$. This may be extended to an arbitrary number of subsystems $\{\mathscr{H}_b\}_b$ with dimensions N_b, where elements of the composition $\mathscr{H} = \bigotimes_b \mathscr{H}_b$ may be written

$$|\psi\rangle = \sum_{b_0, b_1, b_2, \ldots} \alpha_{b_0, b_1, b_2 \ldots} \, |b_0\rangle_0 \otimes |b_1\rangle_1 \otimes |b_2\rangle_2 \otimes \ldots \tag{1.17}$$

with total dimension $\prod_b N_b$.

One can also tensor strings of Pauli operators together to address the individual qubits in the composite system, $P = P_0 \otimes P_1 \otimes P_2 \otimes \ldots$ where $P_i \in \{I, X, Y, Z\}$. For an N-qubit system, the set of all combinations of Paulis (including the identity) up to multiplication by a complex unit $\{1, -1, i, -i\}$ is called the *Pauli group* $\mathcal{P}_N \subset \mathscr{B}(\mathscr{H})$. Explicitly,

$$\mathcal{P}_N = \Big\{ \alpha P \mid \forall \, \alpha \in \{1, -1, i, -i\}, P \in \{I, X, Y, Z\}^{\otimes N} \Big\}. \tag{1.18}$$

The above applies for any Hilbert spaces, even with differing dimensions between tensor factors. To bring this back into the context of two-level qubit systems,

---

[1] Quantum hardware is currently very sensitive to noise, but in the future large-scale error correction will allow quantum computation to be resilient to hardware faults through the detection and correction of errors on-the-fly. This is different to error *mitigation* as discussed later in Chapter 4.

we have $\mathscr{H}_b = \mathbb{C}^2 \; \forall b$. In the bipartite case, adopting again the Z-basis $\{|0\rangle, |1\rangle\}$ for each individual qubit, a general 2-qubit state has the form

$$|\psi\rangle = \alpha_{00} |0\rangle_A \otimes |0\rangle_B + \alpha_{01} |0\rangle_A \otimes |1\rangle_B + \alpha_{10} |1\rangle_A \otimes |0\rangle_B + \alpha_{11} |1\rangle_A \otimes |1\rangle_B \quad (1.19)$$

where $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$. The power of tensoring qubits together in this way is we now have access to states that cannot be expressed with the two systems in isolation, meaning $|\psi\rangle \neq |\psi\rangle_A \otimes |\psi\rangle_B$ for any pure state on each subsystem; the composite system is greater than the sum of its parts.

It is possible to go from the composite system back to one of its subsystems, say $\mathscr{B}(\mathscr{H}_B)$, via a *partial trace* over A, the unique linear map $\mathrm{Tr}_A : \mathscr{B}(\mathscr{H}_A \otimes \mathscr{H}_B) \to \mathscr{B}(\mathscr{H}_B)$ satisfying $\mathrm{Tr}_A(A \otimes B) = \mathrm{Tr}(A)B$ for all $A \in \mathscr{B}(\mathscr{H}_A), B \in \mathscr{B}(\mathscr{H}_B)$. Similarly, we can trace over system B via Tr_B to obtain elements in $\mathscr{B}(\mathscr{H}_A)$. Example 1.1 demonstrates the existence of entangled quantum states, namely those which cannot be described by pure states on the isolated subsystems. The question is, how do we actually prepare such a state in practice?

Example 1.1: Bell State Partial Trace

Take the Bell state

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}\Big(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B \Big); \quad (1.20)$$

our goal is to show that information is distributed across the two constituents qubits and cannot be separated by means of a partial trace.
With $\rho_{AB} = |\psi\rangle\langle\psi|$,

$$\begin{aligned} \mathrm{Tr}_A(\rho_{AB}) &= \frac{1}{2}\sum_{i=0}^{1}(\langle i|_A \otimes I_B)\Big(|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B + |0\rangle\langle 1|_A \otimes |0\rangle\langle 1|_B \\ &\qquad + |1\rangle\langle 0|_A \otimes |1\rangle\langle 0|_B + |1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_B \\ &\quad \Big)(|i\rangle_A \otimes I_B) \\ &= \frac{1}{2}\Big(|0\rangle\langle 0|_B + |1\rangle\langle 1|_B \Big) \\ &= \frac{1}{2} I_B. \end{aligned} \quad (1.21)$$

However, recall from Section 1.1 that this is the maximally mixed state, implying qubits A and B are entangled.
There is a natural entanglement measure for bipartite systems, namely the Von-Neumann entropy of the reduced density matrix of either subsystem:

$$S(\rho_{AB}) \coloneqq -\rho_A \log(\rho_A) = -\rho_B \log(\rho_B). \quad (1.22)$$

The fact that each subsystem yields the same entropy is non-trivial and is a consequence of the Schmidt decomposition. For our Bell state example, we obtain $S(\rho_{AB}) = \log(2) \approx 0.693$, which is maximal for a 2-qubit system.

It is clear that local operations on each subsystem are not going to cut it. What we need is an operation that applies a single-qubit gate V to a target qubit, labelled t, conditional on the state of a control qubit, labelled c. This would allow the qubits to become correlated and means realizing the map

$$\begin{aligned} CV : \mathscr{H}_c \otimes \mathscr{H}_t &\to \mathscr{H}_c \otimes \mathscr{H}_t, \\ \left|0\right\rangle_c \otimes \left|x\right\rangle_t &\mapsto \left|0\right\rangle_c \otimes \left|x\right\rangle_t \\ \left|1\right\rangle_c \otimes \left|x\right\rangle_t &\mapsto \left|1\right\rangle_c \otimes V\left|x\right\rangle_t \end{aligned} \tag{1.23}$$

where $x \in \{0, 1\}$. As a matrix, this looks like

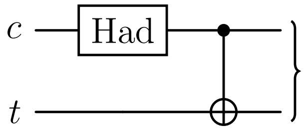
$$\begin{pmatrix} I_t & \mathbf{0} \\ \mathbf{0} & V_t \end{pmatrix}_c = \underbrace{\frac{1}{2}\left(I_c + Z_c\right)}_{:=\mathbb{P}_+} \otimes I_t + \underbrace{\frac{1}{2}\left(I_c - Z_c\right)}_{:=\mathbb{P}_-} \otimes V_t \tag{1.24}$$

where $\mathbb{P}_\pm = \frac{1}{2}(I \pm Z)$ is the projector onto the ± 1-eigenspace of the Pauli Z operator, since $\mathbb{P}_+\left|0\right\rangle = \left|0\right\rangle, \mathbb{P}_-\left|1\right\rangle = \left|1\right\rangle$ and $\mathbb{P}_+\left|1\right\rangle = \mathbb{P}_-\left|0\right\rangle = 0$.

It is very common to take $V_t = X$, otherwise known as the quantum NOT operation, to produce the CNOT, or "controlled-NOT", gate. The $\mathrm{CNOT}_{c,t}$ gate has the action

$$\begin{aligned} &\left|0\right\rangle_c \otimes \left|0\right\rangle_t \mapsto \left|0\right\rangle_c \otimes \left|0\right\rangle_t, \left|0\right\rangle_c \otimes \left|1\right\rangle_t \mapsto \left|0\right\rangle_c \otimes \left|1\right\rangle_t, \\ &\left|1\right\rangle_c \otimes \left|0\right\rangle_t \mapsto \left|1\right\rangle_c \otimes \left|1\right\rangle_t, \left|1\right\rangle_c \otimes \left|1\right\rangle_t \mapsto \left|1\right\rangle_c \otimes \left|0\right\rangle_t. \end{aligned} \tag{1.25}$$

We are now able to construct the Bell state of Example 1.1 from quantum gates: $\frac{1}{\sqrt{2}}\left(\left|0\right\rangle_c \otimes \left|0\right\rangle_t + \left|1\right\rangle_c \otimes \left|1\right\rangle_t\right) = \mathrm{CNOT}_{c,t}(\mathrm{Had}_c \otimes I_t)(\left|0\right\rangle_c \otimes \left|0\right\rangle_t)$. For visual ease, it is often more instructive to present sequences of quantum logical operations in their graphical circuit form. A quantum circuit is characterized by a series of horizontal lines (or *wires*), each representing individual qubits with the relevant sequence of operations applied in blocks placed atop the wires. The gates are to be applied left-to-right and, unless specified otherwise, it is assumed each qubit is initialized in its zero state. To illustrate this, the Bell state preparation can be portrayed as



$$= \frac{1}{\sqrt{2}}\left(\left|0\right\rangle_c \otimes \left|0\right\rangle_t + \left|1\right\rangle_c \otimes \left|1\right\rangle_t\right).$$

A more complete list of quantum gates, their circuit symbols and matrix representations is provided in Table 1.3.

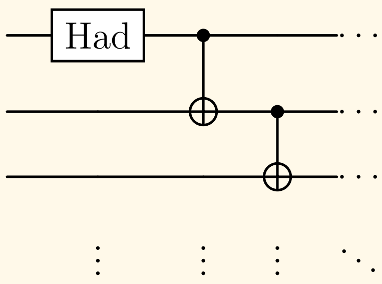
For brevity, throughout the remainder of this work we may drop the tensor symbols $\otimes$ when referring to collections of qubits. The following notations are all equivalent:

$$\left|b_0\right\rangle \otimes \left|b_1\right\rangle \otimes \left|b_2\right\rangle \otimes \cdots = \left|b_0 b_1 b_2 \ldots\right\rangle = \left|\boldsymbol{b}\right\rangle = \left|b\right\rangle \tag{1.26}$$

where $b_i \in \{0, 1\}$ and b is the integer corresponding with the binary expansion $\boldsymbol{b}$, i.e. $b = \sum_i b_i 2^{N-1-i}$. For example, $\left|1\right\rangle \otimes \left|0\right\rangle \otimes \left|1\right\rangle = \left|101\right\rangle = \left|5\right\rangle$. As an aside, note the gate-based approach is not the only form of quantum computation; quantum *annealers* are analog devices that are designed to simulate Ising models in accordance with the adiabatic theorem [13].

Example 1.2: GHZ State Preparation

The process for preparing Bell states can be continued to produce entangled quantum states of arbitrary size N. The circuit



produces states $|\psi_N\rangle = \frac{1}{\sqrt{2}}\Big(\underbrace{|0\rangle \otimes \cdots \otimes |0\rangle}_{N \text{ times}} + \underbrace{|1\rangle \otimes \cdots \otimes |1\rangle}_{N \text{ times}}\Big)$, noting the Bell state is the $N = 2$ case. For $N \geq 3$ we obtain the Greenberger–Horne–Zeilinger (GHZ) states. Preparing these entangled states over increasing numbers of qubits can give a rough indication of the quality of a quantum device, probing both the width and depth of quantum circuit it can sustain.

A useful property of GHZ states is we know exactly what the expected measurement outcome should be: 50% the all zero state and 50% the all one state. Therefore, it is easy to quantify how close the quantum computer is to the expected outcome. The relevant metric here is the *state fidelity*, which for density matrices ρ, σ is defined as $F(\rho, \sigma) = \big(\operatorname{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\big)^2$. If ρ is a pure state, which is the case here with $\rho = |\psi_N\rangle\langle\psi_N|$, then ρ is a projector $\rho^2 = \rho$ and therefore is its own square root, i.e. $\sqrt{\rho} = \rho$. Then, if σ is the noisy state prepared on hardware, we can write

$$F(|\psi_N\rangle\langle\psi_N|, \sigma) = \big(\operatorname{Tr}\sqrt{|\psi_N\rangle\langle\psi_N|\sigma|\psi_N\rangle\langle\psi_N|}\big) = \langle\psi_N|\sigma|\psi_N\rangle\,.$$

Therefore, with $|\psi_N\rangle$ the N-qubit GHZ state as above, we obtain $F(|\psi_N\rangle\langle\psi_N|, \sigma) = \frac{1}{2}[\langle\mathbf{0}|\sigma|\mathbf{0}\rangle + \langle\mathbf{1}|\sigma|\mathbf{1}\rangle] + \Re[\langle\mathbf{0}|\sigma|\mathbf{1}\rangle]$, where $\Re$ denotes the real part of a complex quantity. This can be evaluated by taking measurements of the noisy state σ prepared in hardware. The diagonal part $\frac{1}{2}[\langle\mathbf{0}|\sigma|\mathbf{0}\rangle + \langle\mathbf{1}|\sigma|\mathbf{1}\rangle]$ describes population statistics, while the offdiagonal $\Re[\langle\mathbf{0}|\sigma|\mathbf{1}\rangle]$, evaluated through measurements in the $|+\rangle / |-\rangle$ basis, describes the coherence.

In Figure 4.2 we present the decay in GHZ population fidelity up to $N = 27$ qubits for a suite of IBM Quantum Falcon processors, which is calculated as $\frac{1}{2}(\sqrt{p_\mathbf{0}} + \sqrt{p_\mathbf{1}})^2$ where $p_\mathbf{0}, p_\mathbf{1}$ are the relative frequencies with which we measure the all zero and all one states in the diagonal Z basis, respectively.

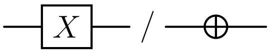
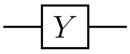
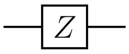
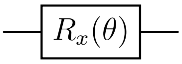
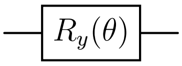
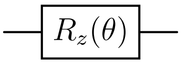
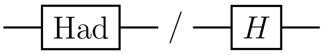
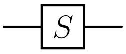
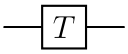
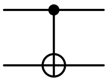
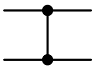
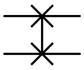
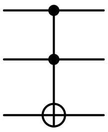
Gate	Circuit symbol	Matrix representation
X/NOT		$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$
Y		$\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}$
Z		$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$
R_x		$\begin{pmatrix}\cos(\theta/2) & -i\sin(\theta/2)\\ -i\sin(\theta/2) & \cos(\theta/2)\end{pmatrix}$
R_y		$\begin{pmatrix}\cos(\theta/2) & -\sin(\theta/2)\\ \sin(\theta/2) & \cos(\theta/2)\end{pmatrix}$
R_z		$\begin{pmatrix}e^{-i\theta/2} & 0\\ 0 & e^{+i\theta/2}\end{pmatrix}$
Had/H		$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$
S		$\begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix}$
T		$\begin{pmatrix}1 & 0\\ 0 & e^{i\pi/4}\end{pmatrix}$
CX/CNOT		$\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}$
CZ		$\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}$
SWAP		$\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$
Toffoli		$\begin{pmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\end{pmatrix}$

Table 1.3: Common quantum gates, their circuit symbols and matrix representations.

1.2.1 Gate Sets and Universality

The operations introduced in Section 1.2 and listed in Table 1.3 are over-complete; many of the gates can be obtained via products of others. This raises the question: what is the minimal set of quantum gates one needs to be able to perform all the operations they might desire? More succinctly, what is the minimal gate set for universal quantum computation? Ideally, one would engineer their quantum device to perform a small subset of gates to a high level of precision and build more complex operations from combinations of this reduced set of native gates.

First of all, we need a definition for what constitutes universality in quantum computation. The *special unitary group* $\mathrm{SU}(d)$ consists of unitary operations U with dimension $d \times d$ having determinant $\det(U) = 1$. A set of quantum gates $\mathcal{G}$ is called *universal* if there exists $d_0 \in \mathbb{N}$ such that for all $d \geq d_0$ and $U \in \mathrm{SU}(d)$ there exists an operation generated by $\mathcal{G}$ that is arbitrarily close to U. In mathematical terms, the subgroup generated by $\mathcal{G}$ is *dense* in $\mathrm{SU}(d)$.

Earlier in this section we introduced the T gate, which is of particular focus in fault tolerant quantum computation. Taken together with the Clifford gates $\{\mathrm{CNOT}, \mathrm{Had}, S\}$, this forms a universal gate set often dubbed "*Clifford* + T". However, there are other examples of universal gate sets, such as $\{X, \sqrt{X}, R_z, \mathrm{CZ}\}$ found as the native operations on IBM Quantum Heron processors; the $\sqrt{X}$ gate here serves a similar purpose as the Hadamard gate, namely creating equal superpositions albeit with an additional phase. The most compact set is $\{\mathrm{Toffoli}, \mathrm{Had}\}$, noting that the Toffoli gate alone is sufficient to capture all of classical computation – adding the ability to create superpositions through Hadamard gates then graduates us to fully universal quantum computation [14, 15].

1.2.2 Quantum Measurement

The beguiling concept of measurement in quantum mechanics is central to the discipline. As we we shall discuss in Chapter 3 at greater length, quantum measurement cannot be thought of in the same way as in classical, "Newtonian" systems. In the classical picture, physical properties are presumed to exist independently of observation and the act of measurement simply uncovers some preexisting truth. In the same way that a ruler reveals the length of an object, it is not controversial to believe that its length is present irrespective of whether we choose to measure it. The same does not hold true of quantum systems. Instead, quantum systems are fundamentally probabilistic and their state is described by a wavefunction, as discussed in the former parts of this Chapter. The act of measurement induces a non-unitary transformation referred to as wavefunction *collapse*, wherein the system transitions from a superposition to a single eigenstate corresponding to the observed quantity. This transition is not predicted by the unitary evolution that governs the Schrödinger equation and consequently introduces a discontinuity that is profound in both its conceptual and practical implications.

The measurement problem has inspired a wide range of interpretive responses, each attempting to reconcile the apparent contradiction between the continuous evolution of quantum systems and the discrete outcomes observed in measurement. The Copenhagen interpretation [16, 17] treats wavefunction collapse as a fundamental feature of quantum mechanical theory and posits the observer as an inseparable element of the measurement process. By contrast, the many-worlds interpretation [18, 19] eliminates collapse by asserting that all possible outcomes occur simultaneously in ever-diverging parallel universes, and measurement aligns us to one of those possible realities. Regardless, the process of quantum measurement is non-reversible and destructive, with further potential for learning anything from the wavefunction being lost. This can occur intentionally through measurement by an observer, or unintentionally through leakage into the environment, thus causing a loss of coherence [2].

It is often stated that the power of a quantum computer lies in the exponentially vast space it can explore, but the nature of quantum measurement means it is not in general possible to learn the internal state of the system. This makes algorithm design very difficult; quantum algorithms need to be designed carefully to circumvent the measurement problem, for example by exploiting quantum interference effects to isolate amplitudes of the wavefunction that have been designed to encode the target solution. The Quantum Fourier Transform (QFT) is typically how this is achieved in practice and is a common circuit primitive found in algorithms such as Quantum Phase Estimation (QPE).

We now describe the general framework of quantum measurement as adopted for this work. Given a set of measurable outcomes, we shall associate with each outcome m a measurement operator $\mathbb{M}_m$. Suppose now that we have a quantum state $|\psi\rangle$, then the probability of observing outcome m is given by

$$p(m) = \langle\psi| \mathbb{M}_m^\dagger \mathbb{M}_m |\psi\rangle . \tag{1.27}$$

Since the $p(m)$ are probabilities they should sum to one, specifically $\sum_m p(m) = \langle\psi| \sum_m \mathbb{M}_m^\dagger \mathbb{M}_m |\psi\rangle = 1$ and consequently the measurement operators must satisfy $\sum_m \mathbb{M}_m^\dagger \mathbb{M}_m = \mathbb{1}$, the so-called *completeness relation*. Post-measurement, the state of our quantum system is now

$$|\psi\rangle \rightarrow \frac{\mathbb{M}_m |\psi\rangle}{\sqrt{p(m)}}. \tag{1.28}$$

Define $F_m := \mathbb{M}_m^\dagger \mathbb{M}_m$, which satisfies $\sum_m F_m = \mathbb{1}$ from the completeness relation above, encoding a probability distribution over measurement outcomes. These F_m are positive operators, meaning $\langle\psi| F_m |\psi\rangle \geq 0$ for all vectors $|\psi\rangle$ (which do not need to be normalized). This positivity may be seen as $\langle\psi| F_m |\psi\rangle = ||\mathbb{M}_m |\psi\rangle ||^2 \geq 0$; the consequence is a non-negative spectrum, again speaking to an underlying probability distribution. The terminology for this generalised measurement scheme is a "positive operator-valued measure" (POVM), where each F_m is referred to as a POVM element, which provides a highly flexible framework and allows one to describe a very broad notion of what measurement means in quantum mechanics. For example, the POVM formalism permits "weak" measurements, which cause only a slight disturbance to the wavefunction, without resulting in total collapse.

A special case arises when the measurement operators $\mathbb{M}_m$ are projections, i.e. $\mathbb{M}_m^2 = \mathbb{M}_m$, which is the unique instance whereby the measurement operators are themselves the POVM elements, namely $F_m = \mathbb{M}_m$. To demonstrate projective measurements, suppose we have a quantum state $|\psi\rangle = \sum_m \alpha_m |m\rangle$. The

---

[2] A semantic point is that, in the case of decoherence, it only *appears* as though the wavefunction has collapsed as it is absorbed into the universal wavefunction encompassing all things.

measurement operator corresponding with outcome m is given by the projector $\mathbb{M}_m = |m\rangle\langle m|$, noting that $\sum_m |m\rangle\langle m| = \mathbb{1}$ as required. Then, the probability of observing m is given by $p(m) = \langle\psi|m\rangle\langle m|\psi\rangle = |\alpha_m|^2$. We see that this encodes the Born rule postulating the probabilistic nature of quantum measurement.

Typically, a quantum computer will facilitate measurement in a fixed basis, usually that of diagonal Pauli Z operators. This means the expansion of $|\psi\rangle$ above should be viewed in the computational basis where $|m\rangle$ is identified with a tensored binary bit string. However, we are often required to make measurements in different bases in order to measure any given observable; to achieve this, we need to perform a basis transformation prior to readout such that diagonal measurements alone are sufficient to probe any desired measurement basis. This is possible using just two quantum operations: the Hadamard and phase gates. Observe that

$$Z = \mathrm{Had} X \mathrm{Had} = \mathrm{Had} S^\dagger Y S \mathrm{Had}. \tag{1.29}$$

Therefore, to measure in the X basis we simply apply Had before readout, while for the Pauli Y measurement we apply $\mathrm{Had} \cdot S^\dagger$; this is depicted in circuit-form in Figure 1.2.

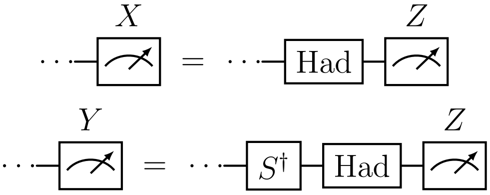


Figure 1.2: Change-of-basis operations facilitating measurement in the Pauli X, Y bases when only diagonal Pauli Z readout is available.

1.2.3 Exponentiating Pauli Strings

Given a Pauli operator $P \in \mathcal{P}_N$ and some angle $\theta \in \mathbb{R}$, we would like to implement the exponential $R_P(\theta) := e^{-i\frac{\theta}{2}P}$ as a quantum circuit; this can be achieved with $\mathcal{O}(N)$ gates. We shall first assume that P consists of just Pauli I, Z operators, i.e. $P \in \{I, Z\}^{\otimes N}$, with the qubit positions acted upon by Z indexed with the set $\mathcal{I}_Z$. Observe that

$$\begin{aligned} e^{-i\frac{\theta}{2}Z^{\otimes \mathcal{I}_Z}} |\boldsymbol{b}\rangle &= \Big(\cos(\theta/2) - i(-1)^{z(\boldsymbol{b})}\sin(\theta/2)\Big)|\boldsymbol{b}\rangle \\ &= e^{-i\frac{\theta}{2}(-1)^{z(\boldsymbol{b})}}|\boldsymbol{b}\rangle . \end{aligned} \tag{1.30}$$

where we have omitted the qubit positions that are identity. Therefore, we may realize this operation by storing $z(\boldsymbol{b})$, the parity of $|\boldsymbol{b}\rangle$ over $\mathcal{I}_Z$, in one of the qubits and applying to it an R_z gate, defined in Table 1.3.

In Equation (1.30), even parity results in a phase $e^{-i\theta/2}$ whereas odd parity yields $e^{+i\theta/2}$; these are obtained by application of $R_z(\theta)$ to each of $|0\rangle, |1\rangle$, respectively. The parity computation is accomplished via a 'cascade' of CNOT gates between adjacent qubits. We are now in a position to explicitly write down a quantum circuit that effects the exponentiation of a Pauli string consisting of I, Z operators, presented in Figure 1.3.

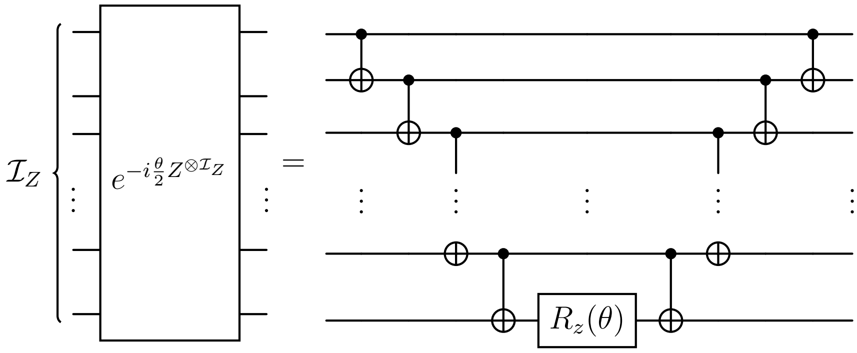


Figure 1.3: Circuit to realize the exponential $e^{i\theta P}$ where P consists of just Pauli I, Z operators. The qubits set to identity are omitted from the diagram.

This is the basic building block for exponentiating an arbitrary Pauli string, as we are able to make a change of basis from Pauli X, Y operators to a Pauli Z via the basis change described in Section 1.2.2. Once this transformation has been applied, the problem is reduced to that of before and we perform the same parity computation and R_z rotation. We end with the reverse transformation taking us back into the original basis, with the complete circuit presented in Figure 1.4. At most $4N + 1$ single-qubit gates and $2(N - 1)$ CNOT gates are required to realize the Pauli exponential ($6N - 1$ gates in total, meaning the depth-scaling is linear), with the worst case encountered when P consists solely of Pauli Y operators acting on all N qubit positions.

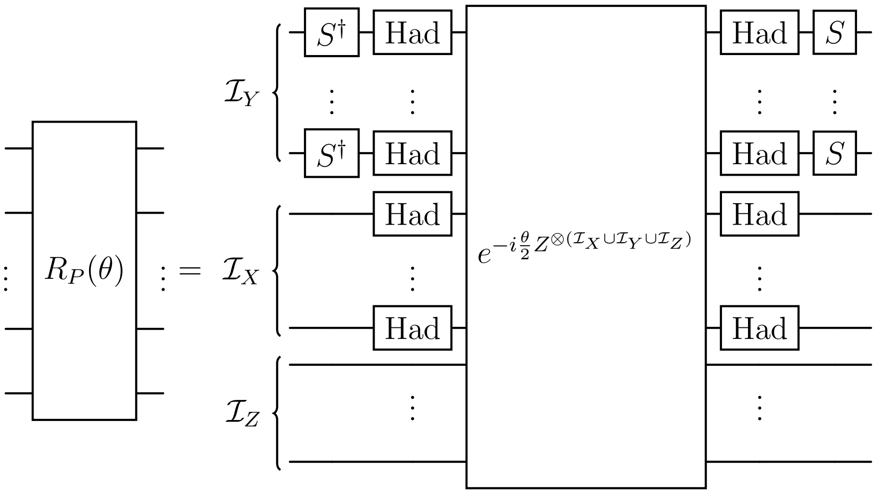


Figure 1.4: Circuit to realize the exponential $e^{i\theta P}$ where P is an arbitrary Pauli string. The qubits set to identity are omitted from the diagram.

It is noted that in the Trotterized circuits of the following Section 1.2.4 there will be many blocks of exponentiated Pauli strings of the form displayed in Figure 1.4, and it is possible that gate cancellation will occur between adjacent blocks where we have $SS^\dagger = \text{Had}^2 = \mathbb{1}$. Furthermore, the implementation presented here is not unique – different circuits yielding the same quantum states are possible.

1.2.4 Trotterization and Hamiltonian Simulation

For operators $A, B \in \mathscr{B}(\mathscr{H})$, we have $e^{A+B} = e^A e^B$ if and only if A commutes with B (i.e. $[A, B] = 0$), contrary to the familiar rules of exponentiation for numbers; in the more general case, the Lie product formula states

$$e^{A+B} = \lim_{n \to \infty} \left(e^{A/n} e^{B/n} \right)^n. \tag{1.31}$$

Alternatively, we may write $e^A e^B = e^C$ where the Baker-Campbell-Hausdorff (BCH) formula expands the generator C as

$$C = A + B + \frac{1}{2}[A, B] + \frac{1}{12}[A, [A, B]] - \frac{1}{12}[B, [A, B]] + \dots, \tag{1.32}$$

in which the commutators become increasingly nested for higher order terms in the expansion [20, 21].

The technique of *Suzuki-Trotter* truncates the BCH expansion to yield an approximation of the exponentiated operator sum. For example, given a sum of operators $\sum_k \theta_k A_k, \theta_k \in \mathbb{R}$, we may write the first-order Trotter expansion

$$e^{i \sum_k \theta_k A_k} \approx \left(\prod_k e^{i \frac{\theta_k}{n_\mathrm{T}} A_k} \right)^{n_\mathrm{T}}, \tag{1.33}$$

where the exponentiated Pauli operator $e^{i \frac{\theta_k}{n_\mathrm{T}} A_k}$ may be performed in-circuit as described in the previous Section 1.2.3 using $\mathcal{O}(N)$ quantum gates. We refer to the parameter $n_\mathrm{T} \in \mathbb{N}$ as the *Trotter number*, which controls the accuracy of the approximation (becoming exact as $n_\mathrm{T} \to \infty$).

Used in combination with Variational Quantum Algorithms (VQAs) of the sort introduced later in Section 1.4, $n_\mathrm{T} = 1$ is often sufficient to achieve high levels of accuracy since we expect the optimizer to counteract the Trotter error via its circuit parameterization [22]. It has also been observed that the ordering of Trotter terms has an impact on errors [23]; as such, there is contention over Trotterized UCC and whether it is well-defined as an ansatz, since the ordering of terms can induce variations in error beyond chemical accuracy, particularly for strongly correlated systems [24].

In Hamiltonian simulation, the goal is the realize the time evolved form e^{iHt} of some Hamiltonian $H = \sum_\ell H_\ell$ in-circuit for the purposes of analyzing its spectral properties through quantum computational techniques. Trotterization is a natural way to achieve this, although one must be mindful of the error that accumulates as we evolve for longer times. This is because, to first order

$$e^{iHt} = \prod_\ell e^{iH_\ell t} + \mathcal{O}(t^2). \tag{1.34}$$

We can combat this error, which is quadratic in time [21], by discretizing the evolution into K time slices of duration τ such that $t = K\tau$. Then, by observing that $e^{iHt} = (e^{iH\tau})^K$ one can approximate

$$e^{iHt} \approx \left(\prod_\ell e^{iH_\ell \tau} \right)^K \tag{1.35}$$

where K serves the same purpose as the Trotter number n_{T} above.

In doing this, we have confined the error accumulated at each step to $\mathcal{O}(\tau^2)$ so that the overall error is proportional to $K\tau^2 = t^2/K$. Therefore, a finer discretization results in a more accurate approximation, but reciprocally leads to increased circuit depths. There are also higher-order Trotter expansions that can be defined recursively [25, 26] and improves the error scaling to higher powers of τ; at order $2k$ for some $k \in \mathbb{N}$ the error in each step is $\mathcal{O}(\tau^{2k+1})$.

A branch of Hamiltonian simulation has pursued the development of stochastic approaches to circuit compilation, initiated with the qDRIFT algorithm [27]. The idea is that, while each circuit layer does not reproduce the Hamiltonian evolution operator exactly, the process approximates the operator *in expectation*. This leads to favourable asymptotic scaling that is decoupled from the number of terms in the Hamiltonian. Since that seminal work, there have been several advances in stochastic circuit compilation [28, 29] that produce incremental improvements in circuit depth scaling.

Trotterization is not the only method of Hamiltonian simulation; there is a rich class of algorithms based on the concepts of *block encoding* and Quantum Signal Processing (QSP). The combination of these techniques gives rise to the Quantum Singular Value Transform (QSVT) algorithm [30, 31], which applies functions (typically Chebyshev polynomials) to the singular values of a block in a unitary matrix. This is usually constructed as a walk operator and the work of Low et al. [32] demonstrates how this can be achieved. Note that while the overall walk operator is unitary, the matrix encoded within one of its blocks need not be – a Hamiltonian matrix, which is Hermitian and not in general unitary, can be encoded in a walk operator and subsequently subjected to some transformation. In the context of Hamiltonian simulation, the time evolution may be realized via an approximation of the exponential function by Chebyshev polynomials. QSVT is more general than Hamiltonian simulation alone, however, and provides a unified framework for algorithm design. For example, it encompasses algorithms such as Unstructured Search (Grover's), Eigenvalue Thresholding/Filtering and Quantum Phase Estimation [33].

1.2.5 Circuit Transpilation

When constructing a quantum circuit, one needs to be mindful of the quantum hardware that it will be executed on. The connectivity between qubits will vary device-to-device and therefore if a circuit tries to entangle qubits for which there does not exist a physical coupling channel, then we will run into problems. Take for example the heavy-hex topology found on IBM Quantum hardware, viewed in Figure 1.5. Each node of the graph represents a qubit, while edges between nodes indicate pairs of qubits that may be coupled natively on the device. The heavy-hex graph consists of qubit cycles of length 12. If a particular quantum circuit calls for the entanglement of qubits that are physically separated across the device, then we are faced with a dilemma.

Despite a physical connection not being available, it is in fact possible to entangle distant qubits via the quantum SWAP operation, given in Table 1.3. The effect of this gate is to exchange the quantum state of two qubits; therefore, if we wish to entangle two non-adjacent qubits and there exists a path of connected

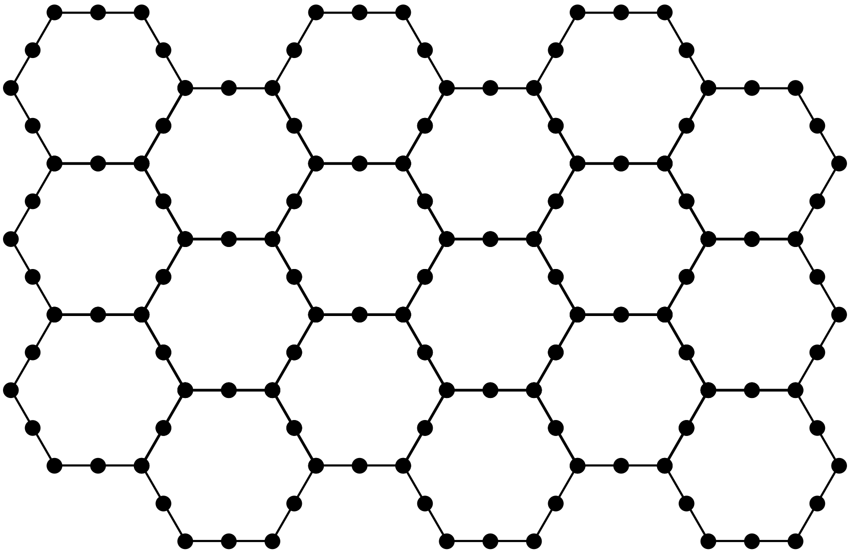


Figure 1.5: The heavy-hex qubit topology found on IBM Quantum hardware.

qubits between them, then iterated application of SWAP gates on neighbouring qubits will eventually bring the two states onto adjacent qubits, at which point we may couple them as desired. However, SWAP operations are expensive; they can increase circuit depths considerably, rapidly consuming the coherence time of the device, as well as being incredibly noisy. SWAP gates are therefore to be avoided at all costs for near-term quantum computation.

Circuit transpilation concerns the mapping of a quantum circuit onto a particular device layout. As discussed in the previous paragraph, if the entangling structure is not natively supported by the target hardware, it is the job of the transpiler to identify a way of routing qubits and performing SWAPs to arrive at a physical realization of the circuit in question, ultimately yielding a set of instructions for the quantum computer to carry out. An effective transpilation routine will attempt to minimize the depth of the resulting effective circuit by possibly reordering qubits and decomposing gate blocks in different ways. In Figure 1.6 we demonstrate a small example of transpiling a 4-qubit cycle of entangling gates onto a linear qubit layout, requiring two SWAP operations.

Moreover, a particular device will have a native gate set (as discussed in Section 1.2.1) that it must adhere to – the transpiler must also decompose the target gates, which may not appear in the quantum computer's gate set, in terms of operations it is able to perform. We provide an example of this in Figure 1.7, which decomposes the circuit of Figure 1.6 when only CNOT and Hadamard gates are available on the hardware.

As with many graph problems in mathematics, circuit transpilation is very challenging to implement in a close-to-optimal way. Algorithms such as the SWAP-based bidirectional heuristic search (SABRE) make an attempt to do so [34], although there is a probabilistic element to this technique that results in a possibly different circuit at each execution. Often, SABRE will be run multiple times and the shallowest depth will be selected, however this is not ideal.

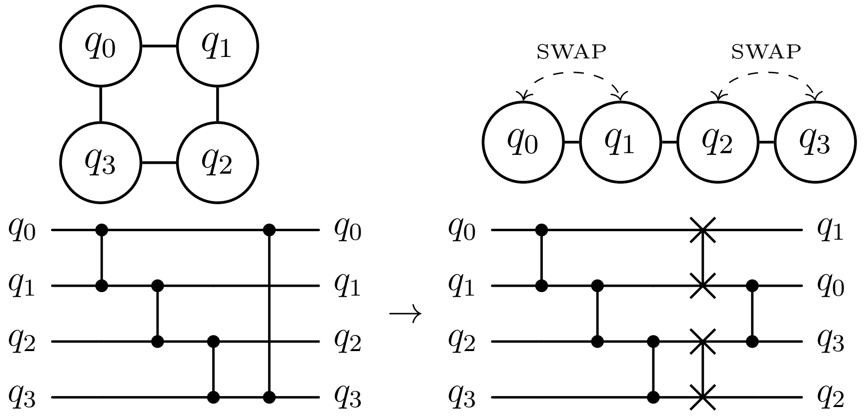


Figure 1.6: Transpilation of 4-qubit circular entanglement in a linear topology. Note the ordering of qubits has changed at the end of transpilation.

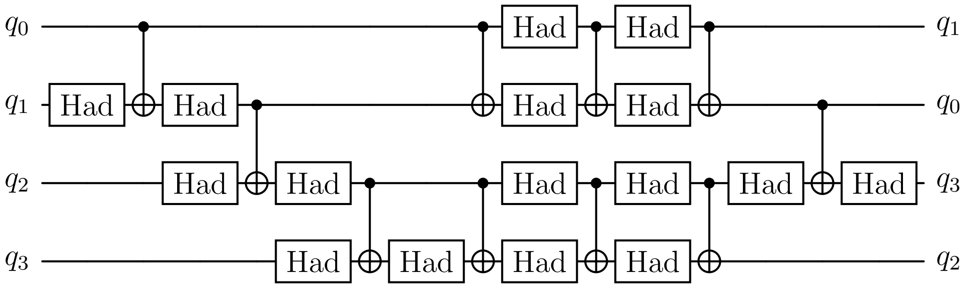


Figure 1.7: Transpilation of the circuit in Figure 1.6 when CZ and SWAP gates are not available, instead decomposing into CNOT and Hadamard gates.

1.3 Quantum Chemistry Preliminaries

Quantum computers are expected to be transformative in many domains of science, offering computational speed-up over conventional algorithms to facilitate the elucidation of physical systems that are otherwise classically intractable. Quantum chemistry has been investigated as an application of quantum computing for almost two decades [35]. Instances of the electronic structure problem are intrinsically quantum mechanical, are challenging for classical methods at the few-hundred qubit scale and are of scientific and commercial importance. Considerable development of quantum algorithms for quantum chemistry has taken place.

Progress in hardware has lead to many small demonstrations of quantum chemical calculations on noisy intermediate-scale quantum (NISQ) devices. These demonstrations evaluate both the practicality of quantum algorithms when deployed on real hardware and also evaluate NISQ hardware against a real-world application benchmark. Variational quantum algorithms have been studied extensively for their shallow circuits, making them appealing for NISQ applications where modest coherence times limit the depth of circuits that can be executed successfully. The current state-of-the-art, summarized by a representative sample of Variational Quantum Eigensolver realizations in Table 1.5 (which is the subject of the following Section 1.4), show that much progress is required before quantum computers can challenge their classical counterparts.

The goal of quantum computing for quantum chemistry is to achieve quantum advantage. This implies the existence of problems where all classical heuristics fail to produce adequate results, while a quantum algorithm succeeds in rendering its chemical features to sufficiently high accuracy to be scientifically useful. This means larger systems and/or basis sets [3] than are classically tractable, although candidate advantage applications are not easy to find, as recently brought to attention [36, 37] . The threshold of "chemical accuracy" is used as a typical indication of success which, in NISQ demonstrations, is taken to mean the quantum calculation achieves an absolute error below 1 kcal/mol (4 kJ/mol, 1.6 mHa, or 43 meV) with respect to the numerically-exact result in the chosen basis. This is an abuse of terminology as small basis set calculations will not typically be chemically-accurate when compared against experiment, which is the ultimate arbiter of computational utility; it has been suggested that *algorithmic* accuracy is more appropriate terminology [38]. The point of interest is whether NISQ devices are able to achieve sufficient resolution such that, with realistic basis sets, we may address chemically-relevant questions and many demonstrations fail to meet this standard.

The NISQ hardware available today limits demonstrations to small basis set calculations on small molecules that do not challenge the classical state-of-the-art. Continued hardware development should enable a sequence of demonstrations of increasing size whose endpoint is quantum advantage. Quantum simulations that address small-scale versions of problems whose large-scale realization is believed to be challenging or out-of-reach of classical chemistry methods are therefore good targets. In the previous Sections 1.1 and 1.2 we introduced the general toolkit for working with qubit systems; the goal now is to situate quantum chemistry within this framework.

1.3.1 The Molecular Hamiltonian

Generally speaking, a Hamiltonian describes the total energy of a physical system and is therefore comprised of terms relating to either kinetic or potential energy:

$$H = T + V. \tag{1.36}$$

The explicit form of these operators is dependent upon the system they model; in quantum chemistry, we are interested in a many-body system describing the distribution of electrons around atomic centres. Specifically, we consider a system of M nuclei with masses m_i and positions $\boldsymbol{R}_i \in \mathbb{R}^3$, $i \in \{1, \dots, M\}$, and N electrons with their positions labelled $\boldsymbol{r}_i \in \mathbb{R}^3$, $i \in \{1, \dots, N\}$.

The kinetic part splits into nuclear and electronic contributions, namely

$$T = T_{\text{nuc}} + T_{\text{elec}} \tag{1.37}$$

where

$$T_{\text{nuc}} = -\sum_{i=1}^{M} \frac{\hbar^2}{2m_i} \Delta_{\boldsymbol{R}_i},\ T_{\text{elec}} = -\sum_{i=1}^{N} \frac{\hbar^2}{2m_e} \Delta_{\boldsymbol{r}_i} \tag{1.38}$$

---

[3] Basis set meaning the atomic orbitals chosen to represent a molecular system. The so-called "minimal" basis set STO-3G (Slater-type orbital, 3 Gaussian functions) is used throughout this work, however larger sets such as the Dunning correlation-consistent bases are necessary to compare results with true chemical experiment.

with $m_e = 9.10938356 \times 10^{-31}$ kg the mass of an electron and $\hbar := \frac{h}{2\pi}$ the reduced form of Planck's constant $h = 6.62607004 \times 10^{-34}$ Js. Here, $\Delta := \nabla^2$ is the Laplace operator.

We may break down the potential term into its constituent parts by considering the possible interactions that may occur: nucleus-nucleus repulsion, nucleus-electron attraction and electron-electron repulsion, leading to the decomposition

$$V = V_{\text{nuc}-\text{nuc}} + V_{\text{nuc}-\text{elec}} + V_{\text{elec}-\text{elec}}. \tag{1.39}$$

In order to write out these terms explicitly we recall that Coulomb's law stipulates the electrostatic potential energy between two point-charges q_1, q_2 with positions $\boldsymbol{r}_i, \boldsymbol{r}_j$ is

$$V_{ij} = \frac{q_i q_j}{4\pi\epsilon_0 |\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{1.40}$$

where $\epsilon_0 = 8.85418782 \times 10^{-12}$ Fm^{-1} is the vacuum permitivity. The charge of a proton is $e = 1.60217662 \times 10^{-19}$ C and $-e$ for an electron, so the charge of an entire nucleus i containing Z_i protons is $Z_i e$. The repulsion terms are therefore given as

$$V_{\text{nuc}-\text{nuc}} = \sum_{i<j}^{M} \frac{Z_i Z_j e^2}{4\pi\epsilon_0 \, |\boldsymbol{R}_i - \boldsymbol{R}_j|}, \quad V_{\text{elec}-\text{elec}} = \sum_{i<j}^{N} \frac{e^2}{4\pi\epsilon_0 |\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{1.41}$$

and the nucleus-electron attraction is

$$V_{\text{nuc}-\text{elec}} = -\sum_{i=1}^{N}\sum_{j=1}^{M} \frac{Z_j e^2}{4\pi\epsilon_0 |\boldsymbol{r}_i - \boldsymbol{R}_j|}. \tag{1.42}$$

By making a change of units these terms can be simplified significantly; in *atomic units* we set $\hbar = e = m_e = 4\pi\epsilon_0 = 1$, so that

$$H = -\sum_{i=1}^{M} \frac{\Delta_{\boldsymbol{R}_i}}{2m_i} - \sum_{i=1}^{N} \frac{\Delta_{\boldsymbol{r}_i}}{2} + \sum_{i<j}^{M} \frac{Z_i Z_j}{|\boldsymbol{R}_i - \boldsymbol{R}_j|} - \sum_{i=1}^{M}\sum_{j=1}^{N} \frac{Z_i}{|\boldsymbol{R}_i - \boldsymbol{r}_j|} + \sum_{i<j}^{N} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}. \tag{1.43}$$

Currently, our molecular Hamiltonian encodes a huge amount of complexity. An effective simplification is the *Born-Oppenheimer approximation*, which exploits the mass between the nuclei and electrons being separated by several orders of magnitude. Relative to the electrons, the nuclei are assumed to be fixed in space so that $T_{\text{nuc}} = 0$ and $V_{\text{nuc}-\text{nuc}}$ is constant. Superficially, the $V_{\text{nuc}-\text{elec}}$ term appears many-body in nature; however, with all nuclei fixed the electrons simply experience a static external potential v which is an aggregate of all the nuclei. In atomic units this is given by

$$v(\boldsymbol{r}_i) = -\sum_{j=1}^{M} \frac{Z_j}{|\boldsymbol{r}_i - \boldsymbol{R}_j|}. \tag{1.44}$$

The molecular Hamiltonian under the Born-Oppenheimer approximation therefore takes the form

$$H = T_{\text{elec}} + V_{\text{elec}-\text{elec}} + \sum_{i=1}^{N} v(\boldsymbol{r}_i). \tag{1.45}$$

where $V_{\text{nuc}-\text{nuc}}$ has been omitted since it represents a constant shift in energy that can be added back in at the end of computation.

1.3.2 Fock Space and Second Quantization

The molecular Hamiltonian as presently formulated in Section 1.3.1 is expressed in the so-called *first quantization*. However, in quantum chemistry we deal with collections of identical particles and so it is nonsensical to ask where a particular particle resides, as it is not possible to distinguish between them. Since the first quantized form makes reference to individual electrons, it encodes redundant information; it is, in a sense, an over-complete representation of the electronic structure problem.

Rather than thinking about the behaviour of each particle as treated individually, a more sensible question is *how many* particles assume a certain state, or *mode*, $|\cdot\rangle$. This leads us to a different description of many-body systems, aptly named *second quantization*. It was developed to rectify the redundancy found in first quantization when treating systems of indistinguishable particles, which moreover requires a complicated (anti)symmetrization procedure that is implicitly baked into the second quantized theory.

The appropriate state-space is spanned by the occupation number basis where $|n_\ell\rangle_\ell$ indicates that $n_\ell \in \mathbb{N}$ particles occupy mode ℓ. To manipulate the occupation states, we introduce the annihilation and creation operators a_ℓ and $a_\ell^\dagger$, respectively, whose action is to either insert or remove a particle into mode ℓ. Specifically,

$$\begin{aligned} a_\ell |n_\ell\rangle_\ell &= \sqrt{n_\ell}\, |n_\ell - 1\rangle_\ell \,, \\ a_\ell^\dagger |n_\ell\rangle_\ell &= \sqrt{n_\ell + 1}\, |n_\ell + 1\rangle_\ell \end{aligned} \tag{1.46}$$

up to a possible sign-flip arising from the commutation relations of these operators, depending on the type of many-body system in question. We may also define the number operator $N_\ell := a_\ell^\dagger a_\ell$ which counts the number of particles in mode ℓ:

$$N_\ell |n_\ell\rangle = a_\ell^\dagger a_\ell |n_\ell\rangle = \sqrt{n_\ell} a_\ell^\dagger |n_\ell - 1\rangle_\ell = n_\ell |n_\ell\rangle \,. \tag{1.47}$$

The total number of particles in the many-body state

$$|\boldsymbol{n}\rangle = \bigotimes_\ell |n_\ell\rangle_\ell = |n_0, n_1, n_2, \ldots\rangle \,, \tag{1.48}$$

otherwise known as a *Fock state*, may then be counted by the operator $N := \sum_\ell N_\ell$, so that $N|\boldsymbol{n}\rangle = \sum_\ell n_\ell |\boldsymbol{n}\rangle$. Mathematically speaking, the Fock space $\mathbb{F}$ is a direct sum of Hilbert spaces representing $0, 1, 2, 3, \ldots$ particle states etc. and may therefore describe systems consisting of any number of particles. All the number states may be obtained via application of creation operators to the so-called *vacuum* state $|\text{vac.}\rangle := |0, 0, 0, \ldots\rangle$ describing a system of zero particles. For example, $|1, 0, 1, 0, 1, 0\rangle = a_0^\dagger a_2^\dagger a_4^\dagger |\text{vac.}\rangle$ (being mindful that there might a sign flip if the creation operators are reordered).

For bosonic system such as photons, characterized by having integer spin, we may have an arbitrary number of particles occupying the same mode. By contrast, in the fermionic case – the setting of quantum chemistry, consisting of electrons with spin-$\frac{1}{2}$ – the Pauli exclusion principle applies and consequently no two particles may occupy the same mode. We are therefore restricted to $n_\ell \in \{0, 1\}$, i.e. each mode is either occupied or unoccupied. The difference between bosons and fermions does not stop here, however; they also manifest in the commutation relations of their respective creation and annihilation operators.

With $b_\ell, b_\ell^\dagger$ the bosonic annihilation and creation operators, we have

$$[b_\ell, b_k] = [b_\ell^\dagger, b_k^\dagger] = 0,\ [b_\ell, b_k^\dagger] = \delta_{\ell,k}. \tag{1.49}$$

Whereas, in the fermionic case with $f_\ell, f_\ell^\dagger$ the corresponding annihilation and creation operators, we instead have

$$\{f_\ell, f_k\} = \{f_\ell^\dagger, f_k^\dagger\} = 0,\ \{f_\ell, f_k^\dagger\} = \delta_{\ell,k}; \tag{1.50}$$

the difference is the anticommutator $\{A, B\} \coloneqq AB + BA$ found in place of the commutator $[A, B] \coloneqq AB - BA$.

Where we presented the generic action of these operators in Equation 1.46, it was noted that a sign had possibly been omitted, which is relevant to the fermionic case. Due to the antisymmetry found in the exchange of fermionic particles, the true action of the annihilation and creation operators in this case is

$$\begin{aligned} f_\ell \left|\boldsymbol{n}\right\rangle &= \begin{cases} s_\ell(\boldsymbol{n}) \left|\ldots, n_{\ell-1}, n_\ell - 1, n_{\ell+1}, \ldots\right\rangle, & n_\ell = 1 \\ 0, & n_\ell = 0 \end{cases}, \\ f_\ell^\dagger \left|\boldsymbol{n}\right\rangle &= \begin{cases} s_\ell(\boldsymbol{n}) \left|\ldots, n_{\ell-1}, n_\ell + 1, n_{\ell+1}, \ldots\right\rangle, & n_\ell = 0 \\ 0, & n_\ell = 1 \end{cases}, \end{aligned} \tag{1.51}$$

where $s_\ell(\boldsymbol{n}) \coloneqq (-1)^{\sum_{k=0}^{\ell-1} n_k}$ counts the occupation of all modes below ℓ. This accounts for the exchange of particles and enforces the antisymmetry requirement. In the bosonic case these operators commute and therefore no sign flip occurs.

Finally, we may rewrite the first-quantized molecular Hamiltonian given in Equation (1.45) as

$$H = \sum_{p,q} h_{p,q} f_p^\dagger f_q + \frac{1}{2} \sum_{p,q,r,s} h_{p,q,r,s} f_p^\dagger f_q^\dagger f_s f_r \tag{1.52}$$

where the transformed Hamiltonian coefficients are obtained by integrating [4] the various kinetic and potential energy terms over first-quantized molecular orbital basis functions $\phi_\ell(\boldsymbol{r})$ corresponding with the Fock mode ℓ. Explicitly, this yields

$$\begin{aligned} h_{p,q} &= \int \phi_p^*(\boldsymbol{r}) \Big[-\frac{1}{2}\Delta + v(\boldsymbol{r}) \Big] \phi_q(\boldsymbol{r}) d\boldsymbol{r} \\ h_{p,q,r,s} &= \int \phi_p^*(\boldsymbol{r}) \phi_q^*(\boldsymbol{r}') \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \phi_r(\boldsymbol{r}) \phi_s(\boldsymbol{r}') d\boldsymbol{r} d\boldsymbol{r}' \end{aligned} \tag{1.53}$$

where $v(\boldsymbol{r})$ was defined in Equation (1.44) and we have again adopted the atomic units [40].

---

[4] These integrals can be calculated efficiently over many choices of atomic orbital basis set via a plethora of available codes. In this work we adopt PySCF [39] as our quantum chemistry driver, which adopts the chemists' notation and thus we transform into the physicists' notation for our applications, transposing array axes as $0, 1, 2, 3 \rightarrow 0, 2, 3, 1$.

Example 1.3: Molecular Hydrogen Fermionic Hamiltonian

The molecular hydrogen (H_2) STO-3G Born-Oppenheimer Hamiltonian, consisting of two electrons in four spin-orbitals, is parametrized with ten coefficients:

1. The nuclear energy $h^{(\text{nuc})}$;
2. Three nucleus - electron attraction integrals $h_0^{(\text{ne})}, \ldots, h_2^{(\text{ne})}$;
3. Six electron - electron repulsion integrals $h_0^{(\text{ee})}, \ldots, h_5^{(\text{ee})}$.

These coefficients are a function of the spatial separation between each hydrogen atom. For example, at the equilibrium bond length they assume values (given in eV): $h^{(\text{nuc})} = 19.591$, $\boldsymbol{h}^{(\text{ne})} = (-17.093, 0.000, -6.420)$, $\boldsymbol{h}^{(\text{ee})} = (18.387, 0.000, 4.923, 18.084, 0.000, 19.009)$.
All in all, the fermionic Hamiltonian for H_2 assumes the form

$$\begin{aligned}
H = \quad & h^{(\text{nuc})} \\
& + h_0^{(\text{ne})} \cdot (f_0^\dagger f_0 + f_1^\dagger f_1) \\
& + h_1^{(\text{ne})} \cdot (f_0^\dagger f_2 - f_2^\dagger f_0 + f_1^\dagger f_3 - f_3^\dagger f_1) \\
& + h_2^{(\text{ne})} \cdot (f_2^\dagger f_2 + f_3^\dagger f_3) \\
& + h_0^{(\text{ee})}/2 \cdot (f_0^\dagger f_0^\dagger f_0 f_0 + f_1^\dagger f_1^\dagger f_1 f_1 + f_0^\dagger f_1^\dagger f_0 f_1 + \text{perms.}) \\
& + h_1^{(\text{ee})}/2 \cdot (f_0^\dagger f_0^\dagger f_0 f_2 + f_1^\dagger f_1^\dagger f_1 f_3 + f_0^\dagger f_1^\dagger f_0 f_3 + f_0^\dagger f_1^\dagger f_1 f_2 + \text{perms.}) \\
& + h_2^{(\text{ee})}/2 \cdot (f_0^\dagger f_2^\dagger f_0 f_2 + f_1^\dagger f_3^\dagger f_1 f_3 + f_0^\dagger f_3^\dagger f_1 f_2 + \text{perms.}) \\
& + h_3^{(\text{ee})}/2 \cdot (f_0^\dagger f_0^\dagger f_2 f_2 + f_1^\dagger f_1^\dagger f_3 f_3 + f_0^\dagger f_1^\dagger f_2 f_3 + \text{perms.}) \\
& + h_4^{(\text{ee})}/2 \cdot (f_0^\dagger f_2^\dagger f_2 f_2 + f_1^\dagger f_3^\dagger f_3 f_3 + f_0^\dagger f_3^\dagger f_2 f_3 + f_1^\dagger f_2^\dagger f_2 f_3 + \text{perms.}) \\
& + h_5^{(\text{ee})}/2 \cdot (f_2^\dagger f_2^\dagger f_2 f_2 + f_3^\dagger f_3^\dagger f_3 f_3 + f_2^\dagger f_3^\dagger f_2 f_3 + \text{perms.}),
\end{aligned} \tag{1.54}$$

with the Hartree-Fock state $|1,1,0,0\rangle = f_0^\dagger f_1^\dagger |\text{vac.}\rangle$. We note that “perms.” above accounts for the 8-fold symmetry present, since permuting the i, j, k, l indices $i \leftrightarrow j, k \leftrightarrow l, (i,j) \leftrightarrow (k,l)$ leaves the integrals invariant. Therefore, these three actions generate the $2^3 = 8$ allowed permutations, e.g. for $i, j, k, l = 0, 1, 2, 3$ we have

$$\begin{gathered}
(0,1,2,3), (0,1,3,2), (1,0,2,3), (1,0,3,2), \\
(2,3,0,1), (3,2,0,1), (2,3,1,0), (3,2,1,0)
\end{gathered}$$

all with the same integral value. Exploiting this fact can simplify the molecular Hamiltonian construction greatly.
One can find further simplification by considering the molecular point group, describing geometrical symmetry. For example, it is no coincidence that in the equilibrium example above we find $h_1^{(\text{ee})} = h_4^{(\text{ee})} = 0$; in fact, this will be true for all bond lengths and the reason comes down to the preservation of symmetry. In Chapter 2 we revisit the idea of simplifying problems through symmetry arguments.

Note the ordering of the annihilation/creation indices in Equation (1.52), where we find p, q, s, r rather than p, q, r, s. This is the physicists' notation, but there is an alternative chemists' notation which results in differing index ordering, namely the permutation $p, q, r, s \to p, r, s, q$. Moreover, the form of the two-body integral in Equation (1.53) is different in the chemists' convention, where the electron positions are grouped on either of the Coulomb operator, rather than being split across as in physicists' [41, p. 67, 95].

The second-quantized formulation of the electronic structure Hamiltonian is appealing as the fermionic antisymmetrization is baked into the Hamiltonian itself, rather than needing to be asserted explicitly in the first-quantized form.

1.3.3 Hartree-Fock and Wavefunction Methods

Having now defined the problem that we are interested in, the goal of quantum chemistry is to solve the Schrödinger equation

$$H |\psi\rangle = \epsilon |\psi\rangle \tag{1.55}$$

where H is the electronic structure Hamiltonian of Equation (1.52). While seemingly inconspicuous, exact solutions are possible only in the simplest of cases. Instead, techniques for computing approximate solutions have been developed and is a core pursuit of quantum computation as well. In this section, we introduce some of the main algorithmic tools of conventional quantum chemistry that shall serve as a benchmark against which the quantum computational techniques developed throughout this thesis will be assessed. In Table 1.4 we list several such methods, including their asymptotic scaling in terms of the number of orbitals, and in Figure 1.8 we compare their performance when stretching the triple bond of molecular nitrogen (N_2). This is a challenging system to describe correctly and in this case highlights the deficiencies in the wavefunction methods that we describe in the remainder of this section. First, we introduce Hartree-Fock theory, from which all proceeding methods follow.

Wavefunction method	Scaling	Variational?
RHF/ROHF/UHF	$\mathcal{O}(M^4)$	Yes
MP2	$\mathcal{O}(M^5)$	No
CCSD	$\mathcal{O}(M^6)$	No
CCSD(T)	$\mathcal{O}(M^7)$	No
CISD	$\mathcal{O}(M^6)$	Yes
FCI	$\mathcal{O}(\exp(M))$	Yes
CASCI($M_{\text{act}}, N_{\text{act}}$)	$\mathcal{O}(\exp(M_{\text{act}}))$	Yes
CASSCF($M_{\text{act}}, N_{\text{act}}$)	$\mathcal{O}(\exp(M_{\text{act}}))$	Yes

Table 1.4: A selection of wavefunction methods of varying complexity that we use to benchmark the quantum algorithms developed throughout this thesis. The number of orbitals is M and for the CAS methods ($M_{\text{act}}, N_{\text{act}}$) indicates an active space consisting of $N_{\text{act}} \leq N$ electrons correlated in $M_{\text{act}} \leq M$ orbitals.

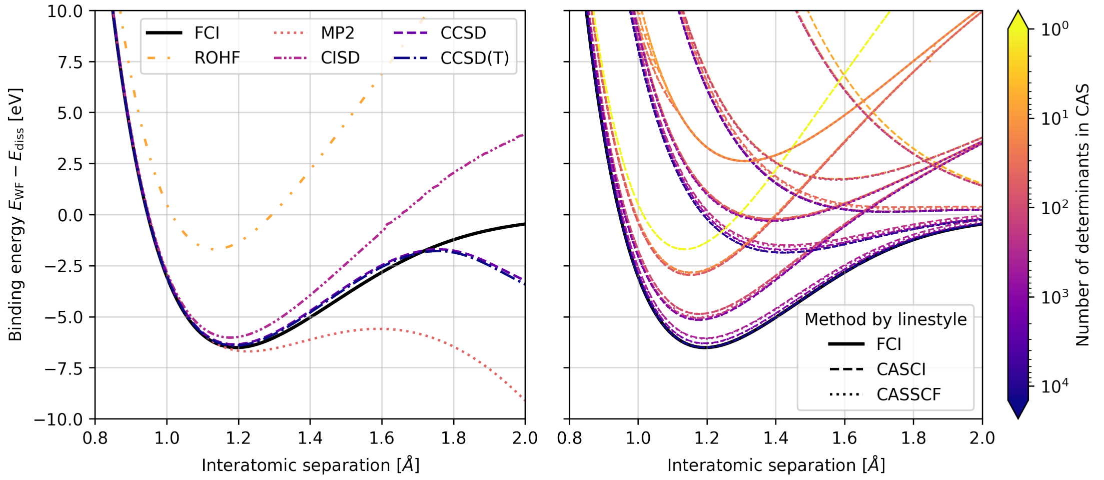


Figure 1.8: A comparison of the wavefunction methods listed in Table 1.4 for dissociation of molecular nitrogen (N_2), a common benchmark problem in electronic structure. We present the CASCI and CASSCF methods for every valid active space consisting of N_{act} electrons in M_{act} orbitals.

Hartree-Fock Theory

Hartree-Fock (HF) is the hallmark of modern electronic structure theory and is the foundation of many more advanced quantum chemistry methods. As mentioned at the end of Section 1.3.2, the molecular Hamiltonian of Equation 1.52 is defined over a set of molecular orbitals $\phi_\ell(\boldsymbol{r})$, with the coefficients $h_{p,q}, h_{p,q,r,s}$ of Equation 1.53 obtained via integration over the coordinates $\boldsymbol{r}$ parametrizing these orbitals. However, we did not discuss how the molecular orbitals are actually obtained from the *atomic* orbital basis set in the first place; this is the purpose of the Hartree-Fock method. In short, Hartree-Fock variationally optimizes the Hamiltonian energy with respect to a single reference state (otherwise known as a *Slater determinant*) via molecular orbital rotations $U(\boldsymbol{\kappa})$. We refer the reader elsewhere for details surrounding the form of this unitary [42]. The particular reference state chosen for Hartree-Fock is that with N electrons filling M orbitals from the lowest canonical molecular orbital energy upwards, which we may write as

$$|\mathrm{HF}\rangle := \prod_{n=0}^{N-1} a_n^\dagger |\mathrm{vac.}\rangle = |\underbrace{1,\dots,1}_{N},\underbrace{0,\dots,0}_{M-N}\rangle \,. \tag{1.56}$$

Note that this formulation assumes alternating spin up/down particles, while some use the convention of grouping all spin up occupancies on the left, and all spin down occupancies on the right, giving a different form of the Hartree-Fock state: $|\mathrm{HF}\rangle = |\underbrace{1,1,\dots,0}_{M/2},\underbrace{1,1,\dots,0}_{M/2}\rangle$. Therefore, the Hartree-Fock optimization problem may be written

$$\epsilon_{\mathrm{HF}} = \min_{\boldsymbol{\kappa}} \langle\mathrm{HF}|\, U(\boldsymbol{\kappa})^\dagger H U(\boldsymbol{\kappa})\, |\mathrm{HF}\rangle \,. \tag{1.57}$$

The resulting molecular orbitals obtained via the optimized orbital rotation $U(\boldsymbol{\kappa})$ then defines the basis in which we write down the molecular Hamiltonian in Equation 1.52. There are also several variants of Hartree-Fock that treat spin up/down electrons in different ways. The simplest, restricted Hartree-Fock (RHF), asserts that spins in the same spatial orbital be treated the same, with the open-shell extension (ROHF) allowing a mismatch of valence electrons. Then there is the unrestricted Hartree-Fock method (UHF) that treats the spin up/down electrons as totally separate; this can produce improved results over RHF, although is prone to symmetry breaking [43].

It should be noted that the Hartree-Fock optimization problem is non-linear and is typically solved using iterative methods with direct inversion [44–47]. The problem is in fact NP-complete [42, 48], although the average-case complexity means Hartree-Fock is broadly applicable across many problems of interest in quantum chemistry. In evaluating $\langle \mathrm{HF}| U(\boldsymbol{\kappa})^\dagger H U(\boldsymbol{\kappa}) |\mathrm{HF}\rangle$, the full two-body electronic structure Hamiltonian contains redundant information. Instead, it is possible to write down a one-body effective Hamiltonian known as the Fock operator. Since Hartree-Fock optimizes for just a single Slater determinant it is an inherently uncorrelated theory. The total number of determinants given N electrons in M orbitals is $\binom{M}{N}$, which grows at worst exponentially with the system size. The goal of "post Hartree-Fock" methods is to then incorporate electron correlations into the molecular wavefunction expansion by various means, which we discuss next.

Perturbation Theory

Perturbation theory provides an appealing approach to problems of many-body physics as it allows one to model a complex system with Hamiltonian H in terms of something simpler, say H_0. The missing contributions of the full system are collected into a so-called perturbation term $V = H - H_0$. The core assumption of the theory is that we may access eigenstates of our "model" Hamiltonian H_0 efficiently, with a view to use them as a basis in which to expand the desired eigenstates of H. It is intuitively true that the closer H_0 is to H – or, conversely, the smaller V is – the better our perturbative treatment of the problem will be. In quantum chemistry, one typically selects the fermionic Fock operator, whose eigenvectors are the occupation number states $|\boldsymbol{n}\rangle = |n_1, n_2, \ldots\rangle$ where $n_i \in \{0, 1\}$ and $\sum_i n_i$ is the total number of electrons. Through subsequent application of the Rayleigh–Schrödinger formulation of perturbation theory we obtain the Møller–Plesset (MPK) theory, where $K \in \{2, 3, 4, \ldots\}$ denotes the expansion order.

The true solution to the Schrödinger equation $H|\psi_0\rangle = \epsilon_0 |\psi_0\rangle$ is expanded as a Maclaurin series

$$\begin{aligned} \epsilon_0 &= \sum_\ell \epsilon_0^{(\ell)} = \epsilon_{\mathrm{HF}} + \sum_{\ell>0} \epsilon_0^{(\ell)}, \\ |\psi_0\rangle &= \sum_\ell |\psi_0^{(\ell)}\rangle = |\mathrm{HF}\rangle + \sum_{\ell>0} |\psi_0^{(\ell)}\rangle \end{aligned} \tag{1.58}$$

where the zeroth-order is the Hartree-Fock solution $\epsilon_0^{(0)} = \epsilon_{\mathrm{HF}}, |\psi_0^{(0)}\rangle = |\mathrm{HF}\rangle$ and higher-order contributions $\epsilon_0^{(\ell)}, |\psi_0^{(\ell)}\rangle$ are calculated in terms of the known eigenvalues/states of the Fock operator H_0. Truncating the expansion order at some $K \in \mathbb{N}$ means evaluating the energy corrections $\epsilon_0^{(1)}, \ldots, \epsilon_0^{(K)}$. The most commonly used perturbative method in quantum chemistry is the second-order

Møller–Plesset theory, where $\epsilon_{\text{MP2}} = \epsilon_{\text{HF}} + \epsilon_0^{(2)}$ with the energy correction given explicitly as

$$\begin{aligned} \epsilon_0^{(1)} &= \langle HF| V |HF\rangle \quad (= 0 \ \text{ here}) \\ \epsilon_0^{(2)} &= - \sum_{\boldsymbol{n} \neq |HF\rangle} \frac{|\langle \boldsymbol{n}| V |HF\rangle|^2}{\epsilon_{\boldsymbol{n}} - \epsilon_{\text{HF}}}. \end{aligned} \tag{1.59}$$

It is important to note the perturbed wavefunction is not normalized in the strict sense, but rather it is *intermediately normalized* since $\langle HF|\psi_0\rangle = 1$. Furthermore, perturbation theory is susceptible to non-variationality and breaks down if the perturbation becomes large in relation to the model system.

Coupled Cluster Theory

In Coupled Cluster theory, we expand the wavefunction with respect to the Hartree-Fock reference as $|\psi\rangle = e^T |\text{HF}\rangle$ where T is referred to as the *cluster operator*. This operator can be broken down into a sum over terms describing single excitations $a_p^\dagger a_q$ (T_1), double excitations $a_p^\dagger a_q^\dagger a_r a_s$ (T_2) and so on, such that $T = \sum_i T_i = T_1 + T_2 + \dots$ with

$$T_1 = \sum_{p,q} t_{p,q} a_p^\dagger a_q, \quad T_2 = \frac{1}{4} \sum_{p,q,r,s} t_{p,q,r,s} a_p^\dagger a_q^\dagger a_r a_s, \quad \text{etc.} \tag{1.60}$$

for cluster amplitudes $t_{p,q}, t_{p,q,r,s}, \dots$. Computing the exponential e^T is intractable in the general case and is moreover non-unitary. Even when truncating the excitations to singles and doubles, the expansion

$$e^{T_1+T_2} = \sum_{k=0}^{\infty} \frac{(T_1+T_2)^k}{k!} = I + T_1 + T_2 + \frac{1}{2}\Big(T_1^2 + T_2^2 + \{T_1, T_2\}\Big) + \dots \tag{1.61}$$

is unwieldy to work with. Instead, coupled cluster solutions are better approached via the similarity transformation $H \mapsto e^{-T} H e^T$ and expanded with the BCH formula introduced in Section 1.2.4. Since the electronic structure Hamiltonian contains no more than two-body terms, the expansion terminates at 4^{th}-order [49, 50]. The Coupled Cluster energy is then obtained as $\epsilon_{\text{CC}} = \langle \text{HF}| e^{-T} H e^T |\text{HF}\rangle$ and if we truncate the cluster operator T at some excitation order we get an approximate ground state energy $\epsilon_{\text{CCSDTQ}\dots}$ where the suffix SDTQ... indicates single, double, triple, quadruple excitations etc.

Unfortunately, calculations in excess of double excitations are typically too computationally demanding and so CCSD is usually as far as we can go via the standard cluster treatment for systems of any respectable size. However, it is possible to add a perturbative triples correction to CCSD, resulting in the CCSD(T) method that is widely considered to be the “gold standard” of quantum chemistry. Like perturbation theory above, the coupled cluster approximation is non-variational, although a unitary formulation of Coupled Cluster theory has been investigated extensively for quantum computing and reintroduces variationality into the method [22, 51–53]. Rather than exponentiating the cluster operator T as in standard Coupled Cluster theory, Unitary Coupled Cluster (UCC) exploits the fact that $T - T^\dagger$ is antihermitian and therefore $e^{T-T^\dagger}$ is a unitary operator. This is revisited in Section 1.4.3 when we discuss chemically-motivated quantum circuit design.

Configuration Interaction Theory

First of all, we shall define by $\mathcal{N}_{M,N}$ the set of occupation (Fock) states $|\boldsymbol{n}\rangle$ defined over M modes and with N particles; practically, $\boldsymbol{n}$ is a binary bit string with Hamming weight N (i.e. the number of nonzero elements). Now, the true ground state of the molecular Hamiltonian in Equation (1.52) will be of the form $|\psi_0\rangle = \sum_{\boldsymbol{n}\in\mathcal{N}_{M,N}} \gamma_{\boldsymbol{n}} |\boldsymbol{n}\rangle$. This state may be identified exactly by directly diagonalizing the Hamiltonian H in the desired particle number sector, an approach called *Full Configuration Interaction* (FCI) as it correlates every possible electron configuration amongst the molecular orbitals. To ensure the diagonalization is carried out in the correct sector, the Hamiltonian is projected into the desired particle space via the projection operator $\mathbb{P}_{M,N} = \sum_{\boldsymbol{n}\in\mathcal{N}_{M,N}} |\boldsymbol{n}\rangle\langle\boldsymbol{n}|$. While diagonalizing the projected Hamiltonian $\mathbb{P}_{M,N} H \mathbb{P}_{M,N} |\psi_{\mathrm{FCI}}\rangle = \epsilon_{\mathrm{FCI}} \mathbb{P}_{M,N} |\psi_{\mathrm{FCI}}\rangle$ is easier than solving the full problem (i.e. over every particle sector), it still requires exponential classical resource. This is because, combinatorially, $|\mathcal{N}_{M,N}| = \binom{M}{N}$, which in general scales exponentially with increasing numbers of electrons and orbitals [5]. To quantify the degree of electron correlation present for a particular problem, one can define the correlation energy $\epsilon_{\mathrm{corr}} = \epsilon_{\mathrm{FCI}} - \epsilon_{\mathrm{HF}}$, which captures the deficiency of Hartree-Fock in describing correlation effects.

However, by restricting ourselves to a subset of determinants $\mathcal{N}'_{M,N} \subset \mathcal{N}_{M,N}$ and forming the projection operator $\mathbb{P}'_{M,N} = \sum_{\boldsymbol{n}\in\mathcal{N}'_{M,N}} |\boldsymbol{n}\rangle\langle\boldsymbol{n}|$, we may project H into a reduced subspace in which diagonalization is feasible with polynomially scaling resources. A common choice is to select the electron configurations that may be reached from either single or double excitations of the Hartree-Fock reference, which we shall denote $\mathcal{N}^{(\mathrm{SD})}_{M,N}$ and has size $|\mathcal{N}^{(\mathrm{SD})}_{M,N}| = \binom{N}{2}\binom{M-N}{2} + N(M-N) + 1$. Diagonalizing the Hamiltonian in the space of single and double excitations yields $\mathbb{P}^{(\mathrm{SD})}_{M,N} H \mathbb{P}^{(\mathrm{SD})}_{M,N} |\psi_{\mathrm{CISD}}\rangle = \epsilon_{\mathrm{CISD}} |\psi_{\mathrm{CISD}}\rangle$, where ϵ_{CISD} is referred to as the Configuration Interaction Singles Doubles (CISD) energy. Another example are Complete Active Space (CAS) methods, in which we consider every determinant within a subset $(M_{\mathrm{act}}, N_{\mathrm{act}})$, indicating an active space consisting of $N_{\mathrm{act}} \leq N$ electrons correlated in $M_{\mathrm{act}} \leq M$ orbitals.

Some of the most cutting-edge methods of quantum chemistry look to choose the configuration space in a better-informed way. For example, one can treat the full system perturbatively to identify determinants that are in some sense "important" and then diagonalize the problem in a subspace defined by the dominant contributions in the perturbed wavefunction expansion; this is the approach of Configuration Interaction using a Perturbative Selection made Iteratively (CIPSI) [54, 55]. This falls under a class of so-called Selected Configuration Interaction (SCI) methods, which are differentiated by how one assigns importance to each configuration candidate.

An SCI technique that has seen recent success is Heatbath Configuration Interaction (HCI) [56–58]. This works by solving the Schrödinger equation in the current configuration space (typically initiated with the Hartree-Fock reference), yielding eigenstate $|\psi\rangle$, before contracting the Hamiltonian onto this state, $H|\psi\rangle$. We then select the configurations therein whose coefficients exceed some thresh-

---

[5] Some sources state the FCI problem as having factorial scaling; that is, *superexponential*. However, note that summing over every particle sector we get $\sum_{N=0}^{M} \binom{M}{N} = 2^M$. In the worst case, where $N = M/2$, we obtain a scaling of $\mathcal{O}(2^{M-1})$.

old δ to append to a growing configuration space. This process is iterated until convergence, followed by a perturbation into the full space from the variationally optimized HCI state. As $\delta \to 0$ HCI approaches the FCI limit since all configurations will be included.

A relatively new development in quantum computation is to use a quantum device to sample configurations for inclusion in the classical diagonalization step, a technique referred to as Quantum-Selected Configuration Interaction (QSCI) [59, 60]. The motivation behind this is that state-sampling is classically hard [61–65]; therefore, any possibility of quantum advantage from QSCI will necessarily be virtue of the specific ansatz choice. Recent work from IBM utilized QSCI to perform a 58-qubit simulation of the dissociation of molecular nitrogen (N_2), where configurations were sampled on a 133-qubit IBM Quantum *Heron* processor and subsequently diagonalized across several thousand nodes of the Fugaku supercomputer [66]. A series of works shortly followed suit [67–71] and marks a shift towards a “Quantum-Centric” approach to supercomputing, as detailed in a recent perspective on the topic in the context of materials science [72].

1.3.4 Fermion to Qubit Transformations

In Section 1.1 we introduced the concept of a Hilbert space $\mathscr{H}$, which encompasses all the states our quantum system might assume. The computational space of N qubits is $\mathscr{H} = (\mathbb{C}^2)^{\otimes N}$, which every finite dimensional vector space is isomorphic to. Therefore, there will exist mappings $q : \mathbb{F} \to (\mathbb{C}^2)^{\otimes N}$ from the fermionic Fock modes, i.e. the occupation number states $|\boldsymbol{n}\rangle$, onto qubit states $|\boldsymbol{b}\rangle$. Moreover, such a mapping should preserve the fermionic annihilation and creation operator anticommutation relations, which we wrote down in Equation (1.50).

Contrast this with the anticommutation relation of the pauli operators (excluding the identity) $\sigma_1 = X, \sigma_2 = Y, \sigma_3 = Z$, we have $\{\sigma_\ell, \sigma_k\} = 2\delta_{\ell,k} I$. Whereas, for a qubit mapping q to respect the fermionic relations, we need

$$\{q(f_\ell), q(f_k)\} = \{q(f_\ell)^\dagger, q(f_k)^\dagger\} = 0 \quad \forall\, \ell \neq k, \;\; \text{and} \;\; \{q(f_\ell), q(f_k)^\dagger\} = \delta_{\ell,k}, \tag{1.62}$$

which the Paulis do not satisfy on their own.

First of all, in order for a qubit mapping to satisfy same-site anticommutativity, we may assume that $q(f)^2 = 0$, since this is the only value x satisfying $x = -x$. This enforces the same behaviour seen for fermionic operators, whereby creation in an already occupied mode or annihilation of an empty one causes the state to vanish. First, we shall consider the possible forms of $q(f)$ on a single qubit position. Decomposing into the Pauli basis we may write $q(f) = wI + xX + yY + zZ$. The off-diagonal elements of $q(f)^2$ yield $2w(x - iy) = 2w(x + iy) = 0$. Since we will most certainly need X and/or Y operators to represent creation/annihilation of fermions, we may assume $w = 0$. Turning now to the diagonal entries of $q(f)^2$ and having asserted $w = 0$, we find the requirement $x^2 + y^2 + z^2 = 0$. We may rewrite this as $(x + iy)(x - iy) = -z^2$ and observe that, letting $\gamma := x + iy$, we must have $x - iy = \frac{-z^2}{\gamma}$. Therefore, the condition $q(f)^2 = 0$ is satisfied for any map q whose action on a single site has the form $q(f) = xX + yY + zZ$ where

$$x = \frac{\gamma^2 - z^2}{2\gamma}, y = \frac{\gamma^2 + z^2}{2i\gamma}, \;\; \text{for any} \;\; z \in \mathbb{C}, \gamma \in \mathbb{C} \setminus \{0\}. \tag{1.63}$$

We may constrain the above requirement further by looking also at the second anticommutation relation of Equation (1.62) for same-site spins, i.e. $\{q(f), q(f)^\dagger\} = 1$. To enforce this, we additionally need $|x|^2 + |y|^2 + |z|^2 = \frac{1}{2}$. The most convenient choice here is to take $\gamma = 1, z = 0$, yielding $q(f) = (X + iY)/2$. This expression adheres to all the same-site anticommutation relations, however one very quickly realizes that it currently commutes off-site, i.e. $[q(f_\ell), q(f_k)] = 0$ for $\ell \neq k$. The task now is to ensure the correct anticommutation relations for differing modes, which can be achieved in several ways. Ultimately, this comes down to a decision over how to store the parity information coming from the different modes. The simplest approach is to append strings of trailing Pauli Z operators to the form above; this results in the ubiquitous Jordan-Wigner transformation [73]

$$q(f_\ell) = \frac{X_\ell + iY_\ell}{2} \otimes Z_{\ell+1} \otimes Z_{\ell+2} \otimes \cdots \otimes Z_{N-1}. \tag{1.64}$$

This is by no means the only way that the parity information could be represented. For example, the Bravyi-Kitaev map [74, 75] stores the parity in a delocalized manner, resulting conversely in more localized observables than Jordan-Wigner [76], since the latter can result in full-weight Pauli strings. There are also additional approaches based on parity trees [77]. Despite all these various approaches to qubit mapping, it should be noted that they act on the single mode in question in the same way, i.e. via $f \mapsto (X + iY)/2$; it is the surrounding qubits and how they encode parity information that they differ. Note that the formulation laid out above implicitly assumes that a single fermionic mode maps onto a single qubit – if this condition is relaxed we are opened up to a wealth of qubit mapping possibilities, for example compact schema [78, 79] or those that encode the conservation of particle number [80].

Example 1.4: Jordan-Wigner Transformation

To encode a single-excitation $f_2^\dagger f_5$ under the Jordan-Wigner transformation, we directly apply Equation (1.64) to obtain

$$\begin{aligned} 2q(f_2^\dagger) &= IIXZZZZZ\cdots + i \cdot IIYZZZZZ\ldots \\ 2q(f_5) &= IIIIIXZZ\cdots - i \cdot IIIIIYZZ\ldots \end{aligned} \tag{1.65}$$

and therefore

$$\begin{aligned} q(f_2^\dagger f_5) = a(f_2^\dagger)q(f_5) = \frac{1}{4}\Big[\\ i \cdot IIXZZYII - i \cdot IIYZZXII - IIXZZXII - IIYZZYII\Big]. \end{aligned} \tag{1.66}$$

We see that, in between the X, Y positions which (de)excite an electron from one mode to another, there is a string of Pauli Z operations which computes the parity of the intermediate modes.

Example 1.5: Molecular Hydrogen Jordan-Wigner Hamiltonian

In Example 1.3 we wrote down the form of the Fermionic Hamiltonian for molecular hydrogen. We now continue the example by demonstrating how we go from the fermionic form to a qubit Hamiltonian via the Jordan-Wigner transformation.

$$\begin{aligned}
4q_{\mathrm{JW}}(H) = {} & (4h^{(\mathrm{nuc})} + 8h_0^{(\mathrm{ne})} + 8h_2^{(\mathrm{ne})} + h_0^{(\mathrm{ee})} - 2h_2^{(\mathrm{ee})} + 4h_3^{(\mathrm{ee})} + h_5^{(\mathrm{ee})}) \cdot IIII \\
& - (4h_0^{(\mathrm{ne})} + h_0^{(\mathrm{ee})} - h_2^{(\mathrm{ee})} + 2h_3^{(\mathrm{ee})}) \cdot (ZIII + IZII) \\
& - (4h_2^{(\mathrm{ne})} - h_2^{(\mathrm{ee})} + 2h_3^{(\mathrm{ee})} + h_5^{(\mathrm{ee})}) \cdot (IIZI + IIIZ) \\
& + h_0^{(\mathrm{ee})} \cdot ZZII \\
& + h_5^{(\mathrm{ee})} \cdot IIZZ \\
& + h_3^{(\mathrm{ee})} \cdot (ZIIZ + IZZI) \\
& + (h_3^{(\mathrm{ee})} - h_2^{(\mathrm{ee})}) \cdot (ZIZI + IZIZ) \\
& + h_2^{(\mathrm{ee})} \cdot (YXXY + XYYX - XXYY - YYXX)
\end{aligned} \tag{1.67}$$

1.4 The Variational Quantum Eigensolver

The algorithms that can be successfully run on quantum hardware today are few and far between. The long-term vision for quantum computing sees algorithms such as Quantum Phase Estimation (QPE) doing the heavy-lifting. While the measurement overhead for QPE is constant, i.e. $\mathcal{O}(1)$, by virtue of interference effects provided by the Quantum Fourier Transform, the circuits end up being very deep. This necessitates extended periods of coherence in order to complete each circuit run before all the quantum information leaks into the environment. In the near-term, this is simply not possible at scale – the few experimental realizations of QPE succeeded only in simulating very small systems such as a 2-qubit example of H_2 [81]. Instead, noisy intermediate-scale quantum (NISQ) research has looked toward algorithms requiring more modest coherent resource. The most extensively studied of these is the Variational Quantum Eigensolver (VQE) [82], which offers short circuit depths at the expense of increased measurement cost $\mathcal{O}(\epsilon^{-2})$ for a target error $\epsilon > 0$. In Table 1.5 we provide a selection of significant VQE experiments conducted in the chemical domain over the decade 2013 – 2023.

Conceptually, VQE is quite straightforward and is grounded in the variational principle: given a Hamiltonian H with ground state energy ϵ_0, all states $|\psi\rangle$ will satisfy $\langle\psi| H |\psi\rangle / \langle\psi|\psi\rangle \geq \epsilon_0$. In other words, the ground state energy bounds all expectation values from below and therefore we can never go below the target energy. This means we can be confident that, as long as the energy is decreasing, our solution will be getting better without fear of overshooting the target. There are also methods for targeting excited states with VQE [113, 114].

While it is somewhat intuitive, we may prove the variational principle holds by decomposing $|\psi\rangle$ over the eigenbasis of H: $|\psi\rangle = \sum_i \alpha_i |\psi_i\rangle$ where $H|\psi_i\rangle = \epsilon_i |\psi_i\rangle$

and $\epsilon_0 < \epsilon_1 < \epsilon_2 < \dots$. Then, we have

$$\begin{aligned} \langle\psi| H |\psi\rangle &= \sum_i \epsilon_i |\alpha_i|^2 \\ &= \epsilon_0 \underbrace{|\alpha_0|^2}_{=\langle\psi|\psi\rangle - \sum_{i\geq 1}|\alpha_i|^2} + \sum_{i\geq 1}\epsilon_i|\alpha_i|^2 \\ &= \epsilon_0 \langle\psi|\psi\rangle + \sum_{i\geq 1}(\underbrace{\epsilon_i - \epsilon_0}_{>0})|\alpha_i|^2 \\ &\geq \epsilon_0 \langle\psi|\psi\rangle\,, \end{aligned} \tag{1.68}$$

with equality naturally when $|\psi\rangle = |\psi_0\rangle$.

The idea behind VQE therefore is to minimize the energy $\langle\psi| H |\psi\rangle$ over different quantum states $|\psi\rangle$. However, the space of possible states is immense – in order to traverse it we need to limit our search-space by restricting the classes

Year	Reference	System(s)	Ansatz	Max qubits	Platform	Hardware
2013	Peruzzo *et al.* [82]	HeH^+	UCC	2	SP	In-house
2015	Shen *et al.* [83]	HeH^+	UCC	1 qudit	TI	In-house
2015	Google Quantum [84]	H_2	UCC	2	SC	Google
2016	Santagati *et al.* [85]	Chlorophyll pair	Parametrized Hamiltonian	2	SP	In-house
2017	Kandala *et al.* [86]	H_2, LiH, BeH_2	Hardware Efficient	6	SC	IBM
2017	Colless *et al.*[87]	H_2 (excited states)	Hardware Efficient	2	SC	In-house
2018	Hempel *et al.* [88]	H_2, LiH	UCC	3	TI	In-house
2018	Kandala *et al.* [89]	H_2, LiH (magnetism)	Hardware Efficient	4	SC	IBM
2019	Nam *et al.* [90]	H_2O	UCC	4	TI	IonQ
2019	Smart & Mazziotti [91]	H_3	custom	3	SC	IBM
2019	McCaskey *et al.* [92]	NaH, RbH, KH	UCC and Hardware Efficient	4	SC	IBM, Rigetti
2020	Rice *et al.* [93]	LiH (dipole moment)	Hardware Efficient	4	SC	IBM
2020	Google AI Quantum [94]	H_6, H_8, H_{10}, H_{12}, HNNH	Hartree-Fock	12	SC	Google
2020	Gao *et al.* [95]	PSPCz	R_y	2	SC	IBM
2021	Kawashima *et al.* [96]	H_{10}	qubit-CC	2	TI	IonQ
2021	Eddins*et al.* [97]	H_2O	Entanglement Forging	5	SC	IBM
2021	Yamamoto *et al.* [98]	Crystalline Iron Model	UCCSD-PBC	2	SC	IBM
2021	Kirsopp *et al.* [99]	Oxazine derivatives	YXXX	4	SC, TI	IBM, Quantinuum
2022	Huang *et al.* [100]	H_2, CO	Linear Response	4	SC	In-house
2022	Lolur *et al.* [101]	HeH^+, LiH	Hardware Efficient	4	SC	IBM
2022	Leyton-Ortega *et al.* [102]	H_2	UCCSD	4	SC	IBM
2022	Liang *et al.* [103]	H_2, HeH^+, LiH, H_2O, NaH, CO_2	NAPA	6	SC	IBM
2022	Motta *et al.* [104]	H_3S^+	Entanglement Forging	6	SC	IBM
2022	O'Brien *et al.* [105]	Cyclobutene Ring	upCCD	10	SC	Google
2022	Khan *et al.* [106]	CH_4	UCCSD	6	TI	Quantinuum
2022	Zhao *et al.* [107]	Li_2O	oo-upCCD	12	TI	IonQ
2022	Guo *et al.* [108]	H_2, LiH, F_2	UCCSD	12	SC	Zuchongzhi 2.0
2023	Weaving *et al.* [3]	HCl	Hardware Efficient	3	SC	IBM
2023	Liu *et al.* [109]	H_2, HeH^+	Hardware Efficient	1 qudit	SC	In-house
2023	Dimitrov *et al.* [110]	CH_3F	pUCCD	11	TI	IonQ
2023	Jones *et al.* [111]	H_2O	UCCD	8	SC	IBM
2023	Liang *et al.* [112]	NH, BeH^+, F_2	SpacePulse	6	SC	IBM
2023	Weaving *et al.* [5]	N_2	Hardware-Aware ADAPT	5	SC	IBM

Table 1.5: A decade of experimental realizations of VQE for quantum chemistry; the list is not exhaustive. The works are listed chronologically by the date of initial preprint availability, not the final publication date. The platform keys are silicon photonic (SP), superconducting (SC) and trapped-ion (TI).

of state we consider. This motivates the concept of an *ansatz* state, the literal meaning of the German being to 'approach'. Ideally, we would fix a family of quantum states $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|\psi_{\mathrm{ref}}\rangle$ described by some number of real parameters $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots)$; for the purposes of quantum chemistry, it is customary to take the Hartree-Fock state as the reference $|\psi_{\mathrm{ref}}\rangle$. The energy we wish to optimize would then be a function of these parameters, $E(\boldsymbol{\theta}) \coloneqq \langle\psi(\boldsymbol{\theta})|\, H\, |\psi(\boldsymbol{\theta})\rangle$, and the minimization problem is then better posed as $E_0 = \min_{\boldsymbol{\theta}} E(\boldsymbol{\theta})$.

The overall VQE workflow is depicted in Figure 1.9, consisting of a quantum subroutine that outputs an energy expectation value $E(\boldsymbol{\theta}_k)$ at step k, embedded within a broader classical optimization scheme that selects a parameter update for the next iteration of the algorithm at $k+1$. Each operator $P_\ell \in \mathcal{P}_N$ in the Pauli decomposition of the Hamiltonian $H = \sum_{\ell=1}^{L} h_\ell P_\ell$ where $h_\ell \in \mathbb{R}$ must be estimated in separate prepare-and-measure routines.

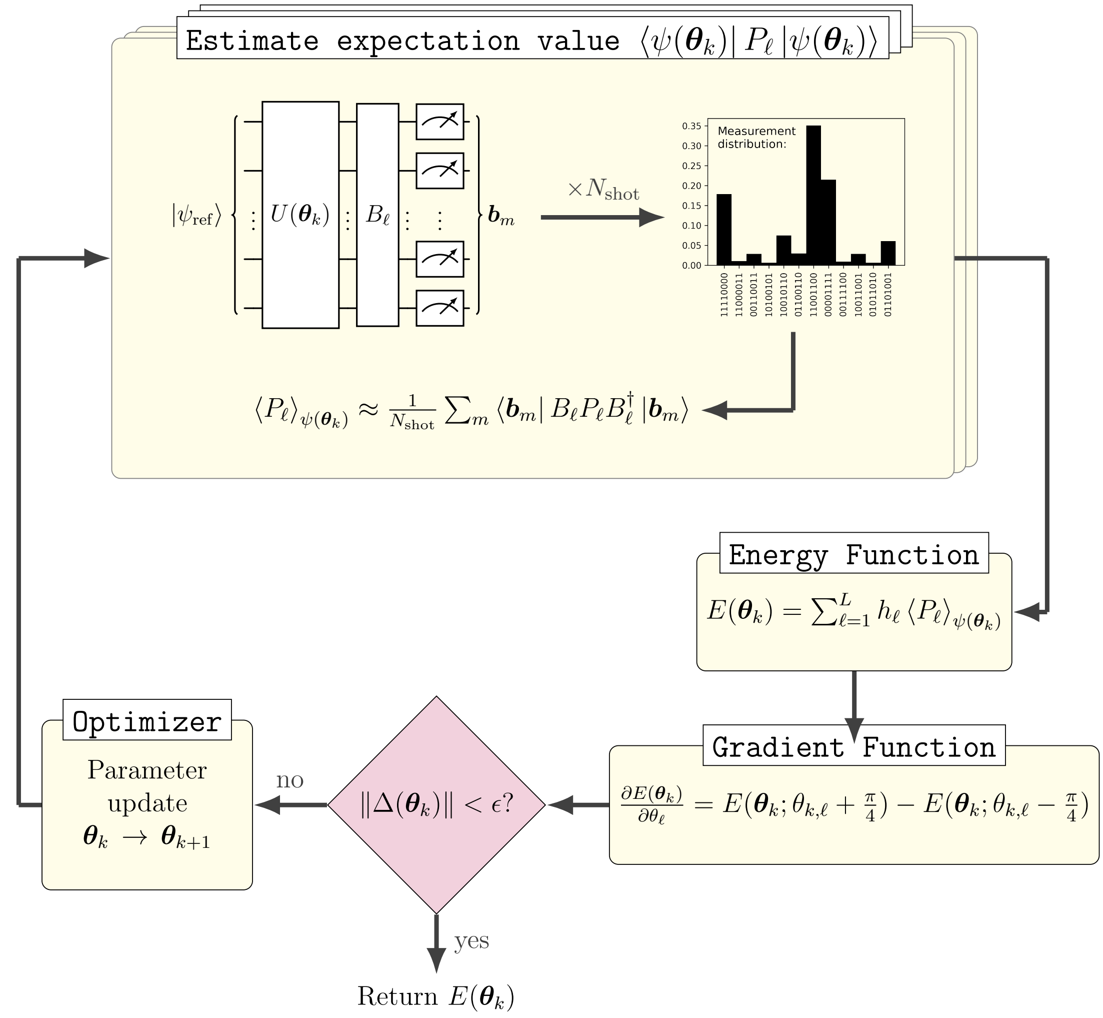


Figure 1.9: Variational Quantum Eigensolver workflow, involving the sampling of an ansatz circuit $|\psi(\boldsymbol{\theta}_k)\rangle$ on the quantum computer to extract an energy estimate that is fed back into a classical optimizer. The unitary B diagonalizes P or, in other words, it is a change-of-basis so that we may evaluate the expectation value of P using only diagonal measurements (i.e. in Pauli Z basis).

1.4.1 Evaluating Energies on a Quantum Computer

The essential routine of VQE is estimating energy expectation values by means of circuit sampling. Ultimately, we need to compute the energy $\langle\psi(\boldsymbol{\theta})| H |\psi(\boldsymbol{\theta})\rangle$ for some ansatz state $|\psi(\boldsymbol{\theta})\rangle$ and Hamiltonian observable operator decomposed over the Pauli basis: $H = \sum_{\ell=1}^{L} h_\ell P_\ell$ where $P_\ell \in \mathcal{P}_N, h_\ell \in \mathbb{R}$. The requirement of real Pauli coefficients h_ℓ ensures the operator is Hermitian and therefore has only real eigenvalues. This can be seen as, for $H|\psi\rangle = \epsilon|\psi\rangle$ we have $H^\dagger|\psi\rangle = \overline{\epsilon}|\psi\rangle$ and therefore $(H - H^\dagger)|\psi\rangle = (\epsilon - \overline{\epsilon})|\psi\rangle$. If $H = H^\dagger$ then it must be the case that $\epsilon - \overline{\epsilon} = 0 \implies \epsilon = \overline{\epsilon}$, which can hold only if $\epsilon \in \mathbb{R}$. The inner product is a linear map, so

$$\begin{aligned}\langle H\rangle_{\psi(\boldsymbol{\theta})} &:= \langle\psi(\boldsymbol{\theta})| H |\psi(\boldsymbol{\theta})\rangle \\ &= \langle\psi(\boldsymbol{\theta})| \sum_{\ell=1}^{L} h_\ell P_\ell |\psi(\boldsymbol{\theta})\rangle \\ &= \sum_{\ell=1}^{L} h_\ell \langle\psi(\boldsymbol{\theta})| P_\ell |\psi(\boldsymbol{\theta})\rangle \\ &= \sum_{\ell=1}^{L} h_\ell \langle P_\ell\rangle_{\psi(\boldsymbol{\theta})}\end{aligned} \tag{1.69}$$

and consequently we may assemble the full expectation value $\langle H\rangle_{\psi(\boldsymbol{\theta})}$ by evaluating its constituent Pauli terms individually. The consequence of this is that expectation value estimation is easily parallelized to multiple quantum devices. One may even run several such routines concurrently on different parts of the same chip given a surplus of qubits.

The only missing piece is the evaluation of $\langle P_\ell\rangle_{\psi(\boldsymbol{\theta})}$. However, as discussed in Section 1.2.2, quantum computers typically only allow diagonal measurement, namely in the Pauli Z basis. This means that all measurements extracted from the device will be eigenstates of Pauli Z operators and therefore the expectation value for terms containing X or Y factors will be incorrect if we proceed naïvely. This is easily remedied by recalling the basis change operations $Z = \mathrm{Had} X \mathrm{Had} = \mathrm{Had} S^\dagger Y S \mathrm{Had}$. Letting $\mathcal{I}_X, \mathcal{I}_Y$ be the sets indexing tensor factors of P_ℓ containing a Pauli X or Y, respectively, the change-of-basis operator B_ℓ, as previously seen in Figure 1.9, takes the explicit form $B_\ell = \bigotimes_{\mathcal{I}_X \cup \mathcal{I}_\mathcal{Y}} \mathrm{Had} \bigotimes_{\mathcal{I}_\mathcal{Y}} S^\dagger$. Application of this to P_ℓ yields $B_\ell P_\ell B_\ell^\dagger \in \{I, Z\}^{\otimes N}$, i.e. it has been diagonalized. We can also apply the change-of-basis to the ansatz circuit $B_\ell|\psi(\boldsymbol{\theta})\rangle$ and then sample diagonal elements as normal, obtaining a set of binary measurement outcomes $\mathcal{M}_\ell = \{\boldsymbol{b}_m\}_{m=1}^{N_{\mathrm{shot}}}$ where N_{shot} is the number of times we sample the state. These data are fed into a statistical estimator

$$\mathcal{E}_\ell(\mathcal{M}_\ell) = \frac{1}{N_{\mathrm{shot}}} \sum_{m=1}^{N_{\mathrm{shot}}} \langle\boldsymbol{b}_m| B_\ell P_\ell B_\ell^\dagger |\boldsymbol{b}_m\rangle \tag{1.70}$$

that reproduces the true value in expectation $\mathbb{E}(\mathcal{E}_\ell) = \langle P_\ell\rangle_{\psi(\boldsymbol{\theta})}$ with no bias [6].

To evaluate an energy estimate for the whole Hamiltonian, this process is repeated for each Pauli term independently (possibly in parallel, as mentioned

---

[6] While this estimator is unbiased, the presence of quantum noise introduces bias that we attempt to counteract through error mitigation, discussed in Chapter 4.

above). This results in measurement batches $\mathcal{M} = \{\mathcal{M}_\ell\}_\ell$ and the overall Hamiltonian estimator follows by linearity:

$$\begin{aligned}\mathcal{E}(\mathcal{M}) &= \sum_{\ell=1}^{L} h_\ell \mathcal{E}_\ell(\mathcal{M}_\ell) \\ &= \frac{1}{N_{\text{shot}}} \sum_{\ell=1}^{L} h_\ell \sum_{m=1}^{N_{\text{shot}}} \langle \boldsymbol{b}_m | B_\ell P_\ell B_\ell^\dagger | \boldsymbol{b}_m \rangle .\end{aligned} \tag{1.71}$$

Just as in the individual Pauli case, this estimator is unbiased, $\mathbb{E}(\mathcal{E}) = \langle H \rangle_{\psi(\boldsymbol{\theta})}$, and the error is on the order $\mathcal{O}(L N_{\text{shot}}^{-\frac{1}{2}})$ [82] where L is the number of terms in the Hamiltonian.

In fact, such a measurement scheme is not restricted to the Pauli basis. It is applicable whenever we can decompose the Hamiltonian over components

$$H = \sum_{\ell=1}^{L} H_\ell \tag{1.72}$$

such that H_ℓ is classically efficient to diagonalize, with U_ℓ the unitary diagonalizing H_ℓ. The benefit of this is it is often possible to express the Hamiltonian using fewer components H_ℓ than in the full Pauli decomposition and therefore reduces the measurement overhead. This observation underpins all measurement reduction techniques in quantum computation and plays out exactly as the Pauli case above, just with a different change-of-basis operator U. We sample from the circuit $U |\psi(\boldsymbol{\theta})\rangle$ and obtain an estimate $\frac{1}{N_{\text{shot}}} \sum_{m=1}^{N_{\text{shot}}} \langle \boldsymbol{b}_m | U_\ell H_\ell U_\ell^\dagger | \boldsymbol{b}_m \rangle$ for each H_ℓ.

Common choices for H_ℓ include qubit-wise commuting [115], commuting [116] or anticommuting [117–120] clique covers of the Hamiltonian Pauli terms. The downside of the latter two is the added coherent resource necessary to realize the change-of-basis unitary U in-circuit, which usually precludes their practical application. For quantum chemistry, where the number of Hamiltonian terms grows as $\mathcal{O}(N^4)$ with N the number of spin-orbitals (qubits), the commuting or anticommuting grouping strategies yield a linear improvement in the number of simultaneously measurable components H_ℓ in the decomposition of Equation (1.72). This results in $L = \mathcal{O}(N^3)$ and thus an asymptotic improvement to the measurement overhead [116, 118].

The qubit-wise commuting grouping strategy is an extension of the Pauli case, where the diagonalizing unitary is of the same form as B_ℓ above. We simply gather Hamiltonian terms for which B_ℓ is the same (up to identity positions) and measure them all simultaneously. Specifically, two Pauli operators P, Q qubit-wise commute if $[P^{(i)}, Q^{(i)}] = 0$ for all qubit positions $i \in \{1, \ldots, N\}$. This is a stricter condition than regular commutation, for example the Paulis XX and ZZ commute but not qubit-wise. Owing to the negligible increase in circuit depth, this is the measurement reduction strategy that we favour for all the quantum simulations in this work. While it does not produce an formal asymptotic improvement like the commuting or anticommuting cases above, it still saves an appreciable pre-factor of ≈ 3 [115] that makes near-term simulations more achievable.

1.4.2 Calculating Gradients for Parametrized Circuits

Now that we are able to evaluate energy values on a quantum computer, as per the previous Section 1.4.1, we have an objective function to optimize in a VQE routine. However, to accelerate the convergence of the optimizer we would ideally like to be able to compute parameter gradients on the hardware as well.

In Section 1.4 we expressed the ansatz state as a unitary $U(\boldsymbol{\theta})$ applied to some reference state $|\psi_{\text{ref}}\rangle$. To expose an explicit dependence on the parameters $\boldsymbol{\theta}$, we shall write $U(\boldsymbol{\theta}) = e^{iA(\boldsymbol{\theta})}$ for some operator $A(\boldsymbol{\theta}) = \sum_k \theta_k P_k$ with $\theta_k \in \mathbb{R}, P_k \in \mathcal{P}_N$. We refer to $A(\boldsymbol{\theta})$ as the *ansatz generator* and it is necessary for the parameters θ_k to be real, else $iA(\boldsymbol{\theta})$ would not be anti-Hermitian and thus $e^{iA(\boldsymbol{\theta})}$ could not be unitary.

Our ansatz state now has the form

$$|\psi(\boldsymbol{\theta})\rangle = e^{iA(\boldsymbol{\theta})} |\psi_{\text{ref}}\rangle \tag{1.73}$$

and application of the chain rule yields

$$\frac{\partial}{\partial\theta_k} |\psi(\boldsymbol{\theta})\rangle = iP_k |\psi(\boldsymbol{\theta})\rangle . \tag{1.74}$$

Combining this with the product rule, we have

$$\begin{aligned} \frac{\partial}{\partial\theta_k} \langle H\rangle_{\psi(\boldsymbol{\theta})} &= \langle\psi(\boldsymbol{\theta})| \left[-iP_k H + iHP_k\right] |\psi(\boldsymbol{\theta})\rangle \\ &= i \left\langle [H, P_k]\right\rangle_{\psi(\boldsymbol{\theta})} ; \end{aligned} \tag{1.75}$$

this is the formulation proposed by Grimsley et al. [121]. However, if one observes that

$$\begin{aligned} i[H, P_k] &= \frac{1}{2}\big[(I - iP_k)H(I + iP_k) - (I + iP_k)H(I - iP_k)\big] \\ &= e^{-i\frac{\pi}{4}P_k} H e^{i\frac{\pi}{4}P_k} - e^{i\frac{\pi}{4}P_k} H e^{-i\frac{\pi}{4}P_k} \end{aligned} \tag{1.76}$$

then we recover the parameter shift rule of Parrish et al. [122], namely

$$\frac{\partial}{\partial\theta_k} \langle H\rangle_{\psi(\boldsymbol{\theta})} = \langle H\rangle_{e^{i\frac{\pi}{4}P_k}\psi(\boldsymbol{\theta})} - \langle H\rangle_{e^{-i\frac{\pi}{4}P_k}\psi(\boldsymbol{\theta})} . \tag{1.77}$$

Therefore, we may evaluate the expectation value $\langle H\rangle_{\psi(\boldsymbol{\theta})}$ at $\theta_k + \frac{\pi}{4}$ and $\theta_k - \frac{\pi}{4}$, with the partial gradient with respect to θ_k being their difference. The curious thing here is that, while the parameter shift approach looks to be a finite-difference, it is analytic. We opt for the latter method of gradient calculation; this allows us to avoid storing a large collection of commutators in memory – each requiring decomposition into qubit-wise commuting measurement groups – at the expense of one additional expectation value calculation per partial gradient.

1.4.3 Ansatz Circuit Design

Choosing an effective ansatz circuit for VQE is a challenge in itself. It needs to be sufficiently expressible to capture the ground state energy (or at least get satisfactorily close to it), while not being so over-complete that the optimization becomes unnecessarily cumbersome. There is also the more insidious issue of barren plateaus, which arises from the curse of dimensionality whereby the optimization landscape becomes locally flat, with the parameter gradient vanishing exponentially with the system size [123, 124]. This is a critical problem for variational algorithms; while many works strive to design ansätze devoid of barren plateaus, there is a poignant question of whether their absence implies classical simulability [125]. Some evidence supporting this statement has been presented, although it is not yet entirely conclusive and nor does it disprove quantum advantage in the near term. For example, it is still possible that advantage can be derived from a particular ansatz choice, as state sampling is classically hard. It is also worth noting that barren plateaus can be induced by noise [126], a further obstacle to near-term quantum utility.

Hardware Efficient Circuits

Hardware efficient circuits are typically developed with a specific quantum device in mind so that they may be natively implemented with no transpilation overhead. They are designed to access the largest possible region of Hilbert space for the shallowest circuit depths. A hardware efficient ansatz consists of two ingredients: rotation layers and entangling layers. These building blocks will often be layered many times, each repetition bringing with it a richer computational state-space. There are several decisions that go into constructing such an ansatz: the choice of rotation gates and the entangling pattern, which is usually dictated by the qubit topology of the target hardware. For example, in Figure 1.10 we display

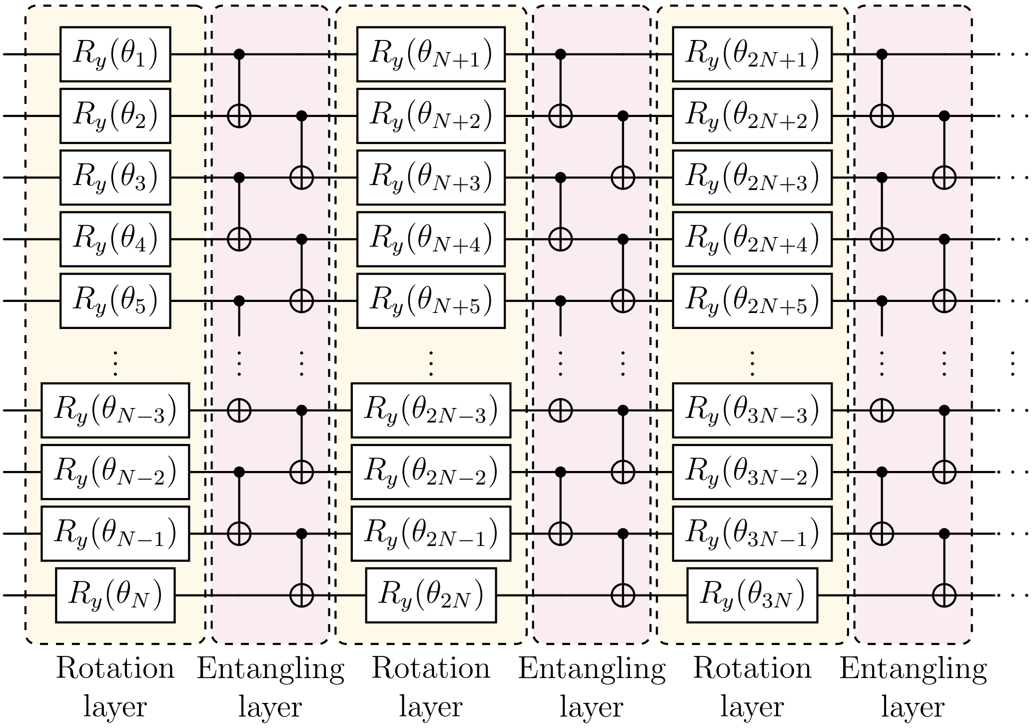


Figure 1.10: A hardware efficient ansatz with R_y rotation layers and linear entanglement. In general, a hardware efficient circuit might consist of different rotations (possibly multiple types per layer) and entangling structures that ideally mirror the device connectivity.

a hardware efficient ansatz consisting of R_y rotations and linear entanglement blocks. Device coherence permitting, this pattern may be repeated as many times L as desired. In this example, we note the number of parameters is LN, therefore one should be wary of the added optimization burden when running variational algorithms such as VQE. In fact, our level of caution should be even greater than that taken just at face value, the reason being this is precisely the regime in which barren plateaus manifest most profoundly.

As mentioned above, barren plateaus are symptomatic of vanishing parameter gradients (which may do so exponentially fast), causing a local flattening of the optimization landscape. For this reason, there is debate over the usefulness of hardware efficient ansätze [127]. At short depths, the issue of barren plateaus is avoided, however the variational flexibility is insufficient to achieve quantum advantage. Conversely, deeper circuits access a vast expanse of the Hilbert space, but are vulnerable to barren plateaus due to sheer system dimensionality. The best we can hope for is a sweet spot between these two extrema.

A further point of note is that this barren plateau problem is not exclusively a feature of the ansatz circuit itself, but also depends on the locality of the observable being measured [127]. To illustrate this, in Figure 1.11 we highlight the gates of our hardware efficient ansatz that are supported for a highly localized observable. We trace back through the circuit and find that only a reduced subset of gates contribute to the measurement outcome. This backpropagation through time is known as the *light-cone* of the observable.

For a local observable measured on a shallow circuit, one may find that the light-cone expansion does not saturate all the qubits, resulting in a small effective circuit and thus alleviating barren plateaus. However, if the circuit is sufficiently deep, then eventually the light-cone will expand to fill the circuit, as seen in Figure 1.11. Therefore, the severity of barren plateaus is a function of both the circuit complexity *and* observable locality. We encounter a light-cone reduction in Section 5.4, in which we simulate time evolution for the magnetization of a single spin in an Ising model; for early times the circuit is shallow and therefore

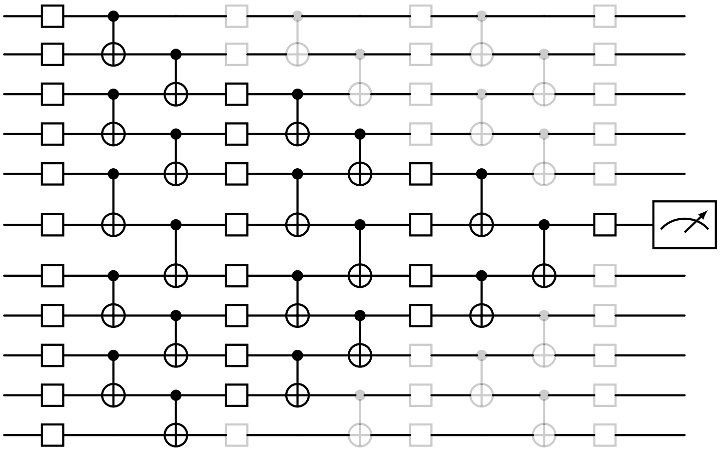


Figure 1.11: The light-cone for a local observable (in this case supported on just a single qubit) over the hardware efficient ansatz from Figure 1.10. Gates that have been grayed-out are external to the light-cone and therefore have no influence on the measured outcome.

the light-cone is small, but as the system evolves the circuit deepens and thus the light-cone expands. This is shown in Figures 5.20 and 5.21 where we plot the number of qubits in the effective light-cone of the single-qubit Z observable. In quantum chemistry, on the other hand, we can encounter complete nonlocality (i.e. observables that act across every qubit), which is the worst-case scenario for a hardware efficient ansatz. For this reason, it is necessary to develop circuits tailored for the electronic structure problem.

Chemically Motivated Circuits

While hardware efficient circuits are convenient when it comes to executing them on a quantum device, they do not take any of the problem structure into account. For example, in quantum chemistry there are numerous symmetries that would ideally be baked into the ansatz circuit. For example, the number of particles is fixed, and therefore computational basis states that possess the incorrect hamming weight are redundant. It is not possible for a hardware efficient ansatz to filter these out, since by design they access the largest feasible region of the Hilbert space. Instead, we need to design ansätze to incorporate such symmetry at the circuit level.

Recall in Section 1.3.3 we introduced the cluster operator $T = T_1 + T_2 + T_3 \ldots$ partitioned into terms relating to single, double, triple excitations etc. above the Hartree-Fock reference $|\mathrm{HF}\rangle = |1, 1, \ldots, 1, 0, \ldots, 0\rangle$. Each of the cluster terms T_i may be written in terms of fermionic creation and annihilation operators

$$T_n = \frac{1}{(n!)^2} \sum_{p_1,\ldots,p_{2n}} t_{p_1,\ldots,p_{2n}} a^\dagger_{p_1} \ldots a^\dagger_{p_n} a_{p_{n+1}} \ldots a_{p_{2n}}. \tag{1.78}$$

It was noted that we may construct a unitary formulation of coupled cluster theory by observing that $T - T^\dagger$ is antihermitian and therefore $e^{T-T^\dagger}$ is a unitary operator. Moreover, from Section 1.3.4 we know how to map fermionic excitation operators onto Pauli operators via a qubit transformation q, so $q(T_n - T_n^\dagger) = i \sum_\ell \theta_\ell^{(n)} P_\ell^{(n)}$ where $P_i^{(n)} \in \mathcal{P}_N$ may act as X or Y on $2n$ qubit positions, with the other qubit positions either identity or Pauli Z for parity calculation.

Truncating the cluster operator at some maximum excitation order, typically to double excitations $T_1 + T_2$, we map onto qubits $q(T_1 + T_2) = \sum_\ell \theta_\ell^{(1)} P_\ell^{(1)} + \sum_\ell \theta_\ell^{(2)} P_\ell^{(2)} = \sum_\ell \theta_\ell P_\ell$ and Trotterize the exponential $\exp\Big(\sum_\ell \theta_\ell P_\ell\Big)$ as per Section 1.2.4 for an approximate decomposition into exponentiated Paulis, which we may realize in circuit form using the approach of Section 1.2.3. For variational algorithms, it is often sufficient to Trotterize only to first-order since the optimization over the ansatz parameters will counteract the induced Trotter error [22]; explicitly, this is $\exp\Big(\sum_\ell \theta_\ell P_\ell\Big) \approx \prod_\ell \exp\big(\theta_\ell P_\ell\big)$. By initializing the circuit in the Hartree-Fock state via an initial layer of X-gates in relevant qubit positions at the beginning on computation, we obtain a circuit that respects the molecular symmetries. This is known as the unitary coupled cluster singles doubles (UCCSD) ansatz.

Example 1.6: Molecular Hydrogen UCCSD

In Examples 1.3 and 1.5 we built fermionic and Jordan-Wigner encoded qubit Hamiltonians for H_2. We now go through the same exercise for the cluster operator. Because of the spin parity symmetry, no single excitations are valid (this is explored in Secion 4.3), leaving just one double excitation $a_2^\dagger a_3^\dagger a_0 a_1$ (and its various permutations) so that

$$T = t_{0,1,2,3}[a_2^\dagger a_3^\dagger a_0 a_1 - a_2^\dagger a_3^\dagger a_1 a_0 - a_3^\dagger a_2^\dagger a_0 a_1 + a_3^\dagger a_2^\dagger a_1 a_0]. \tag{1.79}$$

Mapping this via Jordan-Wigner, we get

$$\begin{aligned} q_{\mathrm{JW}}(T - T^\dagger) = i\frac{t_{0,1,2,3}}{2}[&YYYX + YXXX + YYXY + XYXX \\ &- YXYY - XXYX - XYYY - XXXY]. \end{aligned} \tag{1.80}$$

Note all these Pauli operators commute and therefore Trotterization will introduce no error. This is actually a consequence of a particular feature of the H_2 STO-3G problem, which we discuss in Example 3.1.

The issue is that the UCCSD circuit is usually too deep for practical applications, consisting of $\mathcal{O}(N^4)$ variational parameters and a circuit depth of $\mathcal{O}(N^5)$ [128]. This motivated research into more compact symmetry-preserving circuits based on simple primitives that may be replicated across the circuit, more in the vein of hardware efficiency. Effectively, we need a gate that shuffles the positions of occupied qubits, preserving the Hamming weight of each computational basis state. We have already seen one gate that does this: the SWAP operation. Recall that this has the property of swapping qubit states between sites, namely $|0,1\rangle \rightleftarrows |1,0\rangle$, while the $|0,0\rangle$ and $|1,1\rangle$ states are untouched. For the purposes of a variational circuit, it would be desirable to have a continuous version of the SWAP gate that mixes qubit occupancies together, in other words

$$\mathrm{SWAP}(\theta,\phi) \coloneqq \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & e^{i\phi}\sin(\theta) & 0 \\ 0 & e^{-i\phi}\sin(\theta) & -\cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.81}$$

This is known as a *Givens rotation* and forms the basic building block of low-depth particle-preserving ansatz circuits [129]. Note the discrete SWAP gate is recovered for $\theta = \pi/2, \phi = 0$. This gate is decompose as a circuit in Figure 1.12.

A more recent circuit design that makes use of the Givens rotation above and has received recent attention [66] is the unitary cluster Jastrow [128], favoured for its $\mathcal{O}(N^2)$ parameter scaling while being chemically intuited. This question of ansatz parametrization scaling evokes the previous discussion of the trade-off in expressibility and alleviation of barren-plateaus in hardware efficient circuits; we recall that very shallow circuits could be prone to classical simulability [125], hence we might expect a lower bound on the number of variational parameters we require for quantum advantage in electronic structure. Moreover, there is work that suggests coupled-cluster circuit are also prone to the barren plateau issue [130].

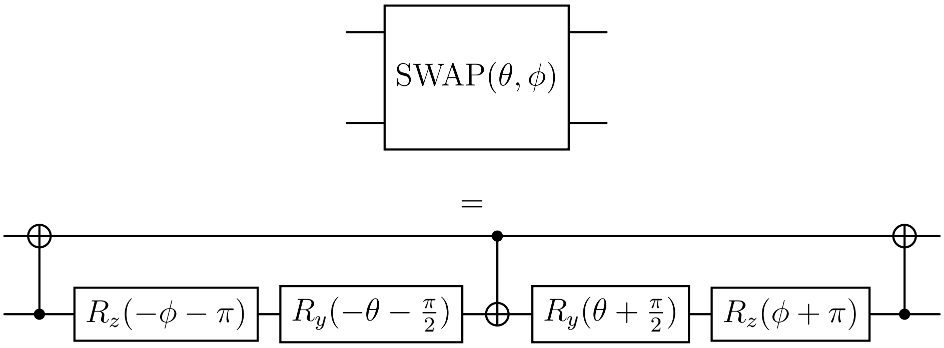


Figure 1.12: A Givens rotation, referred to here as a continuous SWAP parametrized by two angles θ, ϕ.

1.4.4 ADAPT-VQE

Adaptive Derivative-Assembled Pseudo-Trotter (ADAPT) VQE [121] is a promising method of ansatz construction that retains some problem specificity while resulting in dramatically reduced circuit depths compared with conventional chemically-motivated approaches such as unitary coupled cluster with single and double excitations (UCCSD) [22, 51]. The implementation we describe here is the qubit-ADAPT-VQE [131] variant, the central component of which is a pool of Pauli operators $\mathcal{P}$. This pool could, for example, consist of the Jordan-Wigner encoded excitation terms in UCCSD; the key observation is that we typically do not need the full set of excitations in its entirety, and many of the Pauli terms are included simply to enforce the fermionic anticommutation relations (see Section 1.3.4).

While this is a theoretical requirement to correctly represent the problem in the purest sense, in practice we find that fermionic anticommutation constraints can be relaxed at the ansatz level while still obtaining high accuracy results. To expand on this point further, after a mapping onto qubits via the Jordan-Wigner transformation, single, double and triple excitations account for 4, 16 and 64 Pauli operator terms, respectively. Furthermore, referring to Example 1.4, we see that many of the Pauli terms generated from a fermionic excitation are qualitatively the same, in that they map to the same computational basis states – not all of these are necessary to reach the desired level of algorithmic accuracy. By selecting only a subset of such Pauli terms, qubit-ADAPT-VQE can achieve massive reductions in circuit depth.

The general ADAPT framework is provided in Algorithm 1. From the operator pool $\mathcal{P}$, we build an ansatz circuit $|\psi\rangle$ iteratively by appending the term that maximizes some scoring function f at each step. In the standard approach we take the partial derivative at zero after appending a given pool element $P \in \mathcal{P}$, specifically

$$f(P) := \frac{\partial}{\partial\theta} \langle\psi| e^{-i\theta P} H e^{i\theta P} |\psi\rangle \big|_{\theta=0}, \tag{1.82}$$

which may be evaluated either with the parameter shift rule [122] or by measuring the commutator $[H, P]$ [121], as discussed in Section 1.4.2. By calculating the pool scores $f(\mathcal{P})$ and identifying the maximal term, we extend $|\psi\rangle \to e^{i\theta P} |\psi\rangle$ and re-optimize the ansatz parameters via regular VQE before repeating.

We note a few properties of ADAPT-VQE. First of all, the minimized energy E_n will decrease variationally towards the ground state energy with increasing

Algorithm 1: The qubit-ADAPT-VQE algorithm.

```
Input : Operator pool $\mathcal{P}$, initial state $|\psi_0\rangle$, scoring function $f : \mathcal{P} \mapsto \mathbb{R}$,
        score tolerance $\delta_f > 0$, convergence threshold $\delta_c > 0$ and
        maximum number of iterations $n_{\max} \in \mathbb{N}$.
Output: Optimized energy $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})| H |\psi(\boldsymbol{\theta})\rangle$ and ansatz
        $|\psi(\boldsymbol{\theta})\rangle = \prod_n e^{i\theta_n P_n} |\psi_0\rangle$.
$n \leftarrow 0$ ;
$E_0 \leftarrow 0$ ;
while $(\Delta_f > \delta_f) \wedge (\Delta_c > \delta_c) \wedge (n_{\max} > n)$ do
    Identify optimal pool operator:
        $P_{n+1} \leftarrow \underset{P \in \mathcal{P}}{\arg\max} |f(P; \psi_n)|, \quad \Delta_f \leftarrow |f(P_{n+1}; \psi_n)|$
    Append term to growing ansatz:
        $|\psi_{n+1}\rangle \leftarrow e^{i\theta_{n+1} P_{n+1}} |\psi_n\rangle$
    Optimize parameters through VQE:
        $\boldsymbol{\theta}_{n+1} \leftarrow \underset{\boldsymbol{\theta} \in \mathbb{R}^{\times(n+1)}}{\arg\min} E(\boldsymbol{\theta}),$
        $E_{n+1} \leftarrow E(\boldsymbol{\theta}_{n+1}),$
        $\Delta_c \leftarrow |E_{n+1} - E_n|$
    $n \leftarrow n + 1$ ;
end
```

$n \in \mathbb{N}$, since the ansatz becomes increasingly more expressible. However, the ansatz complexity also increases so optimization becomes more demanding and, if running the algorithm on a real quantum device, the increase in circuit depth at each iteration will also induce a greater level of noise. Secondly, ADAPT-VQE is resilient to the barren plateau problem [132], since each iteration of the algorithm is effectively warm-started from the solution of the previous iterate. Finally, the operator pool $\mathcal{O}$ is never exhausted, meaning it is possible to select the same term more than once. This gives rise to the 'Pseudo-Trotter' description of ADAPT-VQE, in the sense that it somewhat resembles the Trotter expansion of Equation (1.33) but with only select terms duplicated.

In Figure 1.13 we demonstrate a noiseless qubit-ADAPT-VQE routine for a 12-qubit instance of molecular oxygen (O_2) at its equilibrium bond length of 1.286Å (in STO-3G), in which we target its triplet ground state. The main point of note is that the adaptive VQE algorithm surpasses the CCSD energy at a fraction of the circuit depth compared with the full UCCSD circuit, just 4%. Once we hit 6% the algorithm achieves the target accuracy of 1.6 mHa, demonstrating that ADAPT-VQE can achieve high levels of accuracy for drastically reduced circuit depths as compared with full chemically-motivated constructions such as UCCSD.

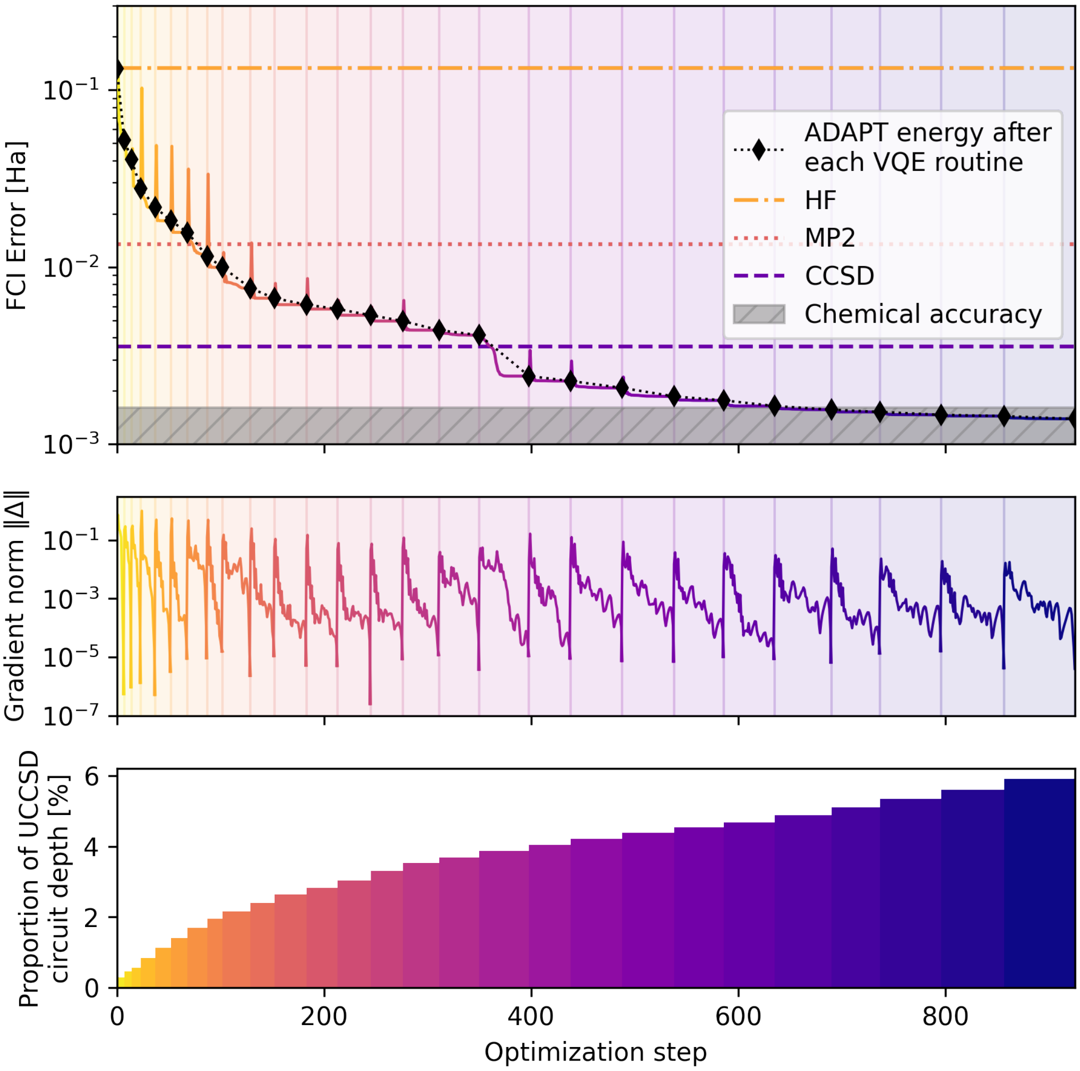


Figure 1.13: Example of ADAPT-VQE for a 12-qubit contextual subspace of molecular oxygen, O_2 STO-3G, in its triplet ground state at the equilibrium bond length of 1.286Å. The diamond symbols in the top panel indicate the energy at the end of each VQE routine, given by the solid line. At each ADAPT step (delineated by the regions of different colour) a term is selected from the pool of coupled cluster singles doubles excitations and appended to a growing ansatz, hence the circuit depth increases and thus the optimization becomes more difficult. Notice that ADAPT-VQE surpasses the CCSD energy with a circuit that is just 4% of the total UCCSD ansatz depth, while at 6% it achieves chemical accuracy (1.6 mHa error).

1.4.5 Hardware-Aware ADAPT-VQE

The ADAPT-VQE algorithm allows us to incorporate chemical intuition into our ansatz circuit for modest circuit depths, representing a middle ground between hardware efficiency and full-blown chemically motivated circuits such as unitary coupled-cluster with single and double excitations. However, unlike hardware efficient ansätze, it does not necessarily respect the qubit topology of the target system – at least, not in its current form.

To address this, we modify the standard pool scoring function above in Equation (1.82) to enforce hardware-awareness in the adaptive circuit construction, thus minimizing the number of SWAP operations incurred through transpilation. We achieve this by ranking approximate subgraph isomorphisms in the hardware topology, described by a graph $\mathcal{G}_{\text{target}} = (\mathcal{N}_{\text{target}}, \mathcal{E}_{\text{target}})$ where $\mathcal{N}_{\text{target}}$ is the set of available qubits and $\mathcal{E}_{\text{target}} \subset \mathcal{N}_{\text{target}}^{\times 2}$ is the edge-set indicating that two qubits may be natively coupled via some nonlocal operation on the hardware. We define an isomorphism between two graphs $\mathcal{G}$ and $\mathcal{H}$ to be a bijective map $g : \mathcal{N}_{\mathcal{G}} \mapsto \mathcal{N}_{\mathcal{H}}$ such that, if $(u, v) \in \mathcal{E}_{\mathcal{G}}$, then $(g(u), g(v)) \in \mathcal{E}_{\mathcal{H}}$. In other words, an isomorphism is a mapping from nodes of $\mathcal{G}$ onto nodes of $\mathcal{H}$ that preserves the adjacency structure. Furthermore, two graphs are said to be *subgraph* isomorphic if $\mathcal{G}$ is isomorphic to a subgraph of $\mathcal{H}$; we use the VF2^{++} algorithm [133] as implemented in the *NetworkX* Python package [134] for subgraph isomorphism matching.

In order to reweight the standard score assigned to a given pool operator $P \in \mathcal{P}$, we construct a weighted graph $\mathcal{G}_{\text{circuit}} = (\mathcal{N}_{\text{circuit}}, \mathcal{E}_{\text{circuit}})$ for the circuit $e^{i\theta P} |\psi\rangle$ and bias with respect to a notion of distance from the nearest subgraph isomorphism, described in Algorithm 2. This works by iteratively deleting collections of qubits $\boldsymbol{n} \in \mathcal{N}_{\text{circuit}}^{\times d}$ from the ansatz circuit and any associated edges in the corresponding coupling graph, terminating once a subgraph isomorphism is identified. Here, d is the search-depth, which begins at $d = 0$ with no qubits deleted and is incremented at each step; since the number of distinct $\boldsymbol{n}$ is $\binom{|\mathcal{N}_{\text{circuit}}|}{d}$, we truncate at some maximum depth D and any pool operator for which no subgraph isomorphism was found with $d \leq D$ receives a score of zero. Otherwise, with the function $s(\boldsymbol{n})$ that sums edge-weights connected to the nodes $\boldsymbol{n}$, our new Hardware-Aware ADAPT-VQE scoring function becomes

$$f(P) \coloneqq \left(1 - \frac{s(\boldsymbol{n})}{W}\right)^{b} \cdot \frac{\partial}{\partial\theta} \langle\psi| e^{-i\theta P} H e^{i\theta P} |\psi\rangle \big|_{\theta=0} \tag{1.83}$$

where $s(\boldsymbol{n})$ is the output of Algorithm 2, $W = \sum_{(u,v,w)\in\mathcal{E}_{\text{circuit}}} w$ is the total sum of edge-weights and $b \geq 0$ is the biasing strength. This allows one to control the severity with which non-subgraph-isomorphic circuits are penalised. While the depth d does not explicitly appear in Equation (1.83), since $|\boldsymbol{n}| = d$ we will have more edge-weights included in $s(\boldsymbol{n})$ for larger depths and hence will be penalized more.

We test this new hardware-aware ADAPT objective function for 12-qubit molecular nitrogen, N_2, at a stretched bond length of 2Å; this system can be challenging for many wavefunction methods. In Figure 1.14, we compare error against the number of CNOT gates in the transpiled circuit for our new scoring function, versus the standard qubit-ADAPT-VQE approach. For the target topology we choose a 12-qubit ring, which is found as a subgraph of the heavy-hex topology presented in Figure 1.5. As described in Section 1.2.5, transpilation is the mapping of a given circuit onto the target quantum device, which may not natively support the required entangling operations and thus expensive SWAP operations are incurred to compensate for discrepancies in the qubit connectivity. The number of two-qubit gates required to transpile the ansatz circuit for the chosen 12-qubit ring is seen to be dramatically reduced, while maintaining similar errors compared with the hardware-agnostic approach. For fairness, both techniques were transpiled using a basic level of circuit optimization (e.g. cancellation of inverse gates).

Algorithm 2: Hardware-aware ADAPT-VQE biasing function.

```
Input : Pool operator $P \in \mathcal{P}$, optimal state $|\psi\rangle$ from previous ADAPT
        iteration, target topology graph $\mathcal{G}_{\text{target}} = (\mathcal{N}_{\text{target}}, \mathcal{E}_{\text{target}})$, bias
        $b > 0$ and maximum search depth $D \in \mathbb{N}$.
Output: $f(P)$, a score for the pool operator $P$.

For two graphs $\mathcal{G}, \mathcal{H}$ the function VF2$^{++}(\mathcal{G}, \mathcal{H})$ [133] returns True if $\mathcal{G}$ is
  subgraph isomorphic to $\mathcal{H}$ and False otherwise;
Build the weighted graph $\mathcal{G}_{\text{circuit}} = (\mathcal{N}_{\text{circuit}}, \mathcal{E}_{\text{circuit}})$ for $e^{i\theta P} |\psi\rangle$. Here,
  $|\mathcal{N}_{\text{circuit}}| < |\mathcal{N}_{\text{target}}|$ represent the circuit qubits and $\mathcal{E}_{\text{circuit}} \subset \mathcal{N}_{\text{circuit}}^{\times 2} \times \mathbb{N}$
  indicate the presence of a nonlocal operation in-circuit, weighted by the
  total number of occurrences. The sum of weights is
  $W = \sum_{(u,v,w) \in \mathcal{E}_{\text{circuit}}} w$.
$\Delta_f \leftarrow \frac{\partial}{\partial \theta} \langle\psi| e^{-i\theta P} H e^{i\theta P} |\psi\rangle \big|_{\theta=0}$, 'standard' score (1.82) ;
$d \leftarrow 0$, the subgraph isomorphism distance ;
if VF2$^{++}(\mathcal{G}_{\text{circuit}}, \mathcal{G}_{\text{target}})$ then
    Already subgraph isomorphic – no biasing;
    return $\Delta_f$
else
    while $D > d$ do
        $d \leftarrow d + 1$ ;
        Order node collections $\boldsymbol{n} \in \mathcal{N}_{\text{circuit}}^{\times d}$ of size $d$ by their summed
          edge-weights
            $s(\boldsymbol{n}) := \sum_{n \in \boldsymbol{n}} \sum_{\substack{(u,v,w) \in \mathcal{E}_{\text{circuit}} \\ n=u \text{ or } n=v}} w$;
        for $\boldsymbol{n} \in \arg\operatorname{sort}_{\boldsymbol{n} \in \mathcal{N}_{\text{circuit}}^{\times d}} s(\boldsymbol{n})$ do
            Form the subgraph $\mathcal{G}(\boldsymbol{n}) \subset \mathcal{G}_{\text{circuit}}$ in which the nodes $\boldsymbol{n}$ have
              been deleted from $\mathcal{G}_{\text{circuit}}$;
            if VF2$^{++}(\mathcal{G}(\boldsymbol{n}), \mathcal{G}_{\text{target}})$ then
                return $\Delta_f \cdot (1 - s(\boldsymbol{n})/W)^b$
            end
        end
    end
    If the maximum depth is reached without finding a subgraph
      isomorphism, the score is set to zero;
    return 0
end
```

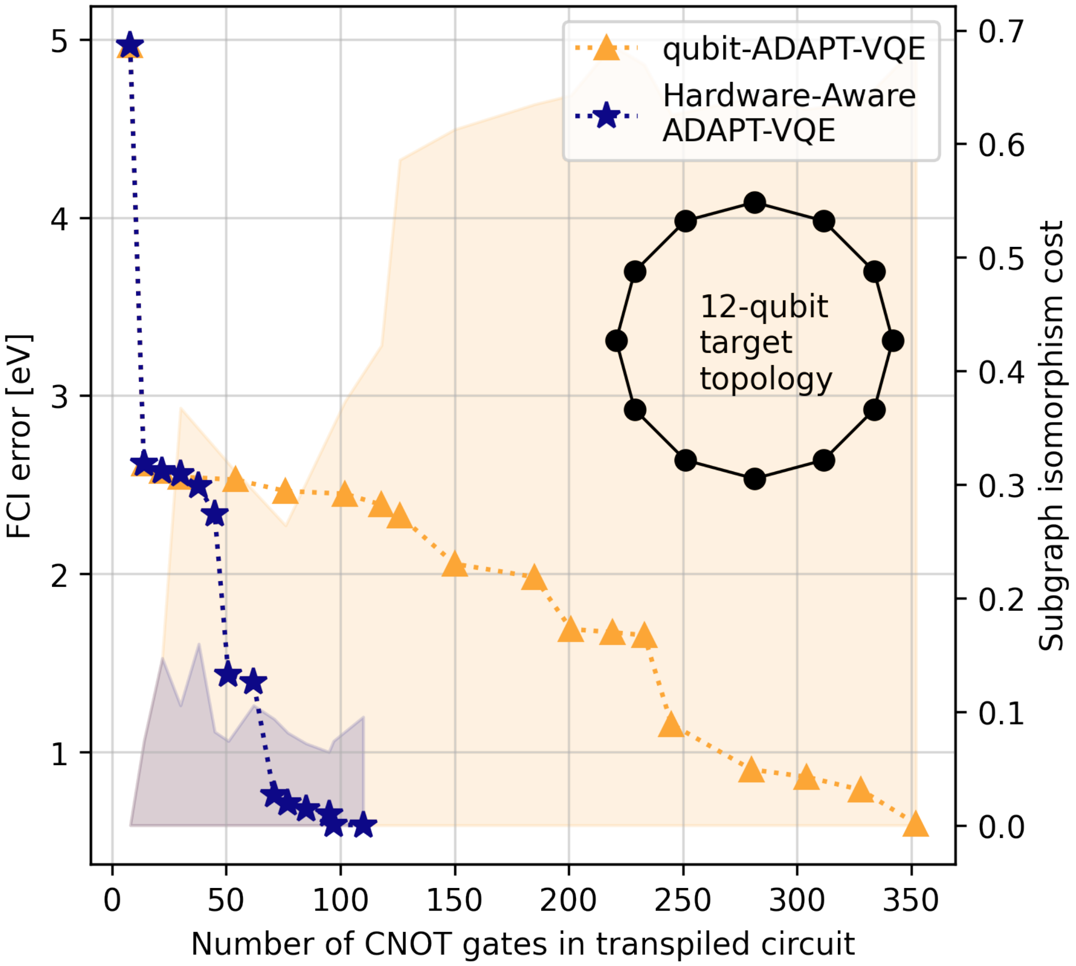


Figure 1.14: Construction of a 12-qubit contextual subspace ansatz for N_2 at 2Å. We show the FCI error per ADAPT cycle against the number of CNOT operations in the corresponding circuits transpiled for a closed loop of the 27-qubit *Falcon* topology as in Figure 4.9. We compare the standard qubit-ADAPT-VQE algorithm versus our hardware-aware approach and observe considerably reduced depths. The ‘subgraph isomorphism cost’ of embedding the ansatz graph in the target is computed as $s(\boldsymbol{n})/W$.

Chapter 2

Exploiting Symmetry for Resource Reduction

> It is only slightly overstating the case to say that physics is the study of symmetry.
>
> *Philip W. Anderson* [135]

The symmetries of a system serve as proxy through which we may probe many of its features. In many-body physics, symmetries manifest as properties such as particle number, total spin, spin parity, or topological characteristics. Oftentimes, the presence of symmetry presents an opportunity for simplification in some sense, effected by projecting the system onto a space in which its symmetries are inherently conserved. Unfortunately, constructing such a projection is not always easy; however, there are restricted settings in which one can do so in a scalable manner. Mathematically, a symmetry of the Hamiltonian H is defined to be an observable S satisfying $[H, S] = 0$; one may alternatively view this as the subset of observables that are constants of motion under the Hamiltonian evolution, governed by Heisenberg's equation of motion $\frac{d}{dt}S(t) = i[H, S]$. A special-case of symmetry arises when S is a Pauli operator, i.e. $S \in \mathcal{P}_N$, which we refer to here as $\mathbb{Z}_2$-type symmetry. Symmetries of this type have the consequence that that they simulataneously symmetric with each of the constituent Hamiltonian terms when decomposed over the Pauli group $H = \sum_{\boldsymbol{\sigma}} h_{\boldsymbol{\sigma}} \boldsymbol{\sigma}$. This may be seen as

$$H = S^2 H = SHS = \sum_{\boldsymbol{\sigma} \in \mathcal{P}} h_{\boldsymbol{\sigma}} S \boldsymbol{\sigma} S = \sum_{\boldsymbol{\sigma} \in \mathcal{P}} h_{\boldsymbol{\sigma}} (-1)^{1-\delta_{[S,\boldsymbol{\sigma}],0}} \boldsymbol{\sigma}. \tag{2.1}$$

Therefore, we must have $h_{\boldsymbol{\sigma}} = h_{\boldsymbol{\sigma}}(-1)^{1-\delta_{[S,\boldsymbol{\sigma}],0}}$, implying $\delta_{[S,\boldsymbol{\sigma}],0} = 1$ and thus $[S, \boldsymbol{\sigma}] = 0$ for each Pauli term $\boldsymbol{\sigma}$ in the Hamiltonian.

The $\mathbb{Z}_2$-type symmetries possess some useful properties that we will later show facilitates a qubit reduction, thus bringing otherwise computationally-intractable problems into reach of near-term quantum technological advances.

2.1 The Symplectic Representation

From an implementation point of view, it is beneficial to represent Pauli operators in the so-called *symplectic representation*. In this framework, operations over the Pauli group reduce to the manipulation of bitstrings, thus facilitating efficient computation via standard binary logic. Furthermore, we can make use of vectorization with Single Instruction, Multiple Sata (SIMD) processing to optimize performance. The `symmer` python package [10], developed to facilitate the work presented in this thesis, is constructed on top of this symplectic foundation.

Here, we identify an N-fold Pauli string $P \in \mathcal{P}_N$ with a pair of N-dimensional binary vectors $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{Z}_2^N$, whose elements are given by

$$x_n = \begin{cases} 1, & P_n \in \{X, Y\} \\ 0, & \text{otherwise} \end{cases}, z_n = \begin{cases} 1, & P_n \in \{Z, Y\} \\ 0, & \text{otherwise} \end{cases}. \tag{2.2}$$

Thus, defining $\boldsymbol{b} \coloneqq \boldsymbol{x}|\boldsymbol{z} \in \mathbb{Z}_2^{2N}$ (where | denotes vector/matrix concatenation) together with the map

$$\begin{aligned} \sigma : \mathbb{Z}_2^{2N} &\to \mathcal{P}_N; \\ \boldsymbol{b} &\mapsto i^{\boldsymbol{x}\cdot\boldsymbol{z}} \bigotimes_{n=0}^{N-1} \Big(X^{x_n} Z^{z_n} \Big), \end{aligned} \tag{2.3}$$

we may reconstruct our Pauli operator $P = \sigma(\boldsymbol{b})$. In other words, the binary vectors $\boldsymbol{x}, \boldsymbol{z}$ indicate tensor factors in which there is a Pauli X, Z respectively, with the additional factor $i^{\boldsymbol{x}\cdot\boldsymbol{z}}$ correcting for any incurred phases from the multiplication $XZ = -iY$.

In this picture, a linear combination $C = \sum_{t=1}^{T} c_t P_t$ of N-qubit Pauli terms $P_t \in \mathcal{P}_N$ may therefore be represented symplectically by a $2N \times T$ binary matrix

$$\boldsymbol{B} \coloneqq \begin{bmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_T \end{bmatrix} \tag{2.4}$$

supplemented by a vector of coefficients $\boldsymbol{c} \coloneqq (c_t)_{t=1}^{T} \in \mathbb{C}^T$. The operator C may be recovered as $C = \sum_{t=1}^{T} c_t \sigma(\boldsymbol{b}_t)$. It will also be useful to decompose the symplectic matrix into its $\boldsymbol{X}$- and $\boldsymbol{Z}$-blocks, where

$$\boldsymbol{X} \coloneqq \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_T \end{bmatrix}, \quad \boldsymbol{Z} \coloneqq \begin{bmatrix} \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_T \end{bmatrix}, \tag{2.5}$$

so that we may write

$$\boldsymbol{B} = \big[\boldsymbol{X} \mid \boldsymbol{Z}\big]. \tag{2.6}$$

In switching to the symplectic representation, many operations over the Pauli group may be parallelized and implemented in a classically-efficient manner - we shall now describe how to perform fast Pauli multiplications, check commutativity, identify $\mathbb{Z}_2$ symmetry generators and factorize a given Pauli operator over a supplied generating set. All of these operations are implemented in the `symmer` python package [10] for scalable manipulation of Pauli operators.

Example 2.1: Symplectic Representation from Pauli Hamiltonian

For illustrative purposes, we may write the 2-qubit Heisenberg XYZ Hamiltonian

$$H = -\frac{1}{2}\Big(J_X \cdot XX + J_Y \cdot YY + J_Z \cdot ZZ + h \cdot ZI + h \cdot IZ \Big), \tag{2.7}$$

where $J_X, J_Y, J_Z \in \mathbb{R}$ are coupling coefficients and $h \in \mathbb{R}$ the strength of an eternal magnetic field, in the symplectic representation:

$$\boldsymbol{B} \coloneqq \begin{bmatrix} \boldsymbol{b}_{XX} \\ \boldsymbol{b}_{YY} \\ \boldsymbol{b}_{ZZ} \\ \boldsymbol{b}_{ZI} \\ \boldsymbol{b}_{IZ} \end{bmatrix} = \left[\underbrace{\begin{matrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix}}_{\boldsymbol{X}} \middle| \underbrace{\begin{matrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{matrix}}_{\boldsymbol{Z}} \right], \boldsymbol{c} = -\frac{1}{2} \begin{bmatrix} J_X \\ J_Y \\ J_Z \\ h \\ h \end{bmatrix}. \tag{2.8}$$

We will revisit this example later.

2.1.1 Multiplication

Multiplication of Pauli operators reduces to binary vector addition in the symplectic representation, however care must be taken to ensure phases are correctly accounted for and is often omitted in much of the stabilizer code literature. Given Pauli operators $P, Q \in \mathcal{P}_N$, we may directly apply the map (2.3) to evaluate their product:

$$\begin{aligned} PQ &= \sigma(\boldsymbol{b}_P)\sigma(\boldsymbol{b}_Q) \\ &= i^{\boldsymbol{x}_P \cdot \boldsymbol{z}_P + \boldsymbol{x}_Q \cdot \boldsymbol{z}_Q} \bigotimes_{n=0}^{N-1} X^{x_{P,n}} Z^{z_{P,n}} X^{x_{Q,n}} Z^{z_{Q,n}} \\ &= i^{\boldsymbol{x}_P \cdot \boldsymbol{z}_P + \boldsymbol{x}_Q \cdot \boldsymbol{z}_Q} \bigotimes_{n=0}^{N-1} (-1)^{x_{Q,n} z_{P,n}} X^{x_{P,n}} X^{x_{Q,n}} Z^{z_{P,n}} Z^{z_{Q,n}} \quad \text{(exchanging inner } Z, X\text{)} \\ &= i^{\boldsymbol{x}_P \cdot \boldsymbol{z}_P + \boldsymbol{x}_Q \cdot \boldsymbol{z}_Q} (-1)^{\boldsymbol{x}_Q \cdot \boldsymbol{z}_P} \underbrace{\bigotimes_{n=0}^{N-1} X^{x_{P,n} + x_{Q,n}} Z^{z_{P,n} + z_{Q,n}}}_{-i^{(\boldsymbol{x}_P \oplus \boldsymbol{x}_Q) \cdot (\boldsymbol{z}_P \oplus \boldsymbol{z}_Q)} \sigma(\boldsymbol{b}_P \oplus \boldsymbol{b}_Q)} \\ &= i^{\boldsymbol{z}_P \cdot (\boldsymbol{x}_P \oplus \boldsymbol{x}_Q) + \boldsymbol{x}_Q \cdot (\boldsymbol{z}_P \oplus \boldsymbol{z}_Q)} (-i)^{(\boldsymbol{x}_P \oplus \boldsymbol{x}_Q) \cdot (\boldsymbol{z}_P \oplus \boldsymbol{z}_Q)} \sigma(\boldsymbol{b}_P \oplus \boldsymbol{b}_Q) \end{aligned} \tag{2.9}$$

2.1.2 Commutation

Given two Pauli operators $P, Q \in \mathcal{P}_N$ with corresponding symplectic vectors $\boldsymbol{b}_P, \boldsymbol{b}_Q \in \mathbb{Z}_2^{2N}$, define the *canonical symplectic form*

$$\Omega \coloneqq \begin{pmatrix} \mathbf{0}_N & \mathbb{1}_N \\ \mathbb{1}_N & \mathbf{0}_N \end{pmatrix} \tag{2.10}$$

and the *symplectic innner product*

$$\langle \boldsymbol{b}_P, \boldsymbol{b}_Q \rangle \coloneqq \boldsymbol{b}_P \Omega \boldsymbol{b}_Q^T = \boldsymbol{x}_P \cdot \boldsymbol{z}_Q + \boldsymbol{z}_P \cdot \boldsymbol{x}_Q. \tag{2.11}$$

Pauli operators commute when they differ on an even number of tensor factors (excluding identity positions); this corresponds with mismatches between the $\boldsymbol{X}$ and $\boldsymbol{Z}$ blocks of each operator. One may count these mismatches using the inner product (2.11), which yields a check for commutation:

$$[P, Q] = 0 \Leftrightarrow \langle \boldsymbol{b}_P, \boldsymbol{b}_Q \rangle \equiv 0 \mod 2. \tag{2.12}$$

Such operations are highly parallelizable in the symplectic representation. Here, for example, we may check commutation between each term of two linear combinations $L = \sum_{t=1}^{T_L} l_t P_t, M = \sum_{t=1}^{T_M} m_t Q_t$ by evaluating the inner product of their symplectic matrices

$$\langle \boldsymbol{B}_L, \boldsymbol{B}_M \rangle = \boldsymbol{X}_L \boldsymbol{Z}_M^\top + \boldsymbol{Z}_L \boldsymbol{X}_M^\top \mod 2. \tag{2.13}$$

The resulting matrix will be of size $T_L \times T_M$ and the (l, m)-th entry is zero (one) when $[P_l, Q_m] = 0$ ($\{P_l, Q_m\} = 0$).

2.1.3 Extracting Symmetry Generators

Unless known *a priori* due to some knowledge of the underlying physical system, identifying general symmetries of an arbitrary Hamiltonian H is usually infeasible. However, in the restricted setting of $\mathbb{Z}_2$-type symmetry as introduced above, one is able to fully characterize the set of such symmetries with surprisingly little computational overhead.

Define by $\mathcal{D}(H) \subset \mathcal{P}_N$ the set of $\mathbb{Z}_2$-type symmetries of H, which forms a group under Pauli multiplication; note this group may not be abelian, which we will later see is important in qubit tapering and leads to further implications regarding degeneracy of the eigenspectrum of H, discussed in Section 2.2. We shall then denote by $\varsigma(H)$ an independent generating set of $\mathcal{D}(H)$, namely $\mathcal{D}(H) = \overline{\varsigma(H)}$; this set is not unique, since multiplication of the set by any $\mathbb{Z}_2$ symmetry not already contained in $\varsigma(H)$ will produce a new independent generating set. We now demonstrate that it is possible to identify $\varsigma(H)$ in $\mathcal{O}(N^2)$ time.

As described in [136], representing an operator C symplectically yields an efficient method for determining symmetry generators, i.e. a set of operators $\mathcal{S}$ that generate the abelian subgroup $\overline{\mathcal{S}} \subset \mathcal{P}_N$ of Pauli operators that commute universally amongst the terms of C. Recalling the commutation equivalence (2.12), the symplectic representation $\boldsymbol{S}$ of $\mathcal{S}$ must therefore satisfy the equation $\boldsymbol{M}\Omega\boldsymbol{S}^\top = \mathbf{0}$. In other words, identification of $\boldsymbol{S}$ corresponds with characterisation of the kernel, or null space, of $\boldsymbol{M}\Omega = \big[\boldsymbol{Z} \mid \boldsymbol{X}\big]$.

This may be achieved by constructing the $(T+2N)\times 2N$ matrix $\left[\begin{array}{c}\boldsymbol{M}\Omega\\ \hline \mathbb{1}_{2N}\end{array}\right]$ and performing column-wise Gaussian elimination over the binary field to obtain a column-reduced echelon form $\left[\begin{array}{c}\boldsymbol{R}\\ \hline \boldsymbol{Q}\end{array}\right]$ (equivalently, row reduction of the transposed matrix). This implies the existence of an (invertible) matrix $\boldsymbol{P}$ such that

$$\left[\begin{array}{c}\boldsymbol{M}\Omega\\ \hline \mathbb{1}\end{array}\right]\boldsymbol{P} = \left[\begin{array}{c}\boldsymbol{R}\\ \hline \boldsymbol{Q}\end{array}\right] \tag{2.14}$$

and consequently $\boldsymbol{M}\Omega\boldsymbol{P} = \boldsymbol{R}, \boldsymbol{P} = \boldsymbol{Q} \implies \boldsymbol{M}\Omega\boldsymbol{Q} = \boldsymbol{R}$; we shall denote by $\mathcal{I}$ the set indexing zero-columns of $\boldsymbol{R}$. Taking $\boldsymbol{S}^\top = \left[\boldsymbol{Q}_{\mathcal{I}_0} \mid \dots \mid \boldsymbol{Q}_{\mathcal{I}_M}\right]$ the matrix formed from the columns of $\boldsymbol{Q}$ that correspond with zero columns in $\boldsymbol{R}$, we therefore have $\boldsymbol{M}\Omega\boldsymbol{S}^\top = \boldsymbol{0}$ and we are done.

Example 2.2: Extracting symmetry generators

We return to Example 2.1 and will identify a symmetry generator for the 2-qubit Heisenberg XYZ Hamiltonian. We first perform Gaussian row reduction to obtain a generating set for the Hamiltonian

$$\boldsymbol{M}\Omega = \left[\underbrace{\begin{matrix}0&0\\1&1\\1&1\\1&0\\0&1\end{matrix}}_{\boldsymbol{Z}}\middle|\underbrace{\begin{matrix}1&1\\1&1\\0&0\\0&0\\0&0\end{matrix}}_{\boldsymbol{X}}\right] \longrightarrow \left[\begin{array}{cc|cc}1&0&0&0\\0&1&0&0\\0&0&1&1\\0&0&0&0\\0&0&0&0\end{array}\right]. \tag{2.15}$$

Then, transposing (to save vertical space) and padding with the identity matrix as in Equation (2.14), this is followed by an additional row reduction (column reduction of the untraposed matrix)

$$\begin{aligned}\left[\Omega\boldsymbol{M}^\top \mid \mathbb{1}_4\right] &= \left[\begin{array}{ccccc|cccc}1&0&0&0&0&1&0&0&0\\0&1&0&0&0&0&1&0&0\\0&0&1&0&0&0&0&1&0\\0&0&1&0&0&0&0&0&1\end{array}\right]\\ &\rightarrow \left[\begin{array}{ccccc|cccc}1&0&0&0&0&1&0&0&0\\0&1&0&0&0&0&1&0&0\\0&0&1&0&0&0&0&0&1\\ \hline 0&0&0&0&0&\multicolumn{4}{c}{\underbrace{0\quad 0\quad 1\quad 1}_{\boldsymbol{S}}}\end{array}\right].\end{aligned} \tag{2.16}$$

Finally, we identify a single symmetry ZZ, which one may verify commutes with each of the Hamiltonian terms in (2.7).

2.1.4 Decomposition Over Generating Sets

Given an independent set of Pauli operators $\mathcal{B} \subset \mathcal{P}_N$, we might wish to determine which terms of an operator C may be reconstructed from products of elements in

$\mathcal{B}$; once again, this is possible in the symplectic formalism. We recall from Section 2.1.1 that multiplication corresponds with binary addition in this framework, so the problem we wish to solve here is to identify a $T \times |\mathcal{B}|$ reconstruction matrix $\boldsymbol{R}$ such that $\boldsymbol{M} = \boldsymbol{R}\boldsymbol{B}$ where $\boldsymbol{B}$ is the symplectic representation of $\mathcal{B}$. Note that $\boldsymbol{B}$ is invertible since its rows are linearly independent.

The interpretation of the reconstruction matrix is that, denoting $\mathcal{I}_i$ the indices of non-zero entries in the i-th row of $\boldsymbol{R}$, we will have $P_i = \prod_{j \in \mathcal{I}_i} B_j$. This assumes the set $\mathcal{B}$ is sufficiently expressible so as to reconstruct the term, however we will demonstrate how our method produces a flag that indicates whether a term of C has been successfully reconstructed from $\mathcal{B}$.

Forming the $(|\mathcal{B}|+T) \times 2N$ matrix $\left[\begin{array}{c} \boldsymbol{B} \\ \hline \boldsymbol{M} \end{array}\right]$ we identify a column-reduced echelon form $\left[\begin{array}{c} \boldsymbol{C} \\ \hline \boldsymbol{D} \end{array}\right]$, related by the invertible matrix $\boldsymbol{P}$ such that

$$\left[\begin{array}{c} \boldsymbol{B} \\ \hline \boldsymbol{M} \end{array}\right] \boldsymbol{P} = \left[\begin{array}{c} \boldsymbol{C} \\ \hline \boldsymbol{D} \end{array}\right], \tag{2.17}$$

just as we did in (2.14). This implies $\boldsymbol{B}\boldsymbol{P} = \boldsymbol{C}$ and $\boldsymbol{M}\boldsymbol{P} = \boldsymbol{D}$; using the invertibility of $\boldsymbol{B}$ and $\boldsymbol{P}$, we obtain $\boldsymbol{M}\boldsymbol{B}^{-1} = \boldsymbol{D}\boldsymbol{C}^{-1} = \boldsymbol{R}$. Because $\boldsymbol{B}$ is independent, we are able to write $\boldsymbol{C} = \left[\mathbb{1}_{|\mathcal{B}|} \mid \boldsymbol{0}_{|\mathcal{B}| \times (2N - |\mathcal{B}|)}\right]$ and consequently $\boldsymbol{D} = \left[\boldsymbol{R} \mid \boldsymbol{F}\right]$ where $\boldsymbol{F}$ is a $T \times (2N - |\mathcal{B}|)$ matrix whose rows are zero when the corresponding term of C may be reconstructed in the supplied generating set. Non-zero entries in $\boldsymbol{F}$ indicate that $\boldsymbol{B}$ is not sufficiently expressible to fully represent the operator.

Example 2.3: 2-qubit XYZ Heisenberg Hamiltonian Reconstruction

We shall use the above algorithm to demonstrate that $\mathcal{B} = \{XX, XY, YX\}$ forms a generating set of the 2-qubit XYZ Heisenberg Hamiltonian (2.7).

$$\begin{aligned} \left[\boldsymbol{B}^\top \mid \boldsymbol{M}^\top\right] &= \left[\begin{array}{ccc|ccccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \end{array}\right] \\ &\to \left[\begin{array}{ccc|ccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{0} & \boldsymbol{F}^\top \\ \hline \mathbb{1}_{|\mathcal{B}|} & \boldsymbol{R}^\top \end{array}\right]. \end{aligned} \tag{2.18}$$

The fact that $\boldsymbol{F}$ contains only zero entries indicates the reconstruction was successful and observe that $\boldsymbol{R}$ indexes the elements of $\mathcal{B}$ required to produce each term of the Hamiltonian up to a phase:

$$\begin{gathered} XX = B_0, \quad YY = B_0 B_1 B_2, \quad ZZ = B_1 B_2, \\ ZI = B_0 B_2, \quad IZ = B_0 B_1. \end{gathered} \tag{2.19}$$

2.2 Link Between Symmetry and Degeneracy

It should be highlighted that it is possible for elements of $\varsigma(H)$ to anticommute amongst themselves, despite forming a generating set of Pauli operators that commute with the Hamiltonian. The presence of anticommuting symmetry generators indicates degeneracy of the energy spectrum, which we shall prove via the equivalent statement that a non-degenerate spectrum implies a fully commuting symmetry generating set.

First of all, we shall assume that $|\psi\rangle \neq \mathbf{0}$ is the unique eigenvector (up to a scaling) satisfying $H|\psi\rangle = \epsilon|\psi\rangle$ and that we have two symmetry elements A, B, i.e. $[H, A] = 0, [H, B] = 0$, with no assumptions of commutativity made between them. We therefore have that $A|\psi\rangle$ and $B|\psi\rangle$ are also eigenvectors of H with eigenvalue ϵ, but since the eigenspace is non-degenerate it must be the case that $A|\psi\rangle = a|\psi\rangle, B|\psi\rangle = b|\psi\rangle$ for scalars $a, b \in \mathbb{C}$; we may assume without loss of generality that $a = b = 1$, since we may redefine $A \mapsto \frac{1}{a}A, B \mapsto \frac{1}{b}B$ and we have $A|\psi\rangle = B|\psi\rangle = |\psi\rangle$. We shall now assume for contradiction that $\{A, B\} = 0$. This implies

$$|\psi\rangle = AB|\psi\rangle = -BA|\psi\rangle = -|\psi\rangle \tag{2.20}$$

and therefore $|\psi\rangle = \mathbf{0}$, contrary to our initial assumptions on $|\psi\rangle$; therefore, A and B cannot anticommute, and if they are $\mathbb{Z}_2$-symmetries then it must be the case that $[A, B] = 0$ (since Pauli operators either commute or anticommute).

We note this result says nothing of the degeneracy for a particular eigenvalue in the case of a fully commuting symmetry generating set. For example, consider the Hamiltonian $H = (XX - YY + ZZ - II)/2$ with symmetry generators $\mathcal{S} = \{XX, ZZ\}$; this system has a degenerate ground state energy $\epsilon = -1$ with multiplicity 3 and corresponding eigenbasis $\{\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), |01\rangle, |10\rangle\}$. This arises as $ZI - IZ$ also commutes with the Hamiltonian, but not with the symmetry generating element XX.

In short, $\mathbb{Z}_2$-symmetries permute the eigenbasis, whereas it is possible for non-$\mathbb{Z}_2$-symmetries to map eigenstates onto the zero vector, as above. This is not possible for $\mathbb{Z}_2$-symmetries as $S|\psi\rangle = \mathbf{0} \implies |\psi\rangle = S^2|\psi\rangle = \mathbf{0}$. Ultimately, degeneracy of the eigenspectrum implies the presence of Hamiltonian symmetry, while the converse is not in general true, since symmetries may coincidentally map eigenstates onto themselves. However, we have shown the case of anticommutation amongst symmetry elements, meaning the symmetry group is non-abelian, guarantees the existence of degenerate eigenvalues – although this observation does not provide a constructive way of identifying such degeneracies.

2.3 Stabilizer Subspace Projection

Let $\mathcal{G} = \varsigma(H)$ be a generating set of $\mathbb{Z}_2$-type symmetries of the Hamiltonian H. Now, recall the Clifford group consists of unitary operators $U \in \mathscr{B}(\mathscr{H})$ (meaning $UU^\dagger = U^\dagger U = \mathbb{1}$) with the property $U\boldsymbol{\sigma}U^\dagger \in \mathcal{P}_N \; \forall \boldsymbol{\sigma} \in \mathcal{P}_N$, i.e., U *normalizes* the Pauli group. We may construct a Clifford operation U mapping each symmetry generator to distinct single-qubit Pauli operators σ_p, where we are free to choose $p \in \{1, 2, 3\}$.

Example 2.4: Anticommuting symmetries

In the absence of an external field, the Heisenberg Hamiltonian model has the form

$$H = \sum_{<i,j>} \left(J_x X^{(i)} X^{(i)} + J_x Y^{(i)} Y^{(i)} + J_z Z^{(i)} Z^{(i)} \right) \tag{2.21}$$

where $i, j \in \{1, \dots, N\}$ are pairs of indices running over the sites of a spin lattice. Such Hamiltonians on an odd number of spins ($N = 2n+1$) exhibit anticommuting symmetry generating sets of the form $\{Z^{\otimes N}, X^{\otimes N}\}$, meaning every single eigenvalue will be degenerate (following the discussion of Section 2.2). Tapering may lift some degeneracy, which might have previously caused issues for optimization etc.

Consider for example a 3-spin lattice, corresponding with the field-free XYZ-Hamiltonian

$$\begin{aligned} H = {} & J_x(XXI + IXX + XIX) \\ & + J_y(YYI + IYY + YIY) \\ & + J_z(ZZI + IZZ + ZIZ) \end{aligned} \tag{2.22}$$

with anticommuting $\mathbb{Z}_2$-symmetries XXX, ZZZ. The eigenvalues are

$$\begin{aligned} \epsilon_0 = {} & -(J_x + J_y + J_z) \\ \epsilon_1 = {} & -\epsilon_0 + \sqrt{2}\sqrt{(J_x - J_y)^2 + (J_x - J_z)^2 + (J_y - J_z)^2} \\ \epsilon_2 = {} & -\epsilon_0 - \sqrt{2}\sqrt{(J_x - J_y)^2 + (J_x - J_z)^2 + (J_y - J_z)^2} \end{aligned} \tag{2.23}$$

with multiplicities $4, 2, 2$, respectively.

Tapering out the XXX symmetry, we obtain a reduced 2-qubit Hamiltonian $H' = J_x A - \frac{J_y}{2}\{A, B\} + J_z B$ where $A = XX + XI + IX$ and $B = ZZ + ZI + IZ$, whose eigenvalues coincide with those above but now with multiplicities $2, 1, 1$, respectively. The fact there is still a degenerate eigenvalue indicates the existence of some non-$\mathbb{Z}_2$ symmetry, in this case the operator $XX + YY + ZZ$, as discussed at the end of Section 2.2.

More precisely, there exists a subset of qubit positions $\mathcal{I}_{\text{stab}} \subset \mathbb{Z}_N$ satisfying $|\mathcal{I}_{\text{stab}}| = |\mathcal{G}|$ and a bijective map $f : \mathcal{G} \to \mathcal{I}_{\text{stab}}$ such that

$$UGU^\dagger = \sigma_p^{(f(G))} \;\; \forall\, G \in \mathcal{G}. \tag{2.24}$$

This is a powerful concept that provides a mechanism for reducing the number of qubits in the Hamiltonian whilst preserving its energy spectrum. This is at the core of qubit tapering [136, 137], in which it is observed that

$$[G, H_{\mathcal{T}}] = 0 \implies [\sigma_p^{(f(G))}, H'_{\mathcal{T}}] = 0 \;\; \forall\, G \in \mathcal{G}, \tag{2.25}$$

implying the rotated Hamiltonian $H'_{\mathcal{T}} := UH_{\mathcal{T}}U^\dagger$ consists solely of identity or Pauli σ_p operators in the qubit positions indexed by $\mathcal{I}_{\text{stab}}$. Taking expectation values, one may replace the qubits $\mathcal{I}_{\text{stab}}$ by their eigenvalues $\nu_i = \pm 1$; each assignment

$$\boldsymbol{\nu} = (\nu_i)_{i \in \mathcal{I}_{\text{stab}}} \in \{\pm 1\}^{\times \mathcal{I}_{\text{stab}}} \tag{2.26}$$

defines a symmetry *sector* and at least one such sector will contain the true solution to the eigenvalue problem. Note the other sectors still have physical significance and may for example relate to solutions with different particle numbers or to excited states.

A quantum state consistent with any such sector must be stabilized by the operators $\nu_i \sigma_p^{(i)}$ and we may define a projection onto the corresponding stabilizer subspace. In general, a projection is defined to be an *idempotent* operator $\mathbb{P} \in \mathscr{B}(\mathscr{H})$, i.e. $\mathbb{P}^2 = \mathbb{P}$; the projection onto the ±1-eigenspace of a single-qubit Pauli operator σ_p for $p \in \{1,2,3\}$ may be written

$$\mathbb{P}_p^{\pm} := \frac{1}{2}\big(I \pm \sigma_p\big). \tag{2.27}$$

States with no component inside the chosen eigenspace are mapped to zero and observe that

$$\mathbb{P}_p^{\pm} \sigma_q \mathbb{P}_p^{\pm} = \pm\delta_{p,q} \mathbb{P}_p^{\pm} \tag{2.28}$$

for $q \in \{1,2,3\}$.

Let $\mathscr{H}_{\text{stab}}$ be the reduced Hilbert space supported by the stabilized qubits $\mathcal{I}_{\text{stab}}$ and $\mathscr{H}_{\text{red}}$ its complement such that $\mathscr{H} = \mathscr{H}_{\text{stab}} \otimes \mathscr{H}_{\text{red}}$. Given an assignment of eigenvalues $\boldsymbol{\nu} \in \{\pm 1\}^{\times \mathcal{I}_{\text{stab}}}$, we may project onto the corresponding sector via

$$\mathbb{P}_{\boldsymbol{\nu}} := \bigotimes_{i \in \mathcal{I}_{\text{stab}}} \mathbb{P}_p^{\nu_i} \tag{2.29}$$

and subsequently perform a *partial trace* over the stabilized qubits $\mathcal{I}_{\text{stab}}$. This is effected by the unique linear map $\text{Tr}_{\text{stab}} : \mathscr{H} \to \mathscr{H}_{\text{red}}$ satisfying the property $\text{Tr}_{\text{stab}}\big(A \otimes B\big) = \text{Tr}\big(A\big)B$ for all $A \in \mathscr{B}(\mathscr{H}_{\text{stab}})$ and $B \in \mathscr{B}(\mathscr{H}_{\text{red}})$.

Finally, we may define the full stabilizer subspace projection map

$$\begin{aligned} \pi_{\boldsymbol{\nu}}^U : \mathscr{B}(\mathscr{H}) &\to \mathscr{B}(\mathscr{H}_{\text{red}}); \\ A &\mapsto \text{Tr}_{\text{stab}}\big(\mathbb{P}_{\boldsymbol{\nu}} U A U^\dagger \mathbb{P}_{\boldsymbol{\nu}}\big) \end{aligned} \tag{2.30}$$

which, using the linearity of Tr_{stab}, yields a reduced Hamiltonian

$$\begin{aligned} H_{\mathcal{T}}^{\text{red}} &:= \pi_{\boldsymbol{\nu}}^U(H_{\mathcal{T}}) \\ &= \sum_{\boldsymbol{\sigma} \in \mathcal{T}} h'_{\boldsymbol{\sigma}} \boldsymbol{\sigma}'_{\text{red}} \end{aligned} \tag{2.31}$$

where $\boldsymbol{\sigma}' = U\boldsymbol{\sigma}U^\dagger = \bigotimes_{i=0}^{N-1} \sigma_{q_i}$ and we have written $\boldsymbol{\sigma}' = \boldsymbol{\sigma}'_{\text{stab}} \otimes \boldsymbol{\sigma}'_{\text{red}}$. The new coefficients $h'_{\boldsymbol{\sigma}} := h_{\boldsymbol{\sigma}} \prod_{\substack{i \in \mathcal{I}_{\text{stab}} \\ q_i \neq 0}} \nu_i$ differ from $h_{\boldsymbol{\sigma}}$ by a sign dependent on the chosen symmetry sector.

In qubit tapering U is taken as (2.24), with the corresponding basis $\mathcal{G}$ a generating set for a full Hamiltonian symmetry [136, 137]. Assuming identification of the correct sector, the ground state energy of the $(N - |\mathcal{G}|)$-qubit reduced Hamiltonian $H_{\mathcal{T}}^{\text{red}}$ will coincide with the true value of the full system $H_{\mathcal{T}}$.

This stabilizer projection procedure is straightforward with respect to the Hamiltonian, since the stabilized qubits contain only operators with non-zero image under conjugation with $\mathbb{P}_{\boldsymbol{\nu}}$. However, suppose we were to take another observable $A \in \mathscr{B}(\mathscr{H})$ and wish to determine a reduced form on $\mathscr{B}(\mathscr{H}_{\text{red}})$ that is consistent with the reduced Hamiltonian $H_{\mathcal{T}}^{\text{red}}$. This may be achieved by following precisely the same process that was applied to $H_{\mathcal{T}}$, but the symmetry $\mathcal{S}$ will not in general be a symmetry of A and therefore the 'symmetry-breaking' terms (those which anticommute with the generators $\mathcal{G}$) will vanish under projection onto the stabilizer subspace, as per (2.28). Letting $\mathcal{A} \subset \mathcal{P}_N$ be the set of terms in the Pauli-basis expansion of A, observe that

$$\begin{aligned}
A^{\text{red}} &:= \pi^{U}_{\boldsymbol{\nu}}(A) \\
&= \sum_{\boldsymbol{\sigma}\in\mathcal{A}} h_{\boldsymbol{\sigma}} \operatorname{Tr}(\mathbb{P}_{\boldsymbol{\nu}} \boldsymbol{\sigma}'_{\text{stab}} \mathbb{P}_{\boldsymbol{\nu}}) \boldsymbol{\sigma}'_{\text{red}} \\
&= \sum_{\boldsymbol{\sigma}\in\mathcal{A}} h_{\boldsymbol{\sigma}} \operatorname{Tr}\Big(\bigotimes_{\substack{i\in\mathcal{I}_{\text{stab}} \\ q_i=0}} \mathbb{P}_p^{\nu_i} \otimes \bigotimes_{\substack{i\in\mathcal{I}_{\text{stab}} \\ q_i\neq0}} \underbrace{\mathbb{P}_p^{\nu_i} \sigma_{q_i} \mathbb{P}_p^{\nu_i}}_{=\nu_i\delta_{p,q_i}\mathbb{P}_p^{\nu_i}} \Big) \boldsymbol{\sigma}'_{\text{red}} \\
&= \sum_{\boldsymbol{\sigma}\in\mathcal{A}} h_{\boldsymbol{\sigma}} \boldsymbol{\sigma}'_{\text{red}} \underbrace{\operatorname{Tr}(\mathbb{P}_{\boldsymbol{\nu}})}_{=1} \prod_{\substack{i\in\mathcal{I}_{\text{stab}} \\ q_i\neq0}} \nu_i \delta_{p,q_i} \\
&= \sum_{\substack{\boldsymbol{\sigma}\in\mathcal{A} \\ q_i\in\{0,p\} \\ \forall\, i\in\mathcal{I}_{\text{stab}}}} h'_{\boldsymbol{\sigma}} \boldsymbol{\sigma}'_{\text{red}}.
\end{aligned} \tag{2.32}$$

The resulting form is identical to (2.31), except we are explicit that the terms surviving projection are only those whose qubit positions indexed by $\mathcal{I}_{\text{stab}}$ consist exclusively of identity and Pauli σ_p operators; this is trivially true for the Hamiltonian by construction.

Example 2.5: Tapering Diatomic Molecules

All molecular systems possess at least two $\mathbb{Z}_2$-symmetries which enforce the parity of spin up or down particles. Additional symmetry arises from the molecular point group that describes the geometrical symmetry of the system. In the setting of diatomic molecules there are two relevant point groups: the cyclic point group $C_{\infty v}$ consisting of continuous rotations around the inter-nuclear axis and the dihedral point group $D_{\infty h}$ that also includes the reflection and inversion symmetries of the diatomic. Heteronuclear molecules consisting of two distinct atomic centres, such as HCl, lie within the former group, while homonuclear molecules such as N_2 fall under the latter.

The specific $\mathbb{Z}_2$-symmetries one exploits through tapering come from abelian subgroups of the above point groups that describe a restriction to 2-fold symmetry. In particular, the relevant group generators of $C_{2v} \subset C_{\infty v}$ are 180° rotations around the bond axis, denoted C_2, and vertical reflections σ_v. In the case of $D_{2h} \subset D_{\infty h}$ we have the same rotational symmetry C_2, in addition to the group generators σ_h, corresponding with horizontal reflections, and the inversion symmetry i. In all, qubit tapering enables the removal of four qubits from heteronuclear molecules (two point group generators C_2, σ_v plus spin up/down parity) and five from homonuclear molecules (three point group generators C_2, σ_h, i plus spin up/down parity).

Most importantly, this extends the stabilizer subspace projection to ansätze defined on the full system for use in variational algorithms. It should be noted that the above operations are classically tractable and can be implemented efficiently in the symplectic representation of Pauli operators [138, 139].

We would be remiss not to draw attention to the likeness of (2.30) with Positive Operator-Valued Measures (POVM) [140]; indeed, the projectors (2.29) define a complete set of *Kraus* operators [141]. The stabilizer subspace projection procedure is reduced to a matter of enforcing a partial measurement over some subsystem of the full problem, for which the relevant outcomes have been determined via an auxiliary method. We discuss this in the following Section 2.4.

2.4 Sector Identification via Reference States

Given a generating set of $\mathbb{Z}_2$ symmetries $\mathcal{S}$ (these might be derived from physical symmetries as in qubit tapering, or pseudo-symmetries in the contextual subspace approach), it is key that one identifies the relevant symmetry sector. This is defined by a vector of eigenvalue assignments $\boldsymbol{\nu} \in \{\pm 1\}^{\times|\mathcal{S}|}$ to the element of $\mathcal{S}$; the states of the subspace specified by this symmetry sector are those subsequently stabilized by the operators $\{\nu_S S | S \in \mathcal{S}\}$. This defines a stabilizer subspace. We remark that the linear algebra interpretation of this is that, for a given choice of $\boldsymbol{\nu}$, this specifies the eigenstate of each element of $\mathcal{S}$ and thus specifies a given block in the Hamiltonian. This is because $[H, S] = 0 \;\; \forall S \in \mathcal{S}$ and thus H is block diagonal with respect to the eigenstates of each symmetry operator in $\mathcal{S}$ as the eigenstates of each S must be a joint eigenstate of H.

In qubit tapering, $\boldsymbol{\nu}$ must be chosen such that the desired eigenstate lies within a given subspace. Whereas, in the contextual subspace approach $\boldsymbol{\nu}$ is selected such that it minimizes the ground state of a noncontextual Hamiltonian. In each case finding $\boldsymbol{\nu}$ is non-trivial and was not addressed in the previous literature [136, 137]. The number of eigenvalue assignment combinations is exponential in the number of stabilizers being imposed over the target Hamiltonian, namely $2^{|\mathcal{S}|}$. It should be noted that the sector identification problem is in general NP-complete [142, 143]. For example, given a diagonal Hamiltonian the corresponding symmetry generators form a generating set of the Hamiltonian itself; as such, sector assignment is equivalent to fully solving the problem in this case. However, we will show next that for some problems it is possible to choose the correct symmetry sector based on knowledge of the underlying problem.

Given an eigenstate $|\psi\rangle$ of H with corresponding $\mathbb{Z}_2$ symmetries $\mathcal{S} = \varsigma(H)$, we formalize the assignment of $\boldsymbol{\nu} = (\nu_S)_{S \in \mathcal{S}}$ with $\nu_S = \langle\psi| S |\psi\rangle$. Note that, by definition, $|\psi\rangle$ is an eigenstate of each $S \in \mathcal{S}$ since commuting operators share a common eigenbasis. However, we will not in general know the eigenstate $|\psi\rangle$ *a priori* as this requires solving the problem. Instead, if one has knowledge of a reference state $|\phi\rangle$, the problem of sector identification can be solved subject to non-zero overlap with the desired eigenstate, $|\langle\psi|\phi\rangle| > 0$.

There are two cases to consider when using a reference state $|\phi\rangle$. First, when $|\phi\rangle$ is an joint eigenstate of the symmetry operators in addition to having non-zero overlap with the desired eigenstate, one simply has to 'measure' the symmetry elements in the reference state to determine $\boldsymbol{\nu} = (\langle\phi| S |\phi\rangle)_{S \in \mathcal{S}}$. This leads to no ambiguity in the sector choice, which is the ideal scenario. Note if $| \langle\psi|\phi\rangle | = 0$ then the wrong sector will be selected, as the wrong joint eigenstate has been used to choose the sector.

Next, consider the case when $|\phi\rangle$ has non-zero overlap with $|\psi\rangle$ but is not an eigenstate of at least one symmetry operator $S \in \mathcal{S}$. In such a situation, whereby $|\phi\rangle$ spans both ±1-eigenspaces of S, we assign the dominant eigenvalue if it exceeds some predefined threshold $0 < \delta \leq 1$:

$$\nu_S = \begin{cases} +1, & \langle\phi| S |\phi\rangle \geq \delta \\ -1, & \langle\phi| S |\phi\rangle \leq -\delta \\ 0, & \text{otherwise} \end{cases}, \tag{2.33}$$

where $\delta = 1$ forces the exact-eigenstate case introduced previously. Any zero assignments will then need to be determined via alternative means, either via brute force search or some other optimization scheme if the reference state is not sufficiently weighted towards either ±1 eigenvalue. Note for $\delta < 1$ there is some ambiguity in the sector choice and care must be taken as the wrong sector can be selected in this heuristic approach even with $| \langle\psi|\phi\rangle | > 0$.

Given that we may decompose any Pauli operator S over its ±1-eigenspace projectors $\mathbb{P}_\pm = \frac{1}{2}(I \pm S)$ (introduced in the previous Section 2.3) as $S = \mathbb{P}_+ - \mathbb{P}_-$, it is possible to rewrite the above expectation value

$$\begin{aligned} \langle\phi| S |\phi\rangle &= \|\mathbb{P}_+\phi\|^2 - \|\mathbb{P}_-\phi\|^2 \\ &= \frac{1}{4}\Big[\|\phi + S\phi\|^2 - \|\phi - S\phi\|^2\Big], \end{aligned} \tag{2.34}$$

where $\|\phi\| = \sqrt{| \langle\phi|\phi\rangle |}$ and we have used the fact $\mathbb{P}_\pm^2 = \mathbb{P}_\pm$. The significance of this is that direct evaluation of the left-hand side involves multiplication over $\mathcal{O}(M^2)$ cross-terms of ψ (defined over M basis states), whereas the right can be implemented in $\mathcal{O}(M)$ operations (evaluating the product $S\psi$) and therefore is more suitable for efficient implementation.

The sector assignment problem is now reduced to the generation of a suitable reference state that exhibits sufficient overlap with the desired eigenstate. We re-iterate that the best scenario is if the reference state is an eigenstate of every symmetry operators and has any non-zero overlap with the target eigenstate, as this will result in always the correct subspace being selected. For many quantum chemistry applications, the Hartree-Fock state provides such a reference for the ground space. However, there are scenarios in which it may fail, for example if the true ground state is fully excited away from the Hartree-Fock reference. In such cases, one may refer to a hierarchy of increasingly precise classical quantum chemistry methods such as Møller–Plesset perturbation theory (MP2, MP3, ...) [144, 145], Coupled-Cluster theory (CCD, CCSD, ...) [49], Configuration Interaction (CID, CISD,...) [146] etc. in the hope that a sufficient overlap with the ground is obtained at some point, suggested in [2]. The sector selection may then be motivated using the resulting wavefunctions. We remark that if these these methods produce reference states that select different symmetry sectors (equation 2.33) compared to those predicted by Hartree-Fock, then there is strong evidence to suggest Hartree-Fock is insufficient for selecting the correct symmetry sector.

Example 2.6: Charged molecules

Direct diagonalization of the full second quantized Hamiltonian for charged molecules such as HeH^+ or ${H_3}^+$ is prone to identifying the incorrect subspace (the neutral species) if the eigenstate with the lowest eigenvalue is taken. However, if we first taper the problem using the Hartree-Fock state to determine the symmetry sector (Equation (2.33)) we find that direct diagonalization now results in the lowest eigenvector being the correct ground state. The reason for this is that two of the $\mathbb{Z}_2$ symmetries tapered out in such problems are the spin-up and spin-down parity operators. By fixing the assignment of these symmetries it stops the neutral species from being obtained as the minimum eigenvalue.

Importantly, we note that if an ansatz circuit is used, such as when using the Variational Quantum Eigensolver (VQE) algorithm, assuming the ansatz circuit is number and spin-parity preserving, the trial state will always remain in the correct sector/subspace. Such an example is the Unitary Coupled Cluster ansatz [22], introduced in Section 1.4.3, or those constructed from single-particle Givens rotations of the form in Figure (1.12).

On the other hand, if the symmetries are not accounted for by the ansatz then a state that is in the wrong symmetry sector could be obtained. In this example, if a fully expressive hardware efficient ansatz is used, such that any state in the Hilbert space could be obtained, then it is likely that the neutral species will again be obtained, rather than the charged system that we are aiming to simulate.

Example 2.7: H_4 molecule

One of the simplest molecules that exhibits a breakdown of the standard computational methods is H_4 arranged in a planar ring. In fact, this system has been used to benchmark different computational chemistry methods [147–150]. Interestingly, when described in the STO-3G basis, the ground state obtained via direct diagonalization has zero overlap with the Hartree-Fock solution due to a difference in symmetry sectors (B_{1g} vs A_g).The ground state energies of $H_{B_{1g}}$ and H_{A_g} are -1.8743 Ha and -1.8643 Ha respectively. Tapering with the Hartree-Fock state incorrectly chooses the A_g symmetry sector. For such cases, a user needs to be careful in choosing the desired symmetry sector.

If the Hartree-Fock input state was used in this case as input for the Quantum Phase Estimation algorithm [35, 151] it would be impossible to get the minimum eigenvalue of the molecular Hamiltonian. This example highlights that care must be taken even in the fault tolerant regime. The chosen reference state should always be selected such that it resides in the correct symmetry sector, or at the very least selected to span multiple sectors (the scenario when $\langle\phi| S |\phi\rangle \neq \pm 1$ in equation 2.33).

Chapter 3

Exploiting Contextuality for Resource Reduction

> What we observe is not Nature itself, but Nature exposed to our method of questioning.
>
> *Werner Heisenberg* [152]

A long-held belief in physics was one of *causal determinism*, that all bodies follow predetermined trajectories governed by a set of universal principles; it is the physicists's prerogative to discover these laws through acute observation of their surroundings. This was first articulated in 1814 by Laplace, who imagined an "intelligence which could comprehend all the forces by which Nature is animated... for it, nothing would be uncertain and the future, as the past, would be present to its eyes" [153, p.4]. Here, Laplace describes a Newtonian picture of Nature, in which the properties of a system are determined for any desired time by the position and momentum of its comprising particles. But, Laplace's oracle was later denounced a *demon*, rather than the intelligence he had first envisaged. Another demon followed five decades later when Maxwell presented a thought-experiment in 1867 that purportedly violated the second law of thermodynamics [154]. However, what Maxwell's demon really revealed was a subtle connection between energy and information.

Advance forward another half century and we see physics undergoing a period of turbulence in the fallout of the ultraviolet catastrophe, whereby classical electromagnetism predicted unbounded intensity of black body radiation [155, 156]. The ensuing decades saw a number of *ad hoc* corrections to the mechanics of the day in order to reconcile such troublesome results. However, the resulting theories were inconsistent and incomplete, serving more as heuristic appendages to an otherwise insufficient description of Nature. This "old quantum theory", as it came to be known, was finally settled with the Heisenberg interpretation of 1925 [157] and concurrently the wave mechanics formulation developed by Schrödinger [158]. While these theories were praised for their ingenuity, the following year bought with it a shocking revelation in the form of Born's rule [159], predicating that quantum mechanics must be understood probabilistically and therefore devoid of Laplace's causal determinism. This notion was met with adversity by many prominent physicists of the time, including none other than Einstein, and came to a head in 1927 at the prestigious Solvay conference where the pivotal Bohr-Einstein debates were sparked. The debates consisted of a series of thought-experiments designed to refute the other's point of view, but Bohr emerged triumphant, vindicating the Copenhagen interpretation of quantum physics and Heisenberg's uncertainty principle, that asserts a fundamental limit on the simulataneous knowledge of both position and momentum [16]. The following decade saw further rebuttal from Einstein, most notably the Einstein-Podolsky-Rosen (EPR) paradox of 1935 [160] which was left unresolved for three decades until Bell's theorem [161] that has since received extensive experimental validation.

The implication of Bell's inequalities is a deficiency of realistic hidden variable theories – the understanding that measurement uncovers some predetermined property of Nature that exists independent of the observer – to capture quantum mechanical behaviour. This is most concisely demonstrated by means of a simple game. It starts with a 3×3 grid in which players must enter ± 1 values in each cell, with the goal of satisfying the conditions that products along rows equal $+1$ and products down columns equal -1. It does not take long to arrive at the conclusion that this game cannot be won; an example of a possible configuration is presented in Table 3.1, where one finds it is not possible to fill the final cell such that both the row and column conditions are satisfied.

However, we may modify the game to instead grant players access to bipartite Pauli operators. Now, we find it actually is possible to satisfy the success criteria that rows multiply to $+\mathbb{1}$ and columns to $-\mathbb{1}$, owing to the richer algebraic structure of the Pauli group. One such satisfying configuration may be viewed in Table 3.2 and, while seemingly inconspicuous, hides an unsettling realisation. The entries $\{IZ, ZI, ZZ, XI, IX, XX, -XZ, -ZX, YY\}$ form a set of

	c_1	c_2	c_3	
r_1	$+1$	-1	-1	$+1$
r_2	-1	$+1$	-1	$+1$
r_3	$+1$	$+1$	?	×
	-1	-1	×	

Table 3.1: It is not possible to fill a 3×3 grid with ± 1 values such that products along rows yield $+1$ and products down columns yield -1.

	c_1	c_2	c_3	
r_1	IZ	ZI	ZZ	$+\mathbb{1}$
r_2	XI	IX	XX	$+\mathbb{1}$
r_3	$-XZ$	$-ZX$	YY	$+\mathbb{1}$
	$-\mathbb{1}$	$-\mathbb{1}$	$-\mathbb{1}$	

Table 3.2: While it was not possible to fill the grid in Table 3.1, relaxing the entries to bipartite Pauli operators we find it is possible to complete the square such that row products yield $+\mathbb{1}$ and column products yield $-\mathbb{1}$.

observables that one could feasibly measure [1] in the laboratory, thus revealing a set of ±1 eigenvalue assignments and effectively superimposing our Pauli grid onto the unsatisfiable grid above.

This presents a contradiction, noting that taking products *across* rows and columns yields

$$r_1 r_2 r_3 = +\mathbb{1} \neq -\mathbb{1} = c_1 c_2 c_3. \tag{3.1}$$

Newtonian mechanics is ill-equipped to resolve this observation; it may, however, be reconciled with the concept of measurement contextuality. In this setting, a *context* is defined to be a set of commuting observables, which each of the rows and columns in Table 3.2 forms. Commuting observables share a common eigenbasis, the consequence of which being it is possible to perform simultaneous measurements [163]. Contextuality of the type presented here is of the so-called *strong* form [164–166]. This is a restricted setting compared to contextuality in the Bell inequality sense [161, 167].

Since each row and column of Table 3.2 specifies a different context, the final measurement outcome is consequently *context-dependent*. Namely, the observer's choice to view a particular context has implications on the measured value. But this directly opposes Laplace's view that we are simple observers of the universe *in vacuo*; instead, quantum mechanics suggests that the two cannot be disentangled, that the observer and observee are inextricably linked.

3.1 Detecting Contextuality

Grids of the form in Table 3.2 are referred to as Peres-Mermin magic squares [168, 169]. Focussing on the upper left quadrant $\mathcal{T} = \{IZ, ZI, XI, IX\}$, the remaining operators in row r_3 and column c_3 are products of elements in $\mathcal{T}$, or in other words $\mathcal{T}$ is a generating set of the full grid; therefore, this reduced set should be sufficient to reproduce the context-dependence demonstrated above. It is instructive to consider the compatibility graph of $\mathcal{T}$, displayed in Figure 3.1, where we find that commutation is non-transitive, one example being $[IZ, ZI] = [ZI, IX] = 0$ but $[IZ, IX] \neq 0$. This property of Pauli measurements hints at the root cause of contextuality; to better understand this phenomenon we shall adopt a more rigorous mathematical language. First of all, it is instructive to introduce the notion of *determining trees* to depict diagrammatically the product decomposition(s) of an operator R over a set $\mathcal{T}$.

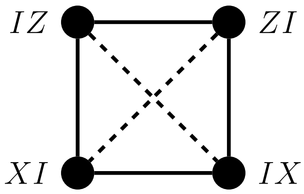


Figure 3.1: Compatibility graph of the set $\mathcal{T} = \{IZ, ZI, XI, IX\}$ shown with solid edges and its complement anti-compatibility graph with dashed edges.

---

[1] The process of measurement in quantum mechanics is a difficult one to capture precisely and carries with it a great deal of nuance [162]; the framework of measurement adopted for this work is described in Section 1.2.2.

By this we mean an acyclic undirected graph with a root node (uniquely characterised by its connection with precisely two edges) labelled R that satisfies the following criteria:

1. The 'children' nodes residing under a common parent must commute;
2. Each parent is the operator product of its children.

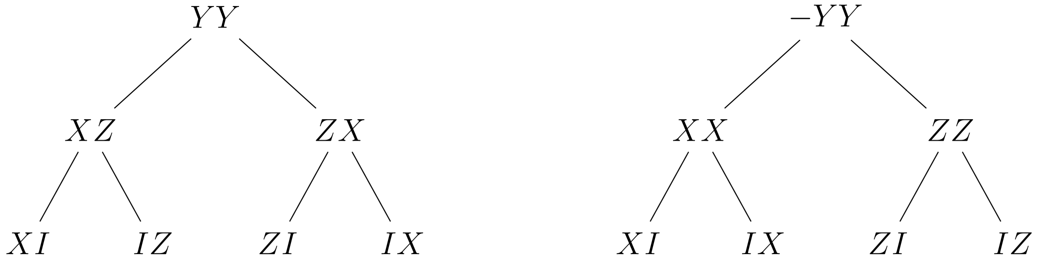
We refer to the *leaves* $L \in \mathcal{T}$ of our determining tree as the nodes with only a single adjacent edge, where $m_L \in \mathbb{N}_0$ denotes the multiplicity of each leaf (we permit $m_L = 0$). If such a determining tree exists then we may write

$$R = \prod_{L \in \mathcal{T}} L^{m_L}. \tag{3.2}$$

Furthermore, given a value assignment $\nu_L = \pm 1$ to the leaves $L \in \mathcal{T}$ we can infer the value assigned to R:

$$\nu_R = \prod_{L \in \mathcal{T}} \nu_L^{m_L} = \prod_{L \in \mathcal{D}_R} \nu_L, \tag{3.3}$$

in which $\mathcal{D}_R := \{L \in \mathcal{T} \mid m_L \text{ is odd}\}$ is the *determining set* of R; since any leaf whose multiplicity is even will make a $+1$ contribution in the product (3.3) they may be omitted, which is the motivation behind this definition.



(a) Determining tree for YY over $\mathcal{T}$.

(b) Determining tree for $-YY$ over $\mathcal{T}$.

Figure 3.2: Determining trees over the Peres-Mermin magic square of Table 3.2

As seen in Figure 3.2, it is possible to construct determining trees for both $\pm YY$ from the elements of $\mathcal{T}$, which moreover share the same determining set. This is problematic, as any joint value assignment to the leaves of the tree infer a contradictory assignment to $\pm YY$, which we observed previously with the incompatibility of row and column products in Table 3.1. This motivates the following definition of strict contextuality:

> We shall say that a set of Pauli operators $\mathcal{T}$ is *contextual* if and only if there exists an operator R that admits determining trees for both $\pm R$ whose determining sets coincide, i.e. $\mathcal{D}_R = \mathcal{D}_{-R}$.

We may reformulate this definition by noting that the determining set of a product of Pauli operators PQ is given by $\mathcal{D}_P \cup \mathcal{D}_Q$. Applying this observation to the trivial statement $-T = -\mathbb{1}T$ yields $\mathcal{D}_{-T} = \mathcal{D}_{-\mathbb{1}} \cup \mathcal{D}_T$, so in the instance that $-\mathbb{1}$ admits a determining tree over $\mathcal{T}$ which additionally has empty determining set

$\mathcal{D}_{-\mathbb{1}} = \varnothing$, we recover the above definition of contextuality. Another observation is that symmetries of a particular set, discussed in Chapter 2, hold no bearing on its contextual status, meaning they can be ignored when trying to assess whether a particular collection of Pauli measurements exhibits contextuality.

With this in mind, we may write down a number of equivalent statements of (non)contextuality. Suppose we are given a set of Pauli measurement terms $\mathcal{T} \subset \mathcal{P}_N$ we wish to classify and let $\mathcal{G} = \varsigma(\mathcal{T})$ be an independent generating set of its $\mathbb{Z}_2$-type symmetries, obtained efficiently via the algorithm described in Section 2.1.3. With $\mathcal{S} = \mathcal{T} \cap \overline{\mathcal{G}}$ the subset of symmetry elements present within $\mathcal{T}$, the following statements are equivalent:

NC.1 $\mathcal{T}$ is a noncontextual set of Pauli measurements;

NC.2 For $A, B, C \in \mathcal{T}$, if $[A, B] = [B, C] = 0$ but $[A, C] \neq 0$, then $B \in \mathcal{S}$ (in other words, commutation transitivity violations must pass through the symmetry subset);

NC.3 Commutation is transitive on $\mathcal{T} \setminus \mathcal{S}$;

NC.4 Commutation forms an equivalence relation over $\mathcal{T} \setminus \mathcal{S}$;

NC.5 $\mathcal{T} = \mathcal{S} \cup \mathcal{C}_1 \cdots \cup \mathcal{C}_L$ where $\{C_i, C_j\} = 0$ for $C_i \in \mathcal{C}_i, C_j \in \mathcal{C}_j$, $i \neq j$ and $[C_i, C_i'] = 0$ for $C_i, C_i' \in \mathcal{C}_i$.

NC.6 The compatibility graph of $\mathcal{T} \setminus \mathcal{S}$ is comprised of disjoint cliques;

NC.7 The anti-compatibility graph of $\mathcal{T} \setminus \mathcal{S}$ is multipartite complete.

Statement **NC.6** above is illustrated in Figure 3.3 for maximal 6-qubit noncontextual sets, providing a convenient graphical representation of strong measurement contextuality.

In the interest of computational efficiency, it is necessary to devise a programmatic approach to assessing whether an arbitrary Pauli set is contextual. Given that we have identified strong measurement contextuality with the violation of commutation transitivity, this provides a first algorithm for testing contextuality: search through triples of elements to check whether they satisfy the transitivity condition. If a non-transitive relation is found, the set is contextual, whereas if no such violations are present it must therefore be *noncontextual*. This approach was the first to be suggested [170] and scales as $\mathcal{O}(M^3)$ where $M = |\mathcal{S}|$ is the size of the set we wish to assess; this algorithm is implemented in the OpenFermion Python package [171].

However, this scaling is not ideal and is impractical even for Pauli sets of modest size. Instead, we propose an improved approach that leverages the symplectic representation of Section 2.1. Recall that we may identify a set $\mathcal{T} \subset \mathcal{P}_N$ of size M with a binary matrix $\hat{T} \in \mathbb{Z}_2^{2N \times M}$, up to a reordering of rows. Then, with the canonical symplectic form $\Omega = X \otimes \mathbb{1}_N$ of Equation (2.10) and following the commutation equivalence of Section 2.1.2, we may efficiently compute the adjacency matrix $A \in \mathbb{Z}_2^{M \times M}$ of the compatibility graph for the set $\mathcal{T}$ as

$$A = \hat{T} \Omega \hat{T}^\intercal \mod 2 \tag{3.4}$$

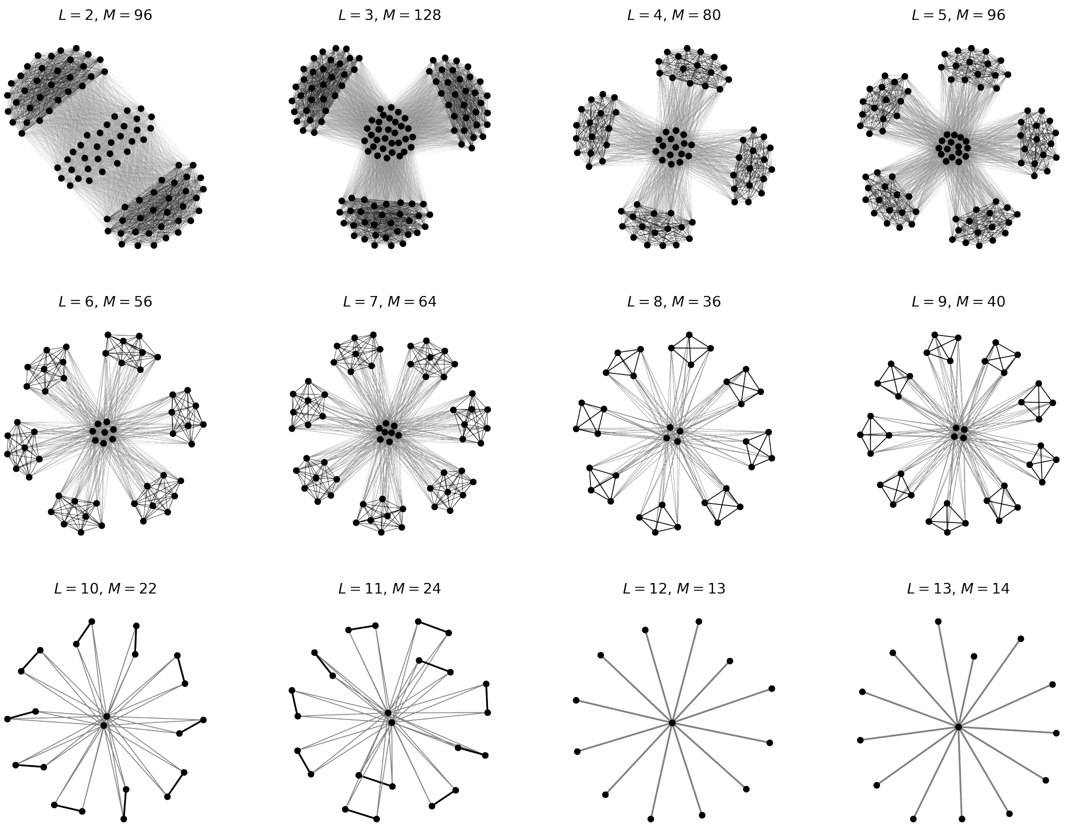


Figure 3.3: Compatibility graphs for maximal 6-qubit noncontextual sets. This highlights the decomposition into $\mathcal{S} \cup \mathcal{C}_1 \cdots \cup \mathcal{C}_L$ where $\mathcal{S}$ is the symmetrical component of the set, while the $\mathcal{C}_i$ form disjoint cliques (Statement **NC.6**). Drawing on the equivalent definitions of noncontextuality above, commutation transitivity violations are permitted only when passing through a symmetry element.

by performing $\mathcal{O}(M^2)$ XOR operations between N-dimensional binary vectors. The entries of this matrix are

$$A_{ij} = \begin{cases} 1 & \{\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\} = 1 \\ 0 & [\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j] = 0 \end{cases}, \tag{3.5}$$

or, in other words, zero indicates commutation and one anticommutation between Pauli operators i and j. Construction of this adjacency matrix is the dominant source of complexity in this new contextuality check.

Next, we drop any zero rows/columns from A; this corresponds with discarding the globally commuting terms and therefore leaves us with an adjacency matrix A' over the reduced set $\mathcal{T} \setminus \mathcal{S}$. Now, suppose two different rows contain a zero in the same column – for the set to be noncontextual these two rows must be identical, since this implies they belong to the same equivalence class; otherwise, the system must be contextual. This can be achieved by taking the bit-flipped matrix $\mathbb{1} - A'$, identifying the unique rows and summing down columns; if the resulting vectors consists only of ones, the set is noncontextual, else it is contextual. The overall complexity is $\mathcal{O}(M^2)$, thus improving over the original implementation by a factor of M, coming from construction of the adjacency matrix.

In Figure 3.4 we benchmark these two contextuality test algorithms for 20-

qubit noncontextual sets consisting of an increasing number of terms M. The time-to-solution for the symplectic adjacency-based approach achieves a speed-up of three orders of magnitude over the original commutation-transitivity-search approach.

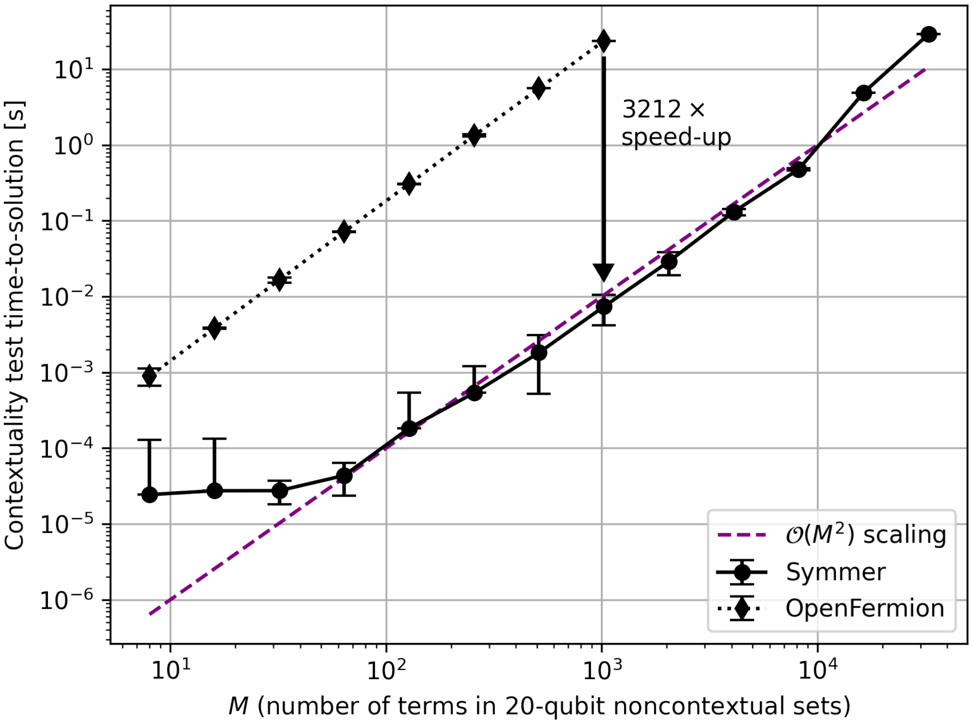


Figure 3.4: Comparison of the contextuality test algorithm from Kirby & Love [170] (implemented in OpenFermion [171]) versus the symplectic adjacency-based approach presented here (as implemented in Symmer [10]).

3.2 The Noncontextual Hamiltonian Problem

Above we attributed contextuality as a property belonging to a set of Pauli measurements $\mathcal{T}$. A Hamiltonian, on the other hand, is a weighted sum of such terms; as it turns out, it is possible to derive an expression for the eigenspectrum of a noncontextual Hamiltonian H_{noncon} in terms of eigenvalue assignments to its symmetry generators (which one might view as the hidden variables describing the system) [143, 172]. Recalling Statement **NC.5**, we obtain a convenient partitioning of a noncontextual Pauli Hamiltonian $\mathcal{T} = \mathcal{S} \cup \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_L$ where the $\mathcal{C}_i$ are equivalence classes with respect to commutation – in other words, elements of the same class commute and across classes they anticommute.

We consider a phase-space formulation of noncontextual quantum models as a means to approximate the behavior of quantum systems using classical structures. Such an approach draws heavily on the conceptual framework introduced by Spekkens [173, 174], in which quantum states are interpreted "epistemically", namely as distributions over a space of hidden variables subject to restrictions on the knowledge that can be gained. Within this framework, we construct a hidden-variable representation consisting of eigenvalue assignments to symmetry generators of a noncontextual set and a real-valued weighting of pairwise anticommuting representative observables selected from noncontextual equivalence classes.

The symmetry subset $\mathcal{S} \subset \mathcal{T}$ can be expanded by taking pairwise products within equivalence classes – since $\{C_i, C_j\} = 0$ for $C_i \in \mathcal{C}_i, C_j \in \mathcal{C}_j$ with $i \neq j$, it is the case that $[C_i C_i', C_j C_j'] = 0$ and we may define $\mathcal{S}' = \mathcal{S} \cup \bigcup_{i=1}^{L} \{C_i C_i' | \; C_i, C_i' \in \mathcal{C}_i\}$. As before, in Section 2.3, $\mathcal{S}'$ induces a symmetry group for which one may define independent generators $\mathcal{G}$ and a Clifford operation $U_{\mathcal{G}}$ mapping the generators to single-qubit Pauli operators; the inferred values over these qubits will again be determined by an assignment $\boldsymbol{\nu} \in \{\pm 1\}^{\times |\mathcal{G}|}$ of eigenvalues, analogous to the selection of a symmetry sector in qubit tapering. From each equivalence class $\mathcal{C}_i$ we select a representative C_i and a weighting r_i, yielding a vector $\boldsymbol{r} \in \mathbb{R}^L$ subject to $|\boldsymbol{r}| = 1$.

This construction defines a reduced phase-space over which a joint probability distribution can be formulated. Explicitly, the distribution

$$P(\boldsymbol{\nu}', \boldsymbol{r}' | \boldsymbol{\nu}, \boldsymbol{r}) := \frac{\delta_{\boldsymbol{\nu}, \boldsymbol{\nu}'}}{2^L} \prod_{i=1}^{L} |r_i + r_i'|, \tag{3.6}$$

which is a special-case of the phase-space distribution given by Spekkens [175], serves to capture the overlap between two hidden-variable configurations. Importantly, while this distribution is non-negative in our restricted setting, it belongs to a broader class of quasiprobabilities, which in general may admit negative values as signatures of contextuality or nonclassicality [176]. However, the present model intentionally avoids these features by construction, resulting in a strictly noncontextual framework.

Expectation values of observables defined within this phase-space are computed analytically. In particular, Kirby & Love [143] proved that the expectation value of a noncontextual symmetry generator is

$$\langle G \rangle_{(\boldsymbol{\nu}, \boldsymbol{r})} = \nu_G \; \forall \, G \in \mathcal{G}, \tag{3.7}$$

while the expectation value of a class representative is

$$\langle C_i \rangle_{(\boldsymbol{\nu}, \boldsymbol{r})} = r_i \; \forall \, i \in \{1, \dots, L\}. \tag{3.8}$$

As a consequence of the above, each hidden-variable configuration $(\boldsymbol{\nu}, \boldsymbol{r})$ may be associated with a valid quantum state $|\psi_{(\boldsymbol{\nu}, \boldsymbol{r})}\rangle \in \mathscr{H}$ that is stabilized by the set of operators $\{\nu_G G \mid G \in \mathcal{G}\} \cup \{C(\boldsymbol{r})\}$ where $C(\boldsymbol{r}) := \sum_{i=1}^{L} r_i C_i$ is a normalized sum over the weighted pairwise anticommuting representative observables. This holds as

$$\langle C(\boldsymbol{r}) \rangle_{(\boldsymbol{\nu}, \boldsymbol{r})} = \sum_{i=1}^{L} r_i \, \langle C_i \rangle_{(\boldsymbol{\nu}, \boldsymbol{r})} = \sum_{i=1}^{L} r_i^2 = |\boldsymbol{r}|^2 = +1, \tag{3.9}$$

through application of Equation (3.8) and guarantees compatibility between the hidden-variable model and a subset of stabilizer states in the full Hilbert space.

Since $\mathcal{G}$ generates the symmetry group $\mathcal{S}'$ and for any $C_i' \in \mathcal{C}_i$ we have $C_i C_i' \in \mathcal{S}'$ by construction, it holds that $\mathcal{S} \subset \overline{\mathcal{G}}$ and $\mathcal{C}_i \subset C_i \overline{\mathcal{G}} := \{C_i P | P \in \overline{\mathcal{G}}\}$. This means we may instead sum over the completion of $\mathcal{G}$, to obtain a formulation of a noncontextual Hamiltonian:

$$H_{\text{noncon}} = \sum_{P \in \overline{\mathcal{G}}} \left(h_P + \sum_{i=1}^{L} h_{P,i} C_i \right) P. \tag{3.10}$$

Putting everything together, we can express the expectation value of our noncontextual Hamiltonian as

$$\begin{aligned}\eta(\boldsymbol{\nu}, \boldsymbol{r}) &\coloneqq \langle H_{\mathrm{noncon}} \rangle_{(\boldsymbol{\nu}, \boldsymbol{r})} \\ &= \sum_{P \in \overline{\mathcal{G}}} \left(h_P + \sum_{i=1}^{M} h_{P,i} \langle C_i \rangle_{(\boldsymbol{\nu}, \boldsymbol{r})} \right) \langle P \rangle_{(\boldsymbol{\nu}, \boldsymbol{r})} \\ &= \sum_{P \in \overline{\mathcal{G}}} \left(h_P + \sum_{i=1}^{M} h_{P,i} r_i \right) \prod_{G \in \mathcal{G}_P} \nu_G,\end{aligned} \tag{3.11}$$

where $\mathcal{G}_P \subset \mathcal{G}$ satisfies $P = \prod_{G \in \mathcal{G}_P} G$.

The noncontextual energy spectrum is therefore parametrized by two vectors: the ±1 eigenvalue assignments $\boldsymbol{\nu}$, determining the contribution of the universally commuting terms, and $\boldsymbol{r}$, encapsulating the remaining pairwise anticommuting classes. In this sense, we may refer to $(\boldsymbol{\nu}, \boldsymbol{r})$ as a state of the noncontextual Hamiltonian itself, abstracted from quantum states of the corresponding stabilizer subspace. Optimizing over these parameters, we obtain the noncontextual ground state energy

$$\epsilon_0^{\mathrm{nc}} \coloneqq \min_{\substack{\boldsymbol{\nu} \in \{\pm 1\}^{\times |\mathcal{G}|} \\ \boldsymbol{r} \in \mathbb{R}^L : |\boldsymbol{r}| = 1}} \eta(\boldsymbol{\nu}, \boldsymbol{r}) \tag{3.12}$$

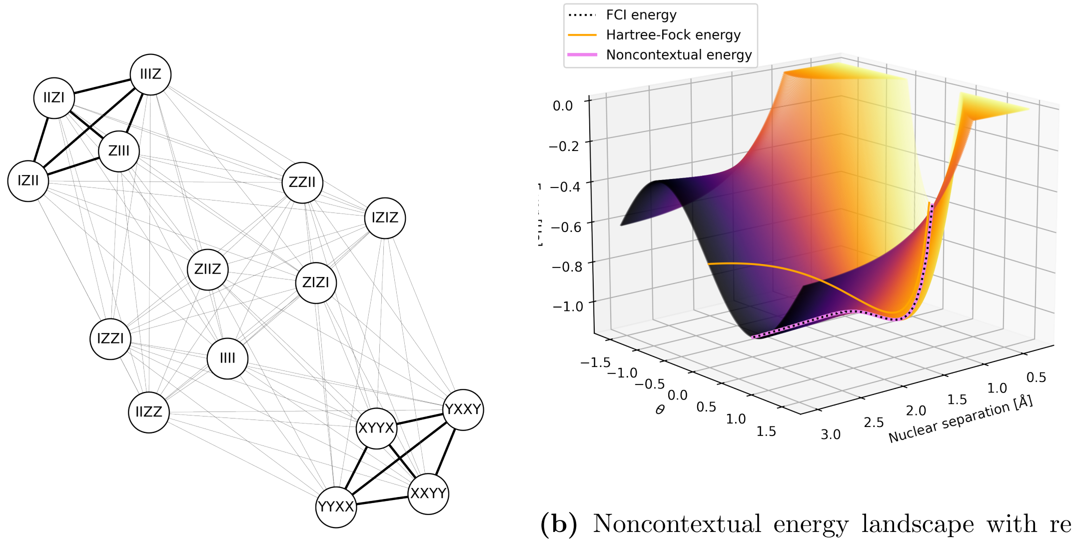
and call an element $(\boldsymbol{\nu}, \boldsymbol{r})$ of the preimage $\eta^{-1}(\epsilon_0^{\mathrm{nc}})$ a noncontextual ground state of H_{noncon}. Let us denote by $\Delta_{\mathrm{nc}} \coloneqq |\epsilon_0^{\mathrm{nc}} - \epsilon_0|$ the absolute error with respect to the true ground state energy.

This framework may be interpreted as a hidden-variable theory for a noncontextual subtheory of quantum mechanics. The configuration $(\boldsymbol{\nu}, \boldsymbol{r})$ plays the role of a classical state over a reduced observable set. The assumption of noncontextuality means that no joint assignment is attempted for incompatible observables; incompatibility refers to observables that are not captured by either the commuting subset or equivalence classes and moreover there is a restriction on expectation values of the latter due to constraints of the vector $\boldsymbol{r}$. This aligns with the idea of epistriction, wherein an observer's access to ontic information is limited to commuting subsets of observables, precluding the types of correlations that would violate classical realism [173].

Attempts to assign definite values to the full set of Pauli operators would reintroduce Kochen–Specker-type contradictions [177]. By narrowing the observable algebra to a classically simulable subset, we avoid contextuality by construction and define a subtheory whose energy landscape can be efficiently explored. In this sense, the noncontextual formalism provides a reduction of the full quantum problem and the epistemic constraints serve to isolate those features of the system amenable to classical interpretation. Though approximate, the model attempts to capture some of the structure present in a quantum system without incurring the overhead of a fully contextual description; in the following Section 3.3, we discuss how to relax noncontextual hidden-variable constraints to introduce contextual features and thus improve the approximation variationally.

Taken as a classical optimization problem, minimizing the objective function of Equation (3.11) is NP-complete in general. Despite this, we expect typical instances to be heuristically solvable by classical methods [143]. The same is true for both Density Functional Theory and Hartree-Fock; while in the worst-case these problems are NP-complete, the average-case hardness renders typical problem instances to be soluble in many cases of chemical interest [48]. If one fixes the ±1 eigenvalue assignments $\boldsymbol{\nu}$ – a case of identifying the correct symmetry sector – the phase-space formulation of noncontextual optimization reduced to a convex optimization problem over points of the unit $(L-1)$-sphere.

In Example 3.1 and Figure 3.5 we demonstrate that minimal basis molecular hydrogen (H_2 STO-3G) describes a noncontextual system and therefore permits a description by hidden variables. Figure 3.5b illustrates the convex optimization part of the phase-space formulation for varying bond lengths, noting that the minimum over this landscape coincides exactly with FCI.



(a) Compatibility graph of the Hamiltonian, which possesses the noncontextual graph structure with $L = 2$ seen previously in Figure 3.3.

(b) Noncontextual energy landscape with respect to the $\boldsymbol{r}$ parameter present in the phase-space formulation of the problem (3.11). The noncontextual states have the form $|\psi(\theta)\rangle = e^{i\theta XYXX}|1100\rangle$ and we plot $\langle H_{\text{noncon}}\rangle_{\psi(\theta)}$.

Figure 3.5: Molecular hydrogen, H_2 STO-3G, under the Jordan-Wigner transformation describes a noncontextual system.

3.3 The Contextual Subspace Method

In the previous Section 3.2 we introduced the noncontextual Hamiltonian problem and how corresponding spectra may be characterized classically. In general, the Hamiltonian of interest will exhibit contextuality amongst its comprising Pauli terms $\mathcal{T}$ and is therefore not directly amenable to the noncontextual approximation above. However, if one can identify some partitioning of the Pauli terms $\mathcal{T}$ into a noncontextual subset $\mathcal{T}_{\text{noncon}}$ and its contextual complement $\mathcal{T}_{\text{context}} = \mathcal{T} \setminus \mathcal{T}_{\text{noncon}}$, we may augment the noncontextual approximation with quantum corrections derived from the contextual Hamiltonian. In particular, this implies the form

$$H = H_{\text{noncon}} + H_{\text{context}} \tag{3.13}$$

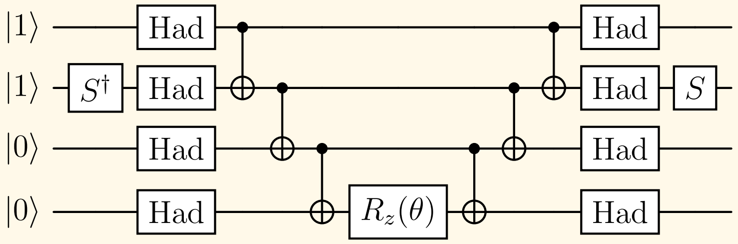
with H_{noncon} as in Section 3.2.

Example 3.1: Molecular Hydrogen is Noncontextual

Molecular hydrogen, or H_2, has been described as the “Hello World” of electronic structure [178]. In the STO-3G atomic orbital basis it consists of just two electrons in two spatial orbitals, thus mapping onto a 4-qubit problem under Jordan-Wigner. The vast majority of quantum algorithms are first validated on this modest system; for example, early demonstrations of Quantum Phase Estimation [81, 179]. However, we would argue that it is not even a ‘quantum’ problem. In fact, its Hamiltonian describes a noncontextual system at all bond lengths and thus can be described using hidden variables. Perhaps a better test benchmark is the isoelectric HeH^+ molecule, which actual does exhibit quantum (i.e. contextual) behaviour.
The H_2 STO-3G noncontextual set (under the Jordan-Wigner transformation as per Example 1.5) is $\mathcal{T} = \mathcal{S} \cup \mathcal{C}_1 \cup \mathcal{C}_2$ where

$$\begin{aligned} \mathcal{S} &= \{IIII, ZZII, ZIZI, ZIIZ, IZZI, IZIZ, IIZZ\}, \\ \mathcal{C}_1 &= \{ZIII, IZII, IIZI, IIIZ\}, \\ \mathcal{C}_2 &= \{YYXX, XYYX, XXYY, YXXY\}, \end{aligned} \tag{3.14}$$

whose compatibility graph is shown in Figure 3.5a. The noncontextual clique rotation generator is obtained by multiplying any two elements across $\mathcal{C}_1$ and $\mathcal{C}_2$, for example $XYXX \propto ZIII \cdot YYXX$ (up to a complex phase). An example circuit traversing the noncontextual energy landscape is:



In Figure 3.5b we present the noncontextual energy landscape with respect to the clique coefficient parameter $\boldsymbol{r}$, equivalent to sweeping the parameter θ in the circuit above, relevant to the phase-space formulation of the noncontextual spectrum in Equation (3.11).

While $C(\boldsymbol{r})$, introduced in Section 3.2, is not a stabilizer in the strict sense (it is not an element of the Pauli group), it is unitarily equivalent to one as a linear combination of anticommuting Pauli elements. Similar to the symmetry generators $\mathcal{G}$, it is possible to define a unitary operation U_C mapping $C(\boldsymbol{r})$ onto a single-qubit Pauli operator, following the approach of unitary partitioning [117–120, 180]. However, unlike the $U_{\mathcal{G}}$ rotation, $U_{\mathcal{C}}$ is not Clifford as it collapses L terms onto a single Pauli operator and can therefore introduce additional terms to the Hamiltonian. Kirby et al. [166] cautioned that, in principle, this increase in Hamiltonian complexity could be exponential in the number of equivalence classes L, namely a scaling of $\mathcal{O}(2^L)$. However, Ralli et al. [120] demonstrated that the general scaling for this sequence of rotations (SeqRot) method is $\mathcal{O}(x^{L-1})$ where $x \in [1, 2]$; that is, still exponential, yet the necessary conditions to obtain the worst-case $x = 2$ are contrived and have not been observed for any molecular Hamiltonians investigated to date. Regardless, one may circumvent this potentially adverse scaling entirely by implementing the linear combination of unitaries (LCU) approach to unitary partitioning [118, 119], which is only quadratic in the number of equivalence classes $\mathcal{O}(L^2)$ [120].

Appending $C(\boldsymbol{r})$ to our set of generators $\tilde{\mathcal{G}} := \mathcal{G} \cup \{C(\boldsymbol{r})\}$ and defining $U := U_C U_{\mathcal{G}}$, there exists a subset of qubit indices $\mathcal{I}_{\text{stab}}$ satisfying $|\mathcal{I}_{\text{stab}}| = |\tilde{\mathcal{G}}|$ and a bijective map $f : \tilde{\mathcal{G}} \to \mathcal{I}_{\text{stab}}$ such that $UGU^\dagger = \sigma_p^{(f(G))}$ for each $G \in \tilde{\mathcal{G}}$. We reiterate that $p \in \{1, 2, 3\}$ may be chosen at will; the approach taken by Kirby et al. [166] is to select $p = 3$ to enforce diagonal generators, i.e. $\sigma_3 = Z$. This partitions the Hilbert space $\mathscr{H} = \mathscr{H}_{\text{stab}} \otimes \mathscr{H}_{\text{red}}$ just as in Section 2.3.

Suppose we have a quantum state $|\psi_{(\boldsymbol{\nu},\boldsymbol{r})}\rangle$ that is consistent with ϵ_0^{nc}; since the rotated state $|\psi'_{(\boldsymbol{\nu},\boldsymbol{r})}\rangle = U\,|\psi_{(\boldsymbol{\nu},\boldsymbol{r})}\rangle$ must be stabilized by $\sigma_p^{(i)}\ \forall\, i \in \mathcal{I}_{\text{stab}}$, the qubit positions $\mathcal{I}_{\text{stab}}$ must be fixed. This implies a decomposition

$$|\psi'_{(\boldsymbol{\nu},\boldsymbol{r})}\rangle = |b_{(\boldsymbol{\nu},\boldsymbol{r})}\rangle_{\text{stab}} \otimes |\varphi\rangle_{\text{red}} \tag{3.15}$$

where $|b_{(\boldsymbol{\nu},\boldsymbol{r})}\rangle$ represents a single basis state of $\mathscr{H}_{\text{stab}}$ and $|\varphi\rangle \in \mathscr{H}_{\text{red}}$ is independent of the parameters $(\boldsymbol{\nu}, \boldsymbol{r})$. The expectation value of the full Hamiltonian may be expressed as

$$\langle H\rangle_{\psi_{(\boldsymbol{\nu},\boldsymbol{r})}} = \epsilon_0^{\text{nc}} + \langle \pi_{\boldsymbol{\nu}}^U(H_{\text{context}})\rangle_\varphi\,, \tag{3.16}$$

where $\pi_{\boldsymbol{\nu}}^U(H_{\text{context}})$ contains only the terms of the contextual Hamiltonian that commute with all the noncontextual generators, just as in (2.32). It was observed by Kirby et al. [166] that any term which anticommutes with at least one noncontextual generator must have zero expectation value and our stabilizer subspace projection captures this fact.

Inspecting (3.16), we may optimize freely over quantum states φ, i.e., we are not constrained by the noncontextual ground state within $\mathscr{H}_{\text{red}}$. In fact, we may absorb the noncontextual ground state energy into the reduced contextual Hamiltonian

$$\tilde{H}_{\text{context}} := \epsilon_0^{\text{nc}} \cdot \mathbb{1} + \pi_{\boldsymbol{\nu}}^U(H_{\text{context}}), \tag{3.17}$$

defining the *contextual subspace Hamiltonian*; this form is obtained naturally when applying the stabilizer subspace projection to the full Hamiltonian, which automatically includes the noncontextual energy by fixing the corresponding eigenvalue assignments.

Now, we may perform unconstrained VQE to obtain a quantum-corrected estimate

$$\epsilon_0^{\text{c}} := \min_{|\varphi\rangle \in \mathscr{H}_{\text{red}}} \langle \tilde{H}_{\text{context}}\rangle_\varphi \tag{3.18}$$

of the true ground state energy with absolute error $\Delta_{\text{c}} := |\epsilon_0^{\text{c}} - \epsilon_0| \leq \Delta_{\text{nc}}$. We have equality when the stabilizers span every qubit position, which is the case when $|\tilde{\mathcal{G}}| = N$ since the generators must be algebraically independent: this means the initial quantum correction is trivial as the noncontextual part determines the entire system.

For instances of the electronic structure problem there is no guarantee that ϵ_0^{c} will achieve chemical precision ($\Delta_{\text{c}} < 1.6\text{mHa} \approx 4\text{kJ/mol}$) and, indeed, it might not improve upon the noncontextual estimate (although it will never be worse, due to the variational principle applying in this case). However, one can easily define a subset of $\mathcal{T}_{\text{noncon}}$ that is again noncontextual – this is achieved by discarding one of the noncontextual generators $G \in \tilde{\mathcal{G}}$, along with the operators that it generates. We now append the discarded operators to the contextual Hamiltonian, relaxing the stabilizer constraint on the qubit position $f(G)$ and permitting a search over its Hilbert space. This process may be iterated until the noncontextual set is exhausted and we recover full VQE. This means that CS-VQE will improve upon the noncontextual energy using less quantum resource than full VQE – this is more rigorously defined in the next section.

In summary, what we have described here is a technique of scaling the relative sizes of the noncontextual (read classical) and contextual (read quantum) simulations in a reciprocal manner. We can therefore trade-off quantum and classical workloads in CS-VQE.

3.3.1 Expanding the Contextual Subspace

Now we describe the process of growing the contextual subspace more rigorously. We select a subset of noncontextual generators $\mathcal{F} \subset \tilde{\mathcal{G}}$ whose stabilizer constraints we mean to fix/enforce and construct a new noncontextual set $\mathcal{T}'_{\text{noncon}} \coloneqq \mathcal{T}_{\text{noncon}} \cap \overline{\mathcal{F}}$; the contextual set is expanded accordingly by appending the terms not generated by $\mathcal{F}$, i.e., $\mathcal{T}'_{\text{context}} \coloneqq \mathcal{T}_{\text{context}} \cup (\mathcal{T}_{\text{noncon}} \setminus \overline{\mathcal{F}})$. As before, there exists a unitary operation $U_{\mathcal{F}}$, a subset of qubit indices $\mathcal{I}_{\text{fix}} \subset \mathcal{I}_{\text{stab}}$ and a bijective map $f : \mathcal{F} \to \mathcal{I}_{\text{fix}}$ satisfying $U_{\mathcal{F}} G U_{\mathcal{F}}^\dagger = \sigma_p^{(f(G))}\ \forall\, G \in \mathcal{F}$ (the rotation $U_{\mathcal{F}}$ may or may not be Clifford depending on whether $C(\boldsymbol{r})$ is among the stabilizers we wish to fix).

Denote by $\epsilon_0^{\text{nc}}(\mathcal{F})$ the ground state energy of the new noncontextual Hamiltonian $\mathcal{T}'_{\text{noncon}}$ with absolute error $\Delta_{\text{nc}}(\mathcal{F}) \geq \Delta_{\text{nc}}$. While this is weaker as an estimate of the true ground state energy of the full system, at the very least we are guaranteed to recover the initial noncontextual ground state energy from performing a simulation of the expanded contextual subspace [166], which we describe below.

The stabilizer constraints of $\mathcal{F}$ are enforced over the Hilbert space $\mathscr{H}_{\text{fix}} = (\mathbb{C}^2)^{\otimes \mathcal{I}_{\text{fix}}}$ of qubits indexed by $\mathcal{I}_{\text{fix}}$, whereas we may perform a VQE simulation over $\mathscr{H}_{\text{sim}} = (\mathbb{C}^2)^{\otimes \mathcal{I}_{\text{sim}}}$, the Hilbert space of the remaining $N - |\mathcal{F}|$ qubits indexed by $\mathcal{I}_{\text{sim}} = \mathbb{Z}_N \setminus \mathcal{I}_{\text{fix}}$. Invoking the stabilizer subspace projection map $\pi_{\boldsymbol{\nu}'}^{U_{\mathcal{F}}}$ with the eigenvalue assignments $\boldsymbol{\nu}' = (\nu_i)_{i \in \mathcal{I}_{\text{fix}}}$ yields an expanded contextual subspace Hamiltonian

$$\tilde{H}_{\text{context}}(\mathcal{F}) \coloneqq \epsilon_0^{\text{nc}}(\mathcal{F}) \cdot \mathbb{1} + \pi_{\boldsymbol{\nu}'}^{U_{\mathcal{F}}}(H_{\text{context}}). \tag{3.19}$$

Performing an $|\mathcal{I}_{\text{sim}}|$-qubit VQE simulation over the contextual subspace we obtain a new quantum-corrected estimate

$$\epsilon_0^{\text{c}}(\mathcal{F}) \coloneqq \min_{|\varphi\rangle \in \mathscr{H}_{\text{sim}}} \langle \tilde{H}_{\text{context}}(\mathcal{F}) \rangle_{\varphi} \tag{3.20}$$

with an error satisfying $\Delta_{\text{c}}(\mathcal{F}) \leq \Delta_{\text{c}}$. Recall that $\Delta_{\text{c}} = \Delta_{\text{c}}(\tilde{\mathcal{G}})$ corresponds with the contextual error when we enforce the full set of noncontextual stabilizers.

Observe that, when $|\mathcal{I}_{\text{sim}}| = N$, we are simply performing full VQE over the entire system – this occurs when we do not enforce the stabilizer constraint for any of the noncontextual generators, i.e. $\mathcal{F} = \emptyset$. Therefore, it must be the case that

$$\Delta_{\mathrm{c}}(\mathcal{F}) \to 0 \text{ as } |\mathcal{F}| \to 0. \tag{3.21}$$

Furthermore, given a nested sequence of generator subsets $(\mathcal{F}_i)_i$ with $\mathcal{F}_{i+1} \subset \mathcal{F}_i$, then $\Delta_{\mathrm{c}}(\mathcal{F}_{i+1}) \leq \Delta_{\mathrm{c}}(\mathcal{F}_i)$ and the convergence is monotonic. In this way, CS-VQE describes an interpolation between a purely classical estimate of the ground state energy and a full VQE simulation of the Hamiltonian. In the context of electronic structure calculations, this often permits one to achieve chemical precision at a saving of qubit resource, as indicated by Kirby et al. [166] for a suite of tapered test molecules of up to 18 qubits. We note in (3.21) that the quality of the chosen ansatz and optimization procedure will limit the actual error one may achieve in practice. This statement instead indicates that, for an appropriate level of contextual subspace approximation, it is possible to construct a reduced Hamiltonian whose exact ground state lies within some error threshold of the true value.

Suppose we wish to find the optimal contextual subspace Hamiltonian of size $N' < N$. The problem reduces to minimizing the error $\Delta_{\mathrm{c}}(\mathcal{F})$ over the $\binom{|\tilde{\mathcal{G}}|}{N-N'}$ generator subsets $\mathcal{F} \subset \tilde{\mathcal{G}}$ satisfying $|\mathcal{F}| = N - N'$. CS-VQE is highly sensitive to this choice and remains a vital open question for the continued success of the technique. For chemistry applications, we grow the contextual subspace until the CS-VQE error attains chemical precision, which means finding the minimal $\mathcal{F}$ such that $\Delta_{\mathrm{c}}(\mathcal{F}) < 1.6\mathrm{mHa}$. In general, we will not have access to a target energy and so will not necessarily know when the desired precision is achieved; instead, we might choose the largest contextual subspace accommodated by the available quantum resource or iterate until the VQE convergence is within some fixed bound.

Greedily selecting combinations of $d \leq N$ generators that yield the greatest reduction in error is an effective stabilizer relaxation ordering heuristic, where iterate $k < N/d$ involves a search of depth d over $N - dk$ elements, thus necessitating $\sum_{k=0}^{\lfloor N/d \rfloor} \binom{N-dk}{d} = \mathcal{O}(N^{d+1})$ CS-VQE simulations. Taking $d = 2$ produces a good balance between efficiency and efficacy [166], but there is room for more targeted approaches that exploit some structure of the underlying problem. For example, in quantum chemistry problems it could be that one should relax the stabilizers that have non-trivial action near the Fermi level – between the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO). Excitations clustered around this gap are more likely to appear in the true ground state and should therefore not be assigned definite values under the noncontextual projection. This idea comes from the theory of pseudopotential approximations [181], in which it is observed that chemically relevant electrons are predominantly those of the valence space, whereas the core may be ‘frozen’, thus reducing the electronic complexity.

Alternatively, one might define a Hamiltonian term-importance metric that considers coefficient magnitudes [182] or second-order response with respect to a perturbation of the Hartree-Fock state [183]. In relation to this, it is also not clear which features of a molecular system mean that it might be more or less amenable to CS-VQE; additional insight here would allow one to predict how many qubits will be required to simulate a given problem to chemical precision.

It is not fully understood how CS-VQE relates to active space techniques more generally, but would be an interesting pursuit for future work. For example, the downfolding technique of subsystem embedding subalgebra coupled cluster (SES-CC) [184] presents a compelling approach that iteratively decouples excitations $\sigma = \sigma_{\text{int}} + \sigma_{\text{ext}}$ into an 'internal' part that belongs to a chosen excitation subalgebra and its 'external' complement that may additionally be combined with the double unitary coupled cluster (DUCC) ansatz [185]. This yields an effective Hamiltonian $H_{\text{ext}}^{\text{eff(DUCC)}} = (P + Q_{\text{int}})e^{-\sigma_{\text{ext}}} H e^{\sigma_{\text{ext}}} (P + Q_{\text{int}})$ where P projects onto the reference state and Q_{int} onto the subspace of excitations generated by σ_{int}. This has a similar form to our stabilizer subspace projection (2.30); indeed, it might be possible to reproduce SES-CC under a qubit mapping within the contextual subspace framework by identifying an appropriate noncontextual sub-Hamiltonian and stabilizer subspace.

3.3.2 Noncontextual Projection Ansatz

Prior to this work, reduced contextual subspace Hamiltonians (3.19) were solved by direct diagonalization [166] – clearly, this will not scale to larger systems, with the required classical memory increasing exponentially. Instead, they must be simulated by performing VQE routines, but defining an ansatz for the contextual subspace provided an obstacle to achieving this in practice.

However, having now placed the problem within the stabilizer formalism described in Section 2.3, we have already introduced (in Sections 3.2 and 3.3.1) the tools necessary to restrict an ansatz of the form

$$|\psi_{\text{anz}}(\boldsymbol{\theta})\rangle := e^{iA(\boldsymbol{\theta})} |\psi_{\text{ref}}\rangle \in \mathscr{H}, \tag{3.22}$$

defined over the full system, to the contextual subspace (3.19). The approach adopted here is equivalent to that which we defined for qubit tapering in (2.32). To restrict a parametrized ansatz operator

$$A(\boldsymbol{\theta}) = \sum_{\boldsymbol{\sigma} \in \mathcal{A}} \theta_{\boldsymbol{\sigma}} \boldsymbol{\sigma} \mapsto \tilde{A}(\boldsymbol{\theta}) \in \mathscr{B}(\mathscr{H}_{\text{sim}}) \tag{3.23}$$

in line with the stabilizer constraints $\mathcal{F} \subset \tilde{\mathcal{G}}$ we may simply call upon the stabilizer subspace projection map $\pi_{\boldsymbol{\nu}'}^{U_{\mathcal{F}}}$ once more, which yields a restricted ansatz state

$$|\tilde{\psi}_{\text{anz}}(\boldsymbol{\theta})\rangle := e^{i\tilde{A}(\boldsymbol{\theta})} |\tilde{\psi}_{\text{ref}}\rangle \in \mathscr{H}_{\text{sim}} \tag{3.24}$$

where

$$\tilde{A}(\boldsymbol{\theta}) := \pi_{\boldsymbol{\nu}'}^{U_{\mathcal{F}}}\big(A(\boldsymbol{\theta})\big). \tag{3.25}$$

Any rotated ansatz term $U_{\mathcal{F}} \boldsymbol{\sigma} U_{\mathcal{F}}^\dagger$ that is not identity or a Pauli σ_p on some subset of the qubit positions indexed by $\mathcal{I}_{\text{fix}}$ will vanish.

The restricted reference state $|\tilde{\psi}_{\text{ref}}\rangle$ is obtained from an effective partial projective measurement of $U_{\mathcal{F}} |\psi_{\text{ref}}\rangle$ (see the discussion on POVMs in Section 2.3) with outcomes defined by $\boldsymbol{\nu}'$, which yields a product state

$$\frac{P_{\boldsymbol{\nu}'} U_{\mathcal{F}} |\psi_{\text{ref}}\rangle}{\sqrt{\langle \psi_{\text{ref}}| U_{\mathcal{F}}^\dagger P_{\boldsymbol{\nu}'} U_{\mathcal{F}} |\psi_{\text{ref}}\rangle}} = |b_{(\boldsymbol{\nu},\boldsymbol{r})}\rangle_{\text{fix}} \otimes |\tilde{\psi}_{\text{ref}}\rangle_{\text{sim}} \tag{3.26}$$

where we have explicitly demarcated the separability across $\mathscr{H}_{\text{fix}}$ and $\mathscr{H}_{\text{sim}}$. The post-measurement state $|b_{(\boldsymbol{\nu},\boldsymbol{r})}\rangle \in \mathscr{H}_{\text{fix}}$ on the noncontextual subspace represents a single basis vector and can therefore be disregarded, leaving just the state of the contextual subspace – this we take as reference for our restricted ansatz. If the unitary partitioning rotations are *not* to be applied, then the $U_{\mathcal{F}}$ rotation is trivial over $\mathscr{H}_{\text{sim}}$ and we incur no expense in coherent resource. However, if one does enforce the operator $C(\boldsymbol{r})$ over the contextual subspace, there might be some non-trivial component of the rotation that must be applied in-circuit to ensure the ansatz lies within the correct subspace; referring to Section 3.3.1, for the SeqRot approach this will consist of at most $\mathcal{O}(N^2)$ CNOT operations in-circuit, whereas LCU is probabilistic due to the nature of block-encoding [119]. Given a hardware-efficient ansatz, one may neglect this since the optimizer should compensate the parameters accordingly.

We may now define the contextual subspace energy expectation function

$$\tilde{E}(\boldsymbol{\theta}) \coloneqq \langle \tilde{H}_{\text{context}} \rangle_{\tilde{\psi}_{\text{anz}}(\boldsymbol{\theta})} \tag{3.27}$$

with $\tilde{H}_{\text{context}}$ as in Equation (3.19), at which point we have reduced the problem to standard VQE, performed over a subspace of the full problem.

The molecular systems that were simulated to benchmark the noncontextual projection ansatz for CS-VQE are given in Table 3.3. The molecule geometries were obtained from the Computational Chemistry Comparison and Benchmark Database (CCCBDB) [186] and their Hamiltonians constructed using IBM's `qiskit-nature` [187] with `pyscf` the underlying quantum chemistry package [39].

Before we evaluate the efficacy of our noncontextual projection ansatz, there are a few features of Equation (3.24) that should be highlighted. First of all, from the discussion following Equation (3.26), we potentially apply some component of the operation $U_{\mathcal{F}}$ in-circuit, introducing further gates that will contribute additional noise. However, when the reference state is taken to be that of Hartree-

Molecular systems			Number of qubits		
Name	Charge	Mult.	Full	Taper	CS-VQE
Be	0	1	10	5	3
B	0	2	10	5	3
LiH	0	1	12	8	4
BeH	+1	1	12	8	6
HF	0	1	12	8	4
BeH_2	0	1	14	9	7
H_2O	0	1	14	10	7
F_2	0	1	20	16	10
HCl	0	1	20	17	4

Table 3.3: The systems investigated to benchmark the noncontextual projection ansatz (all in the STO-3G basis). The CS-VQE column indicates the fewest number of qubits required to achieve chemical precision.

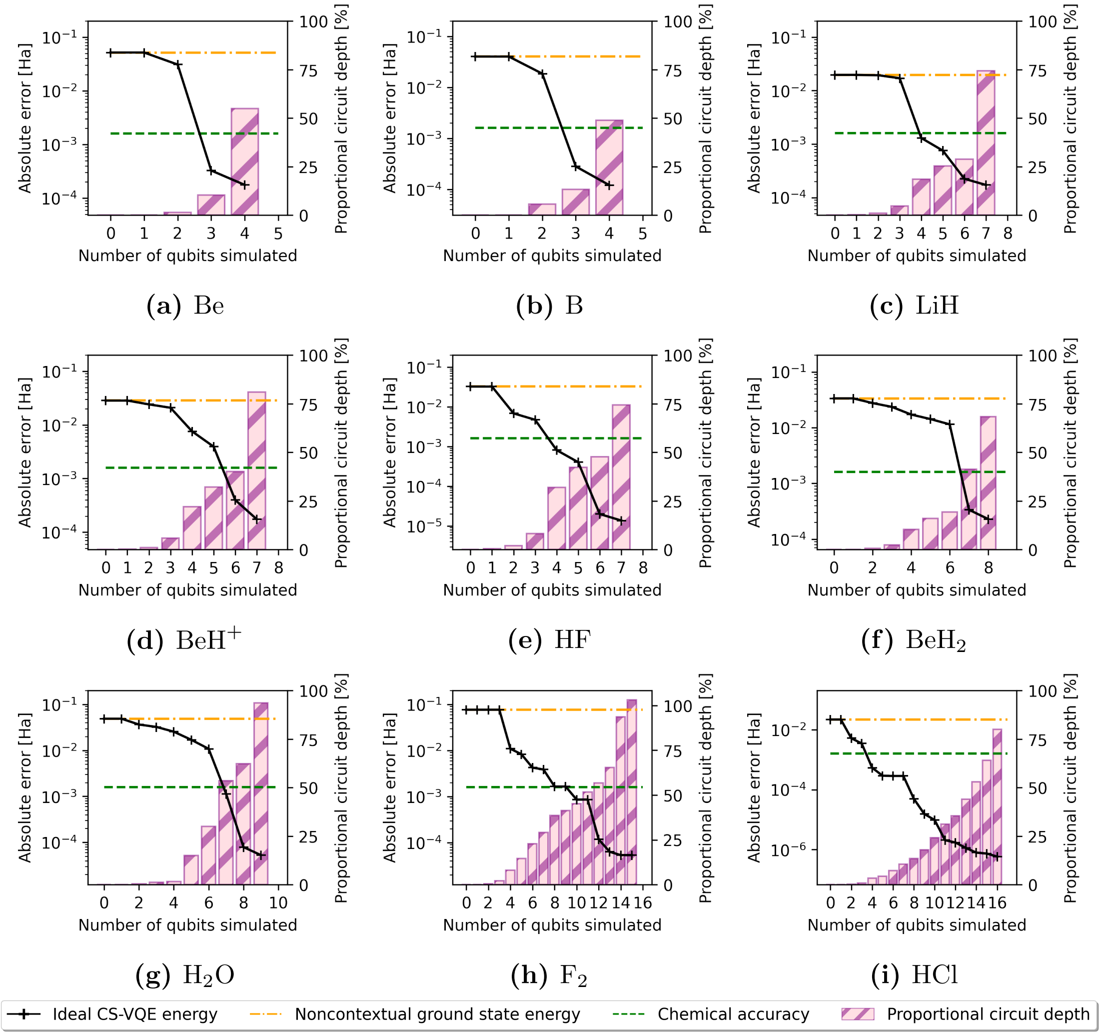
Fock, we observed $U\left|\psi_{\text{ref}}\right\rangle$ to coincide with the noncontextual ground state. This is an artifact of the noncontextual set construction heuristic prioritizing diagonal entries, used within both this work and that of Kirby et al. [166]. This need not always be the case, but for the molecular systems investigated this allows us to avoid performing $U_{\mathcal{F}}$ in-circuit and instead take the noncontextual ground state as our reference. Since we choose to rotate the noncontextual symmetry generators onto Pauli σ_3 operators here, this may be prepared by applying a Pauli σ_1 in each of the qubit positions $i \in \mathcal{I}_{\text{sim}}$ such that $\nu_i = -1$ so that the corresponding reference state is stabilized by the relevant operators $\nu_i \sigma_3^{(i)}$. This is visible in Figure 3.7, in which the VQE routine is initiated with the optimization parameters zeroed, i.e. $\boldsymbol{\theta} = \mathbf{0}$, and since $e^{i\tilde{A}(\mathbf{0})} = \mathbb{1}$ optimization begins at the noncontextual ground state energy.

Secondly, application of the unitary partitioning rotations U_C to the ansatz operator $A(\boldsymbol{\theta})$ may introduce additional terms by a worst-case scaling factor of $\mathcal{O}(2^{L-1})$ where L is the number of equivalence classes in Statement **NC.5**, although the true scaling is rarely this severe as discussed in the introduction to Section 3.3. We obtained $M = 2$ for all of the molecules tested, in which case SeqRot is identical to the asymptotically favourable LCU method. In fact, for small $M \ll N$ SeqRot may generate fewer terms than LCU (Ralli et al. presented a toy problem with $M = 3$ in which this was the case [120]) and therefore our choice of SeqRot here is valid given that the noncontextual set $\mathcal{T}_{\text{nc}}$ construction heuristic prioritizes the universally commuting terms $\mathcal{S}$ in Statement **NC.5**. Different heuristics may lead to larger values for M, in which case we recommend an adoption of LCU for implementations of CS-VQE.

Despite this, upon the subsequent projection of $A(\boldsymbol{\theta})$, it is possible that a significant number of terms will vanish. This is highly dependent on the quality of the initial ansatz and how heavily it is supported on the stabilized qubit positions $\mathcal{I}_{\text{fix}}$. Figure 3.6 presents circuit depths of the noncontextual projection ansatz as a proportion of the base ansatz from which it is derived, in this case the unitary coupled-cluster singles and doubles (UCCSD) operator. A net reduction in circuit depth is observed, which is quite dramatic up to the point of reaching chemical precision in the CS-VQE routine; in Table 3.4 we give the specific number of ansatz terms before and after application of the noncontextual projection to UCCSD and UCCSDT for the fewest number of qubits permitting chemical precision.

In order to identify a compact ansatz that closely captures the underlying chemistry with minimal redundancy, we employ the ADAPT-VQE methodology [121, 131, 188, 189] as introduced in Section 1.4.4. Recall that the algorithm centres around an operator pool from which terms are selected in line with a gradient-based argument and appended to a dynamically expanding ansatz whose parameters are optimized at each cycle via VQE. The particular approach we implement here is that of qubit-ADAPT-VQE [131], following on from iterative qubit coupled cluster [190], which searches at the level of Pauli operators rather than fermionic excitations [121].

To leverage ADAPT-VQE in the context of CS-VQE, we define an operator pool $\mathcal{O} \subset \mathcal{P}_N$ and apply to it the stabilizer subspace projection (2.30) to define a reduced pool $\pi_{\boldsymbol{\nu}'}^{U_{\mathcal{F}}}(\mathcal{O})$ for the corresponding contextual subspace. Projecting the full pool in this way will ensure that any symmetries S present will be preserved, since $[U_{\mathcal{F}} H U_{\mathcal{F}}^\dagger, U_{\mathcal{F}} S U_{\mathcal{F}}^\dagger] = U_{\mathcal{F}}[H, S]U_{\mathcal{F}}^\dagger = 0$, allowing us to incorporate some chem-



(a) Be **(b)** B **(c)** LiH

(d) BeH$^+$ **(e)** HF **(f)** BeH$_2$

(g) H$_2$O **(h)** F$_2$ **(i)** HCl

Figure 3.6: Ideal CS-VQE errors (left-hand axis) and corresponding noncontextual projection ansatz circuit depths as a proportion of the full UCCSD operator from which it is derived (right-hand axis) against the number of qubits simulated.

ical intuition into the contextual subspace despite an abstraction from the original problem; one could define a reduced pool directly, but care should be taken to avoid the inclusion of symmetry-breaking terms that may needlessly increase the complexity of the ADAPT-VQE procedure The algorithm is then executed as normal, only terminating once the ADAPT-VQE energy is chemically precise with respect to the FCI energy; for scalability, one should terminate computation when the largest gradient in magnitude falls below some predefined threshold, since the true ground state energy will not in general be known. In the Supporting Information, we provide a detailed description of the specific ADAPT-VQE implementation used within this work.

For the following, we take our pool $\mathcal{O}$ to be the terms of the UCCSD operator for each of the molecules in Table 3.3 before tapering and projecting into the relevant contextual subspace. In Figure 3.7, we present the ADAPT-VQE convergence data with expectation values obtained via exact wavefunction (statevector) calculations (i.e. no statistical/hardware noise); chemical precision is achieved in each instance. We used the adaptive moment estimation (Adam) [191] classical optimizer and computed parameter gradients as per the parameter shift rule [122]. Adam has been adopted for previous research in VQE for its resilience to noise, although it exhibits relatively slow convergence compared with other optimizers [192, 193] such as Broyden-Fletcher-Goldfarb-Shanno (BFGS) [194] and quantum natural gradient (NatGrad) [195]; the latter might be preferable for future work.

The number of ADAPT-VQE cycles (and therefore the number of terms in the resulting ansatz operator) is presented In Table 3.4, alongside the size of the projected UCCSD operator pool used; one observes a significant reduction in the number of terms. Extracting the optimal parameter configuration $\boldsymbol{\theta}_{\mathrm{min}}$ from the wavefunction simulations in Figure 3.7, we subsequently assess the effect of sampling noise on the simulation error with our ansatz circuit preparing the optimal quantum state $|\tilde{\psi}_{\mathrm{anz}}(\boldsymbol{\theta}_{\mathrm{min}})\rangle$.

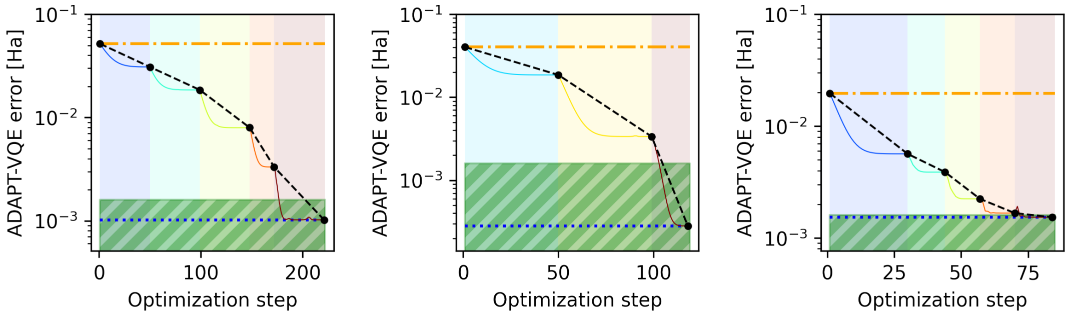
		Number of terms in ansatz operator		
Molecule	$\lvert\mathcal{I}_{\mathrm{sim}}\rvert$	UCCSDT (full/proj)	UCCSD (full/proj)	ADAPT-VQE
Be	3	(48/6)	(48/6)	5
B	3	(48/12)	(32/4)	3
LiH	4	(704/53)	(192/53)	5
BeH^+	6	(646/191)	(166/79)	11
HF	4	(92/57)	(92/57)	4
BeH_2	7	(1312/352)	(224/96)	10
H_2O	7	(1892/942)	(324/238)	21
F_2	10	(176/114)	(176/114)	12
HCl	4	(348/40)	(348/40)	4

Table 3.4: The number of Pauli terms $|\mathcal{A}|$ for a selection of (tapered) ansätze. The $|\mathcal{I}_{\mathrm{sim}}|$ column indicates the number of qubits in the contextual subspace over which the ansatz is projected and each tuple (full/proj) gives the number of terms pre and post projection. The final column gives the number of ADAPT-VQE cycles required to achieve chemical precision, with the operator pool consisting of the projected UCCSD terms; each simulation is plotted in Figure 3.7.

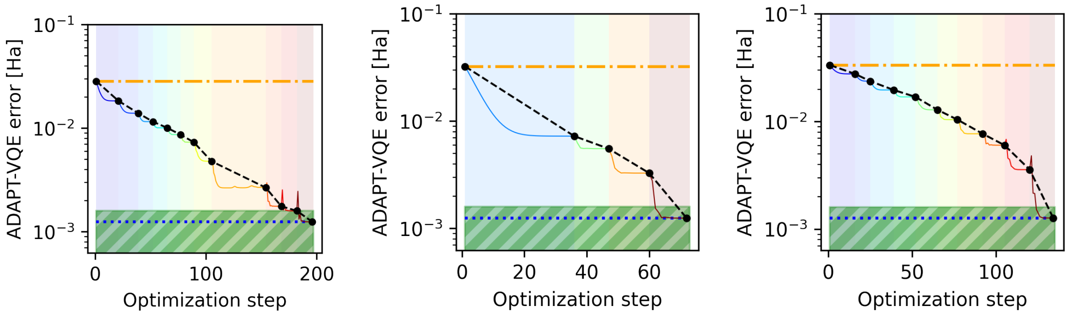
To achieve an absolute error of $\Delta > 0$, one should expect to perform $\mathcal{O}(\frac{1}{\Delta^2})$ shots (for each term of the Hamiltonian) [82]. Conversely, suppose we are allocated a quantity $S \in \mathbb{N}$ of shots – the obtained error should be of the order $\mathcal{O}(\frac{1}{\sqrt{S}})$. In order to increase estimate accuracy, we collected the Pauli terms into qubit-wise commuting (QWC) groups [116] using the graph-colouring functionality of NetworkX [196]; such groups may be measured simultaneously.

In Figure 3.8, the number of shots $S = 2^n$ for $n = 0, \dots, 20$ carried out per QWC group is varied and we observe the root mean-square error (RMSE) over twenty realizations of the ground state energy estimate, plotted on a log-log scale. For clarity, note the **only** source of noise here is that which arises from statistical variation of the quantum circuit sampling – we have **not** introduced hardware noise in the form of imperfect quantum gates or decoherence.

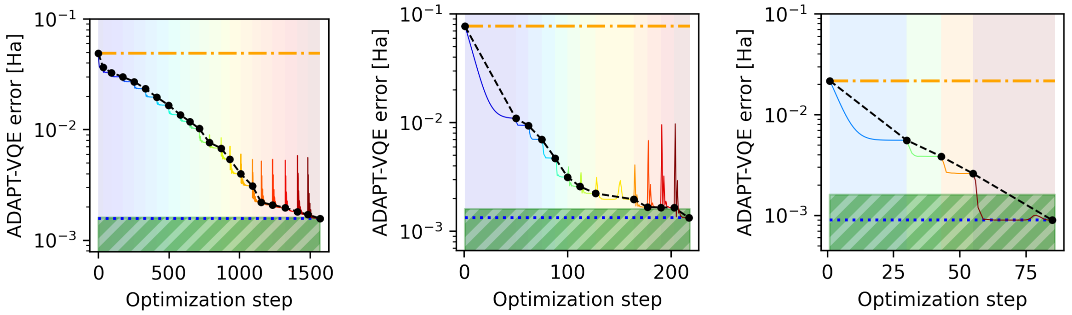
Two error regimes are observed, one of which is quite trivial: at high shot-counts we see a plateau resulting from the optimal error $|\tilde{E}(\boldsymbol{\theta}_{\min}) - \epsilon_0|$ being recovered. To assess the convergence properties outside of this limiting region, we plot a line of best fit $m \cdot \log_{10}(S) + c$ among the data not exhibiting such behaviour;



(a) Be 3-qubit CS-VQE **(b)** B 3-qubit CS-VQE **(c)** LiH 4-qubit CS-VQE



(d) BeH^+ 6-qubit CS-VQE **(e)** HF 4-qubit CS-VQE **(f)** BeH_2 7-qubit CS-VQE



(g) H_2O 7-qubit CS-VQE **(h)** F_2 10-qubit CS-VQE **(i)** HCl 4-qubit CS-VQE

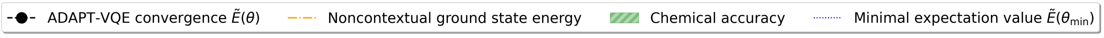
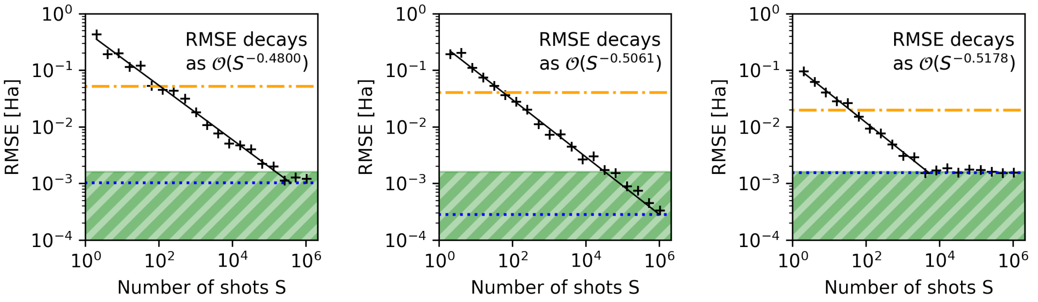


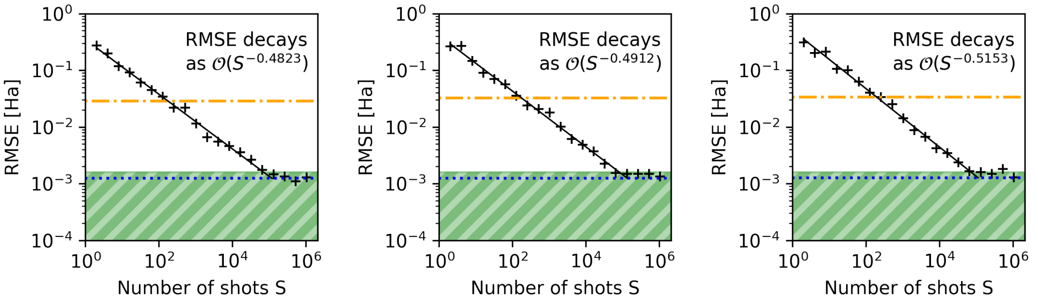
Figure 3.7: Validation of the noncontextual projection approach to ansatz construction for CS-VQE (3.24), used here in conjunction with ADAPT-VQE [121, 131, 188, 189]. We plot (on a $\log_{10}$ scale) the absolute error of wavefunction simulations conducted for the suite of trial molecules outlined in Table 3.3, each shown to achieve chemical precision; the horizontal axis indicates the algorithm step counter with each shaded region a separate ADAPT-VQE cycle. Adaptive moment estimation (Adam) [191] is the classical optimizer taken in the VQE routine performed over the contextual subspace for each ADAPT-VQE cycle and the settings used are as follows: tolerance $= 10^{-4}$, learning rate $= 10^{-2}$, $\beta_1 = 0.4, \beta_2 = 0.999, \epsilon = 10^{-8}$. The parameter gradients $\partial \tilde{E}(\boldsymbol{\theta})/\partial \theta_i$, required for both operator pool term selection and VQE, were computed using the parameter shift rule [122].

since the data is represented on a log-log scale, this corresponds with a decay in error of $\mathcal{O}(S^m)$. In each plot of Figure 3.8 we obtain $m \approx -0.5$, meaning the RMSE follows the predicted decay of $\mathcal{O}(\frac{1}{\sqrt{S}})$.

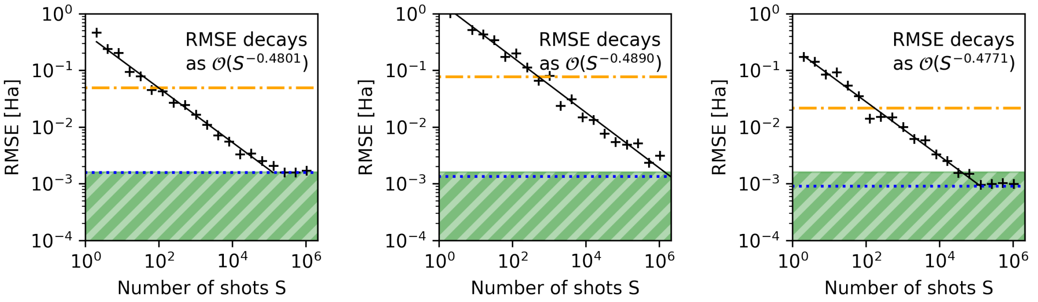
In every simulation bar F_2, chemical precision was achieved within $S = 2^{20} \approx 10^6$ shots per QWC group. However, our shot budget could be reduced by implementing more advanced allocation strategies, for example according to the magnitude of Hamiltonian term coefficients [197] or a classical shadow tomography approach [198, 199].



(a) Be 3-qubit CS-VQE **(b)** B 3-qubit CS-VQE **(c)** LiH 4-qubit CS-VQE



(d) BeH^+ 6-qubit CS-VQE **(e)** HF 4-qubit CS-VQE **(f)** BeH_2 7-qubit CS-VQE



(g) H_2O 7-qubit CS-VQE **(h)** F_2 10-qubit CS-VQE **(i)** HCl 4-qubit CS-VQE

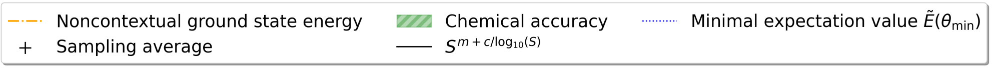


Figure 3.8: Each of the plots 3.8a - 3.8i correspond with 3.7a - 3.7i above and illustrate the statistical effect of sampling noise at the optimal parametrization $\boldsymbol{\theta}_{\min}$ determined from the ADAPT-VQE statevector simulations in Figure 3.7. We plot the root mean-square error (RMSE) for twenty 'realizations' of the ground state energy estimate with $S \leq 2^{20}$ shots executed via IBM's QASM simulator; determining the line of best fit $m \cdot \log_{10}(S) + c$ with respect to the log-log data indicates a decay in error of $\mathcal{O}(S^m)$.

Chapter 4

Quantum Error Mitigation

> With fault tolerance the ultimate goal, error mitigation is the path that gets quantum computing to usefulness.
>
> *IBM Quantum* [200]

Current and near-term quantum devices are plagued by noise and hardware error, limiting their capabilities considerably. Using a quantum computer today is like carving a sculpture into ice – in the desert. The information simply melts into the environment faster than you can use it. As such, the greatest challenge facing noisy intermediate-scale quantum (NISQ) computation is the handling of error, which manifests as imperfection in quantum logical operations or during readout. Gate errors may be categorized either as *coherent*, arising for example as imprecision in gate rotation angles and device miscalibration, or *incoherent*, representing a coupling of the system with its environment. Coherent error is hard to suppress since it leads to a systemic bias in the output, while incoherent error is stochastic in nature and its average effect can often be well-described by the *depolarizing channel* [201].

Error handling in quantum computation can be categorized as suppression (QES), mitigation (QEM) or correction (QEC). The first of these is implemented close to the hardware itself and attempts to deal with flaws in the operation and control of the device. Mitigation, on the other hand, serves to reduce bias in some statistical estimator of interest by executing ensembles of circuits that have been carefully designed to exploit a feature of the quantum noise; this typically comes at the cost of increased uncertainty in the resulting expectation values. Finally, error correction schema engineer redundancy into the system, forming ‘logical qubits’ from many physical qubits such that errors may be detected and corrected on-the-fly during computation. Quantum error correction is a longer-term pursuit of the field and not in the scope of this research; instead, we turn here to techniques of error suppression and mitigation to extract a signal from the noisy quantum systems available to us today.

In Section 1.1 we introduced the notion of a *quantum channel*, which describes any operation one might perform on their quantum system. The general form was given in Equation (1.7), courtesy of Choi’s theorom on completely positive maps

[12]. A special case is the unitary map, the rank-1 channel, which most of our work up until this point has been concerned with (excluding quantum measurement, as it is an irreversible process). However, the remainder of this thesis now aims to deploy the algorithmic developments of the previous chapters on real quantum devices. Consequently, one needs to deal with the inherent faults of these systems, and to do so we need an appropriate language for quantum noise.

In the context of qubit systems, a convenient class of noise is described by the *Pauli channel*, where each Kraus operator $K_i = \sqrt{\lambda_i} P_i$ for $P_i \in \mathscr{P}_N = \{I, X, Y, Z\}^{\otimes N}$ is an element of the Pauli group, which forms a basis of $\mathscr{B}(\mathscr{H})$, and $\sum_i \lambda_i = 1, \lambda_i \geq 0$. The explicit map is therefore

$$\Phi(\rho) = \sum_{i=1}^{d^2} \lambda_i P_i \rho P_i \tag{4.1}$$

and is interpreted as the stochastic application of a Pauli error $\rho \mapsto P_i \rho P_i$ with probability λ_i. Using the convention that $P_1 = I^{\otimes N}$, a special case of the Pauli channel is found when we set $\lambda_1 = 1 - \delta + \frac{\delta}{d^2}$ and $\lambda_i = \frac{\delta}{d^2}$ for $i > 1$: the *depolarizing* channel. This may be described as a convex combination between the noiseless state ρ and the maximally mixed state,

$$\Phi_\delta(\rho) = (1 - \delta)\rho + \frac{\delta}{d} I^{\otimes N}, \tag{4.2}$$

where the parameter $0 \leq \delta \leq 1$ is the rate of depolarization. In geometrical terms, the depolarizing channel represents a contraction of the Bloch sphere towards its centre, as depicted in Figure 4.1, which is the point where the maximally mixed state lies.

Recall in Example 1.2 we introduced the notion of quantum state fidelity, a metric for the similarity of two quantum states. It provides a way to quantify how close an experimental quantum state is to the ideal or expected state. Given two density matrices ρ, σ, their fidelity is defined as $F(\rho, \sigma) := \left(\operatorname{Tr} \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)^2$, a quantity satisfying $0 \leq F(\rho, \sigma) \leq 1$. If $\rho = \sigma$, then we get $F(\rho, \rho) = 1$ since $\sqrt{\sqrt{\rho}\rho\sqrt{\rho}} = \rho$; conversely if the states are orthogonal then $F(\rho, \sigma) = 0$. In other words, high fidelity indicates that the two states are nearly identical, while low fidelity suggests they differ significantly. This concept is particularly useful for evaluating the performance of quantum devices, verifying the success of quantum operations. By assessing fidelity, we can gauge how well a quantum system preserves information and behaves as intended. For example, under the depolarizing channel

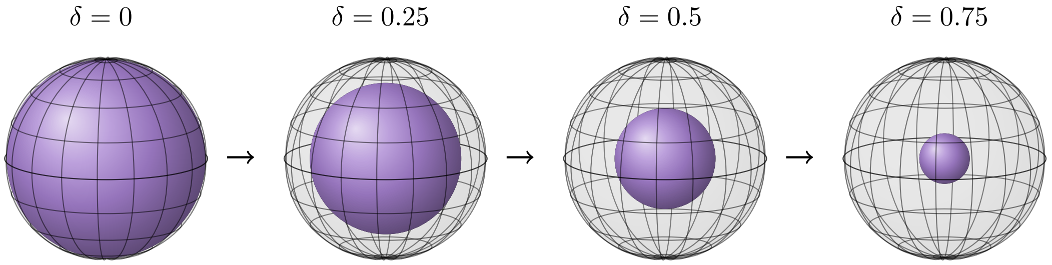


Figure 4.1: Geometrical interpretation of depolarizing noise on the Bloch sphere.

of Equation (4.2), we see the fidelity of an initial pure state $\rho = |\psi\rangle\langle\psi|$ compared with its depolarized form $\sigma = \Phi_\delta(|\psi\rangle\langle\psi|)$ is $F\big(|\psi\rangle\langle\psi|, \Phi_\delta(|\psi\rangle\langle\psi|)\big) = 1 - \delta + \frac{\delta}{d}$.

In Figure 4.2 we observe the decay in GHZ state fidelity against the number of qubits utilized for a suite of IBM Quantum Falcon chips. We superimpose on top of this the expected fidelity for purely depolarizing noise and see the average unmitigated results (solid lines) match quite convincingly up until 22 qubits, where we exceed the longest connected path of qubits on the device and therefore incur expensive SWAP operations. The depolarizing channel is an idealized model of quantum noise and previous works have attempted to learn the depolarization rate in order to mitigate error, for example with Clifford estimation circuits [201]. A more elaborate noise learning scheme was presented in [202] and is, to date, the largest calculation performed on quantum hardware.

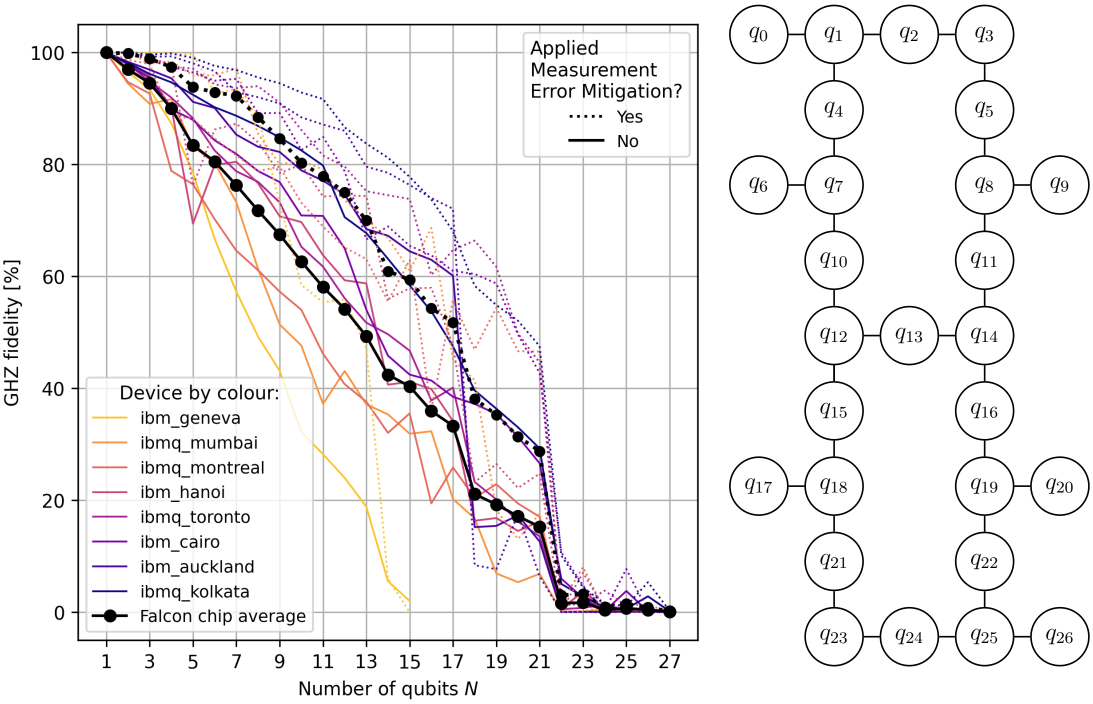


Figure 4.2: Decay in quantum state fidelity versus number of qubits in GHZ preparation across a suite of IBM Quantum Falcon chips. The dotted lines indicate fidelities after mitigating against readout error, which recovers approximately 10% on average (detailed explanation to come later in Section 4.2). We could not utilize more than 15-qubits on *ibm_geneva* due to a faulty qubit. The sharp drop to near-zero fidelity at $N = 22$ is due to the longest connected path of qubits being of length 21; beyond this point, we incur expensive SWAP operations that rapidly consume the remaining fidelity.

4.1 Statistical Tools

Quantum error mitigation typically exploits the properties of noise via ensembles of measurements and is therefore statistically motivated. Before proceeding onto the error mitigation methods we shall investigate in this thesis, we need to introduce the statistical tools that will aid our endeavour.

4.1.1 Estimators

The language we shall use to describe our QEM techniques is that of *estimators*. Suppose that we are interested in some observable O (a Hermitian operator, i.e. $O^\dagger = O$) and have access to a general quantum state ρ; we wish to estimate the quantity $\mathrm{Tr}\,(\rho O)$, but may only probe the state via some finite sample of quantum measurements $\mathcal{M} = \{m_i\}_{i=1}^{M}$ where $m_i \in \mathbb{Z}_2^N$. The way in which we collect and subsequently combine our sample to approximate the desired observable property defines an estimator $\mathcal{E} : \mathcal{M} \to \mathbb{R}$; the goal of QEM is to construct effective estimators that are capable of suppressing errors and extracting some usable data from the noise.

For example, we may define a naïve estimator for the expectation value of a Pauli operator $P \in \mathcal{P}_N$. Given a pure quantum state $|\psi\rangle$, we may sample from the quantum device in a compatible basis (i.e. one that commutes with P) and obtain eigenstates $|m_i\rangle$ such that $P\,|m_i\rangle = m_i\,|m_i\rangle$ where $m_i = \pm 1$ to estimate the expectation value $\langle P \rangle_\psi := \langle\psi|\,P\,|\psi\rangle$. The raw estimator is

$$\mathcal{E}^P_{\mathrm{RAW}}(\mathcal{M}) = \frac{1}{M}\sum_{i=1}^{M} m_i \to \langle P\rangle_\psi \quad (M \to \infty). \tag{4.3}$$

Since any Hermitian operator may be decomposed as $O = \sum_P o_P P$ with $o_P \in \mathbb{R}$, this allows us to extend our estimator to the full observable by linearity

$$\mathcal{E}_{\mathrm{RAW}} = \sum_P o_P \mathcal{E}^P_{\mathrm{RAW}}, \tag{4.4}$$

which shall form a baseline for our QEM benchmark.

We shall use the following metrics to assess the efficacy of QEM techniques:

$$\begin{aligned} \mathrm{var}(\mathcal{E}) &= \mathbb{E}(\mathcal{E}^2) - \mathbb{E}(\mathcal{E})^2 \\ \mathrm{bias}(\mathcal{E}) &= \mathbb{E}(\mathcal{E} - \langle O\rangle_\psi) \end{aligned} \tag{4.5}$$

and the related quantity

$$\begin{aligned} \mathrm{MSE}(\mathcal{E}) &= \mathbb{E}\Big((\mathcal{E} - \langle O\rangle_\psi)^2\Big) \\ &= \mathrm{var}(\mathcal{E}) + \mathrm{bias}(\mathcal{E})^2, \end{aligned} \tag{4.6}$$

or mean squared error. Taking $O = H$ and $|\psi\rangle$ the ground state of H, our objective is to approximate $\mathbb{E}(\mathcal{E}) \approx \langle\psi|\,H\,|\psi\rangle = \langle H\rangle_\psi = E_{\mathrm{FCI}}$. The goal of QEM is to reduce bias as far as possible (ideally within the threshold of chemical precision, i.e. $|\mathrm{bias}(\mathcal{E}_{\mathrm{QEM}})| < 1.6$ mHa) while aiming not to amplify variance severely.

QEM typically involves some classical post-processing of the quantum measurement data, possibly with a modification to the input circuits as we shall see in the cases of zero-noise extrapolation (Section 4.6) and echo verification (Section 4.7). In this way, QEM is equivalent to a modification of the estimator $\mathcal{E}$ by which we aim to estimate some observable O in a general state ρ; given a collection of quantum measurements M obtained from some experiment, we desire $\mathcal{E}(M) \approx \mathrm{Tr}\,(\rho O)$, but due to hardware noise and various other factors there will be non-sampling error that cannot be suppressed by simply increasing the size of our sample.

Although it would be preferable to run multiple instances of each quantum simulation to evaluate $\mathbb{E}(\mathcal{E})$, this is not feasible given the length of time taken to produce each energy estimate. Instead, we rely on the statistical tool of *bootstrapping*, introduced in further detail in Appendix 4.1.2, whereby we generate resampled data from the empirical measurement outcomes. Since this obeys the central limit theorem we may assume normality and evaluate the various statistics we are interested in as usual.

4.1.2 Bootstrapping

To evaluate the uncertainty in our energy estimates we rely on the statistical technique of *bootstrapping* [203]. Ideally, one would perform quantum experiments many times to probe the 'true' population, but from a practical standpoint this is not feasible due to the length of time required for each energy estimate (in our case ≈ 30 minutes for a shot budget of $B = 10^6$). Instead, we perform the experiment just once and generate resampled measurement data from the empirical distribution. This technique is widespread in statistics and makes the statistical analysis very convenient, not least as we may assume normality under the central limit theorem, which we verified using the *normaltest* function in *SciPy* [204] that implements the D'Agostino-Pearson test [205].

The idea underpinning bootstrapping is that we may resample from the observed measurement distribution to generate many additional theoretically-valid distributions that recover the noisy value in expectation. Suppose that we measure a quantum device N_{sample} times and obtain a random sample $\boldsymbol{m} = \{m_1, \dots, m_n\}$ in which the measurement outcome m occurs with frequency f_m. Our various QEM strategies combine these measurements in some way to yield an energy estimate $\mathcal{E}(\boldsymbol{m})$, but we would like to say something about the uncertainty in each estimator *without* having to perform further experiments. More concretely, this defines a random variable M with probabilities $P(M = m) = \frac{f_m}{N_{\text{sample}}}$ from which we may resample. While this process generates no additional information, it allows us to understand how emulated fluctuations in the data affect some estimator $\mathcal{E}$ that we are interested in; extracting resampled data sets $\boldsymbol{m}_i$ with $|\boldsymbol{m}_i| = N_{\text{sample}}$ from M allows us to obtain bootstrapped moments of $\mathcal{E}(\boldsymbol{m}_i)$.

We perform this process as many times as possible given the available compute resource, say $R \in \mathbb{N}$ repetitions, which allows us to approximate for example

$$\operatorname{var}(\mathcal{E}) \approx \frac{1}{R^2} \sum_{r=1}^{R} \sum_{s=r+1}^{R} \Big(\mathcal{E}(\boldsymbol{m}_r) - \mathcal{E}(\boldsymbol{m}_s)\Big)^2; \tag{4.7}$$

this is how we obtained the variances in Table 5.3. One might question whether bootstrapping is well-motivated here. A priori, one has no reason to expect acceptable agreement with the true population parameters, hence we ran 225 instances of our quantum experiment applied just to the diagonal terms of the Hamiltonian (given in Table 5.2), necessitating only computational basis measurements. We performed $10,000$ circuit shots in each experiment, for a combined total of 2.25×10^6 point samples before assessing the quality of the bootstrapped distributions against the overall sample. The 225 quantum experiments provide a *target* standard deviation σ, indicated by the vertical line in Figure 4.4, and we compare with this the *bootstrap* standard deviations obtained per experiment.

In Figure 4.3 we plot the result of our bootstrapping test and see reasonable agreement with the true energy distribution obtained from the NISQ hardware; the standard deviations all coincide with the experimentally-obtained value to $\mathcal{O}(10^{-3})$ (on the order of algorithmic accuracy), as indicated in Figure 4.4, and therefore we employ bootstrapping with confidence.

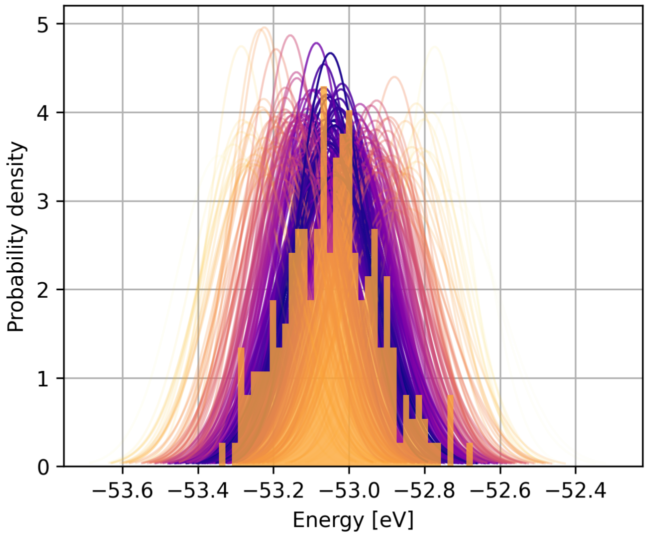


Figure 4.3: The true distribution of energy estimates obtained from 225 quantum experiments on *ibmq_kolkata*, each consisting of 10,000 circuit shots. Overlayed are the bootstrapped distributions for individual measurement sets to understand the relation between bootstrapping and the true population; the colour gradient indicates how far a given sample lies from the true (empirical) mean.

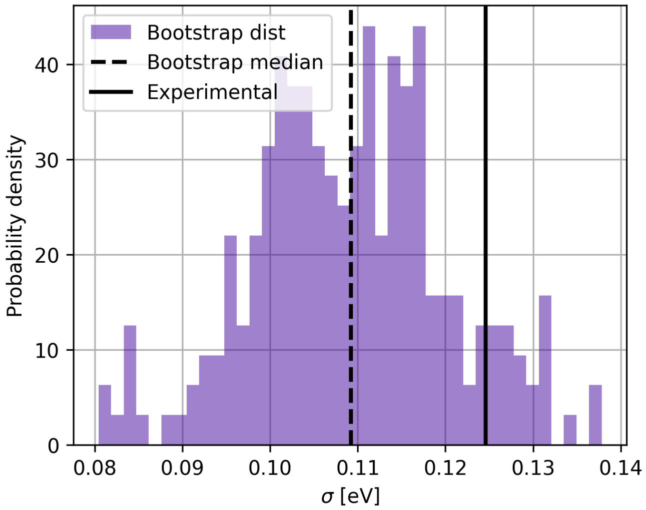


Figure 4.4: Distribution of bootstrapped standard deviations σ versus the experimentally obtained value on *ibmq_kolkata*. We observe good agreement, with the bootstrapped values correct up to 15 meV, nearly a third of algorithmic (chemical) accuracy.

4.1.3 Regression

For a simple linear regression model consisting of data pairs (x_i, y_i), in ordinary least-squares (OLS) one aims to find to find parameters α, β such that the residuals $r_i = y_i - \beta x_i - \alpha$ are minimized with respect to the objective function $S(\alpha, \beta) = \sum_i r_i^2$, the "sum of squared residuals" (SSR). However, OLS treats all data pairs equally, whereas it would be desirable to account for variability in the regression. Weighted least-squares (WLS), on the other hand, biases the fitting procedure towards data of low variance. This is achieved by defining weights W_i and modifying the objective function $S(\alpha, \beta) = \sum_i W_i r_i^2$; taking $W_i = \frac{1}{\sigma_i^2}$ where σ_i^2 is the variance of y_i it can be proved that this approach yields a best linear unbiased estimator (BLUE) [206].

To use WLS, clearly we need to be able to estimate the variances in our data, but this is not always directly accessible. As discussed previously, running quantum experiments multiple times to obtain the requisite distributions is impractical, given the significant overheads already involved in such calculations. Instead, we opt to leverage the bootstrapping procedure of Section 4.1.2 above, which permits us the use of WLS regression with no additional expenditure of coherent resource (only classical postprocessing). As we shall see later on, this produces improvements where regression techniques are required for quantum error mitigation, in particular for Zero-Noise Extrapolation (Section 4.6) and Clifford Data Regression (Section 4.8).

4.2 Measurement Error Mitigation

Measurement-error mitigation (MEM) (or *readout* error mitigation) aims to characterize the errors incurred during the readout phase of a quantum experiment [207]; it treats the state preparation itself as a black box and does not consider errors that occur prior to measurement.

A naive, non-scalable, approach to MEM is to prepare-and-measure each of the 2^N basis states individually; given some $|\boldsymbol{b}_i\rangle$ with $\boldsymbol{b}_i \in \mathbb{Z}_2^N$ we perform measurements to obtain a noisy distribution of binary outcomes $|\mu_{\text{noisy}}^{(i)}\rangle = \sum_j p_{i,j} |\boldsymbol{b}_j\rangle$ where $p_{i,j} = \langle \boldsymbol{b}_i | A | \boldsymbol{b}_j \rangle$ denotes the probability of preparing the state $|\boldsymbol{b}_i\rangle$ and measuring $|\boldsymbol{b}_j\rangle$. The doubly stochastic matrix $A = \sum_{i,j} p_{i,j} |\boldsymbol{b}_j\rangle \langle \boldsymbol{b}_i|$ is referred to as the assignment (or transition) matrix and lies at the core of this technique.

Now, suppose we wish to implement a circuit with noiseless measurement output $|\mu_{\text{ideal}}\rangle = \sum_i m_i |\boldsymbol{b}_i\rangle$; since $A|\boldsymbol{b}_i\rangle = |\mu_{\text{noisy}}^{(i)}\rangle$, then by linearity we have

$$A |\mu_{\text{ideal}}\rangle = \sum_i m_i |\mu_{\text{noisy}}^{(i)}\rangle =: |\mu_{\text{noisy}}\rangle . \tag{4.8}$$

More realistically, what we will actually have access to is $|\mu_{\text{noisy}}\rangle$, the output from some quantum experiment. Therefore, by inverting the assignment matrix we obtain a measurement-error mitigated distribution $|\mu_{\text{ideal}}\rangle = A^{-1} |\mu_{\text{noisy}}\rangle$.

In its current form, it will not be possible to construct the assignment matrix for large numbers of qubits. The "tensored" approach of Nation *et al.* [208] is designed to assess the *qubitwise* measurement assignment error, namely evaluating the probability p_k that qubit k is erroneously flipped $|0\rangle \rightleftharpoons |1\rangle$. The single-qubit assignment matrix for this process is

$$A^{(k)} = \begin{array}{c} \\ \langle 0| \\ \langle 1| \end{array} \begin{array}{c} \begin{array}{cc} |0\rangle & \quad |1\rangle \end{array} \\ \begin{bmatrix} 1-p_k & p_k \\ p_k & 1-p_k \end{bmatrix} \end{array} \tag{4.9}$$

and we subsequently reconstruct the full N-qubit assignment error probability by taking products over the relevant single-qubit transitions

$$A_{i,j} \approx \prod_{k=0}^{N-1} A^{(k)}_{(\boldsymbol{b}_i)_k,(\boldsymbol{b}_j)_k}. \tag{4.10}$$

This expression makes some strong assumptions on the character of the readout errors, in particular that they are predominantly uncorrelated. On the IBM Quantum hardware Nation *et al.* found this to be a reasonable assumption (using *ibmq_kolkata*), with little difference observed between this tensored approach versus a complete measurement calibration until inducing correlations by increasing the readout pulse amplitudes from their optimized values [208]. We note there

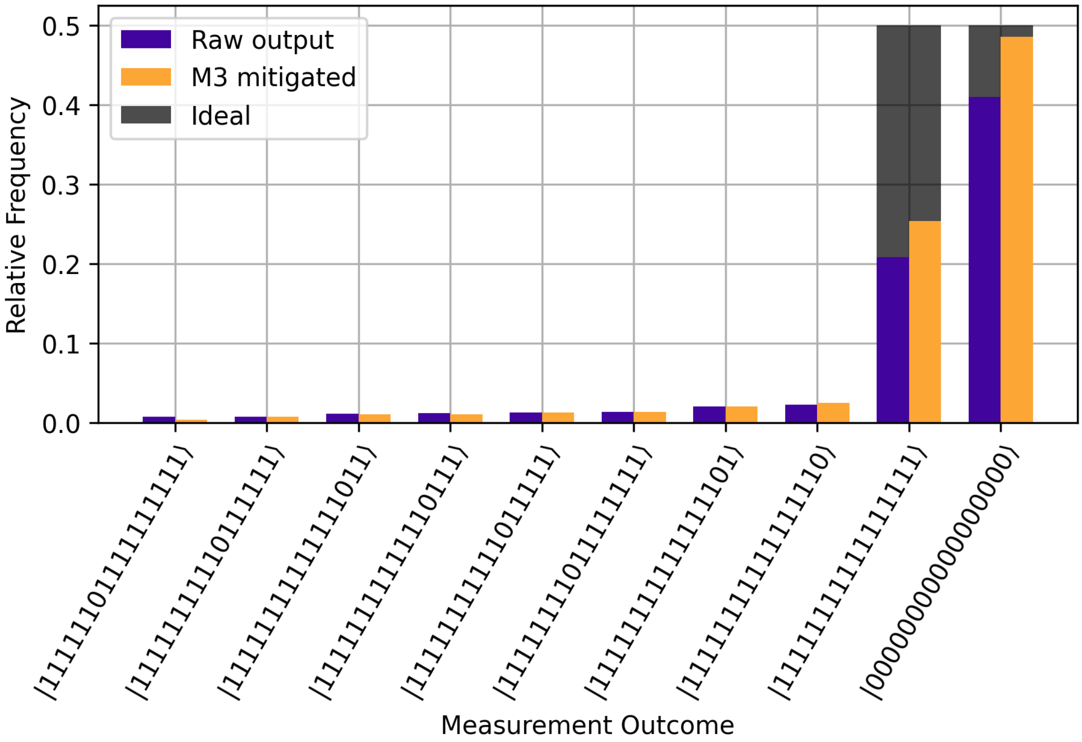


Figure 4.5: Comparing raw and MEM measurement distributions against the ideal output for 17-qubit GHZ preparation on *ibm_auckland* (the greatest number of qubits possible without SWAP operations) with 2^{15} circuit shots. Only outcomes exceeding a frequency of 10^{-2} are plotted here; this contributes 72.7% and 84.5% of the raw and MEM distributions, respectively. T_1 relaxation results in a reduced frequency of $|\mathbf{1}\rangle$ measurement outcomes compared with $|\mathbf{0}\rangle$ whereas they should be observed with equal probability 50%. Note also that the most frequently occurring spurious measurements differ from the correct $|\mathbf{1}\rangle$ outcome by a Hamming weight of just one, i.e. a single bit-flip.

are alternative techniques that do make such assumptions on the nature of the readout error, such as Twirled Readout Error Extinction (TREX), which leverages the idea of twirled measurements [209]

The expression of A in terms of single-qubit readout errors (4.10) requires just $2N$ quantum experiments to be carried out, versus 2^N in a complete measurement calibration. Furthermore, its form is particularly convenient as it is amenable to matrix-free iterative linear algebra techniques [210]. The Python package *mthree* developed through the work of Nation *et al.* is available in Qiskit; we utilized this for our QEM benchmark and is the only technique presented here that we did not implement ourselves. In Figure 4.5 we present the measurement distribution pre- and post-MEM for a 21-qubit GHZ preparation procedure on *ibmq_kolkata*, recalling from Figure 4.2 that we observed an increase from 29.3% to 47.5% in GHZ state fidelity. The effect of T_1 relaxation is also visible in this plot, whereby the $|\mathbf{0}\rangle$ state occurs with considerably greater probability than $|\mathbf{1}\rangle$ since the former is energetically favourable.

To investigate this further, in Figure 4.6 we plot the difference between probabilities of observing 0 or 1 measurement outcomes, p_0 or p_1 respectively, for the GHZ experiment of Figure 4.2. In the ideal setting we should have $p_0 = p_1 = 0.5$ and therefore the difference should be zero. However, due to the effect of T_1 relaxation, we see a drift in favour of 0 measurements as the number of qubits is increased in the GHZ state preparation. Interestingly, we actually see this effect exacerbated further by measurement error mitigation, which perhaps comes as a surprise. However, one should note that the MEM calibration circuits, i.e. preparing bitstrings $|\boldsymbol{b}_i\rangle$, trivially have a depth of one and therefore the calibration data contains no knowledge of T_1 relaxation. Furthermore, since MEM is oblivious to errors occurring during circuit execution, it cannot possibly rebalance the mea-

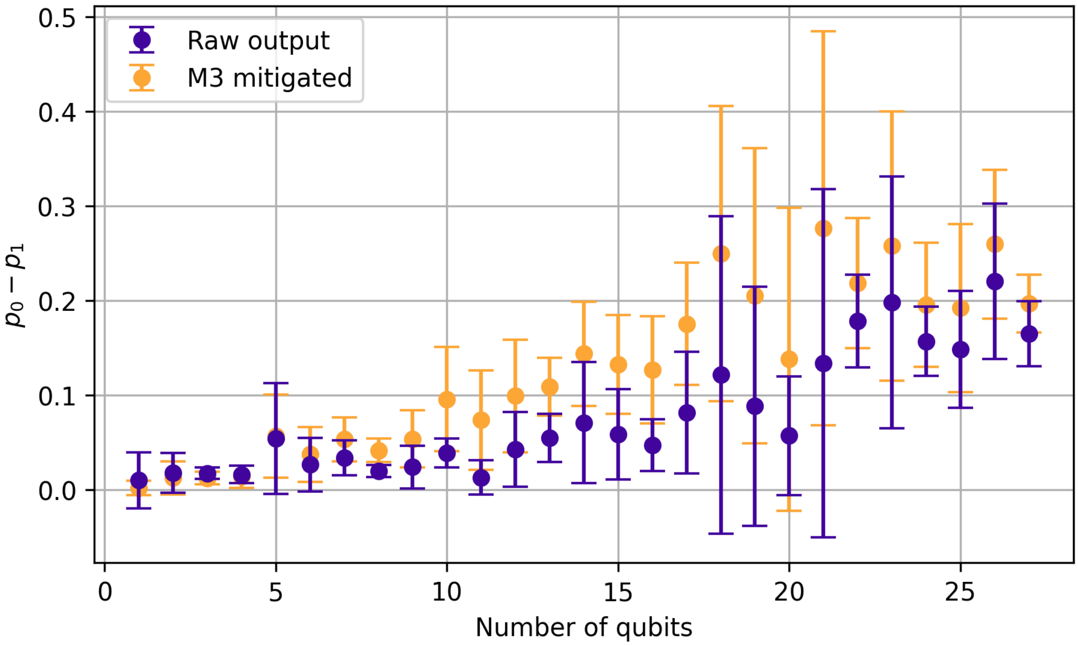


Figure 4.6: Difference, $p_{\mathbf{0}} - p_{\mathbf{1}}$, between the probabilities of observing 0 or 1 measurements for the GHZ experiment of Figure 4.2.

surement distribution towards the true distribution; additional error mitigation is required to achieve this.

4.3 Symmetry Verification

An inexpensive method of error mitigation is to take known symmetries of the Hamiltonian H and enforce stabilizer constraints on the measured binary strings resulting from a quantum experiment; we shall refer to this as *symmetry verification* (SV) [211–214]. Commuting operators share simultaneous eigenstates, i.e. given an S satisfying $[H, S] = 0$ and $|\psi\rangle$ such that $H|\psi\rangle = \epsilon|\psi\rangle$, we must have $S|\psi\rangle = s|\psi\rangle$ for some eigenvalue s of S. Therefore, after extracting a set of measurements from a quantum device, we may discard outcomes that violate the allowed symmetry.

This is not dissimilar to the themes presented in Chapters 2 and 3; the existence of symmetry offers opportunities for simplification, either in terms of a qubit reduction (vis-à-vis tapering and contextual subspace) or restriction of the state-space as is the case here. Indeed, both of these ultimately represent a reduction in dimensionality, thus making the problem easier to solve. In the case of $\mathbb{Z}_2$ symmetry, i.e. Pauli operators that commute termwise across the Hamiltonian, there is a choice to be made – does one taper the symmetry out of the Hamiltonian entirely, ensuring that all states in the reduced space are valid with respect to the projected symmetries, or do we retain some symmetry for the purposes of error mitigation as laid out in this section? It is perfectly valid to do both; for example, in the Bayesian phase estimation work by Yamamoto *et al.*, [81] they decided to taper two of the $\mathbb{Z}_2$ symmetries present in the H_2 problem of Example 4.1, while the third was retained for symmetry verification.

If we opt to taper out the physical $\mathbb{Z}_2$ symmetries of a Hamiltonian, we may no longer use them for the purposes of postselection of allowed measurement outcomes, as the reduced Hamiltonian has been projected into a subspace in which they are enforced *a priori*. However, there still exist symmetries of a more general nature that need not commute with each term individually, but do so with respect to the full Hamiltonian. Examples in the setting of electronic structure are the particle number and spin multiplicity operators

$$S_N = \sum_{i=1}^{N} a_i^\dagger a_i, \quad S_z = S_N^\uparrow - S_N^\downarrow, \tag{4.11}$$

where $S_N^{\uparrow/\downarrow}$ counts the number of up/down electrons, respectively. Under the Jordan-Wigner transformation, these take the form

$$S_N = \frac{1}{2}\left(N \cdot I^{\otimes N} - \sum_{i=1}^{N} Z_i\right), \quad S_z = \frac{1}{4}\sum_{i=1}^{N}(-1)^i Z_i; \tag{4.12}$$

note how the latter differs from the up/down spin parity operators of Equation (5.1). Furthermore, this expression assumes alternating spins, i.e. even/odd indices correspond with up/down electrons.

An important point is that we may only mitigate errors of terms that commute with the number and spin operators which, in the case of electronic structure, means only the diagonal ones; this may still yield significant improvements in error since these terms have the greatest coefficient magnitude and errors here will be amplified proportionally.

Given an ensemble of measurements $\{\boldsymbol{b}\}$, we discard any binary strings $\boldsymbol{b} \in \mathbb{Z}_2^N$ that do not respect the number and spin symmetries; given that we know the number of particles n in the system and the allowed spin values $\{s_0, \dots, s_{M-1}\}$ where $s_i = s - i$ for quantum number s (multiplicity $M = 2s + 1$), we require that $S_N \left|\boldsymbol{b}\right\rangle = n \left|\boldsymbol{b}\right\rangle$ and $S_z \left|\boldsymbol{b}\right\rangle = s_i \left|\boldsymbol{b}\right\rangle$ for some $i \in \{0, \dots, M-1\}$. For example, a molecule in a singlet configuration has only one allowable spin value $s = 0$ and thus valid quantum measurements are those in the kernel of S_z.

In Example 4.1 we illustrate this postselection for a simple molecular hydrogen system, in which just two computational basis states are valid electronic configurations respecting the particle number, spin parity and point group symmetries of the molecule. Complementing this example, in Figure 4.7 we observe how the energy error is affected under depolarizing noise when applying symmetry verification for H_2.

This postselected form of symmetry verification requires no additional coherent overhead and only minor postprocessing, yet we observe respectable error suppression from enforcing number and spin symmetries on the diagonal Hamiltonian terms. This can be seen in Table 5.3 where we prepare the equilibrium ground state of hydrogen chloride (HCl) on superconducting quantum hardware. It should be noted that there also exist techniques of symmetry verification *in hardware* [211, 212, 214], although such methods require additional coherent resource

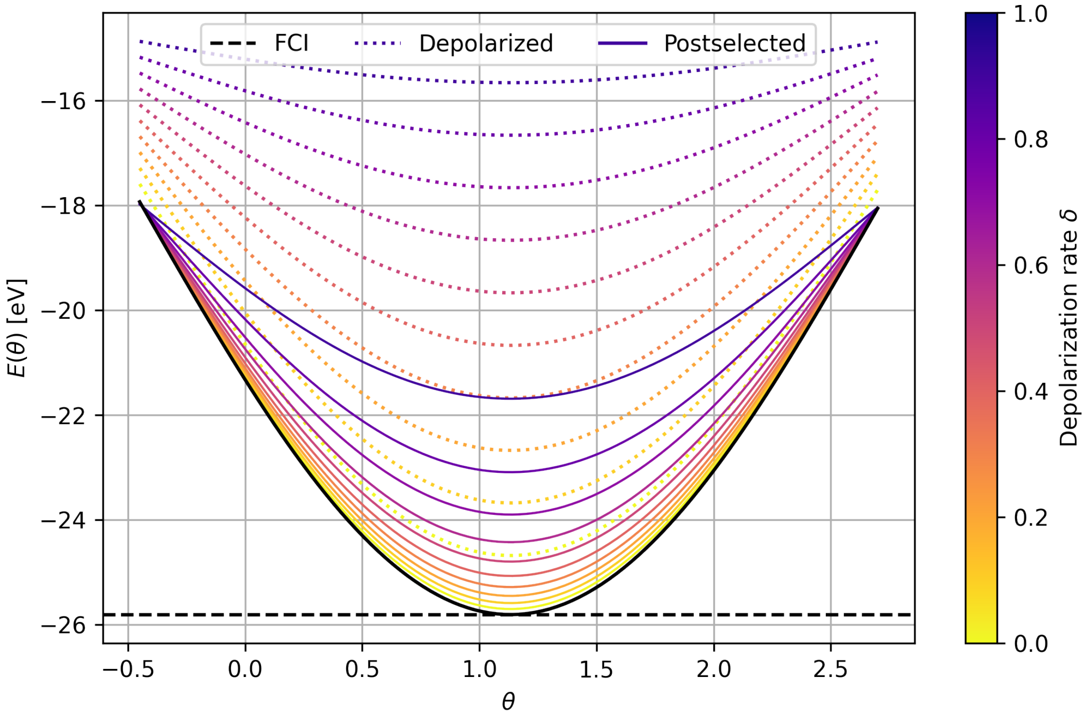
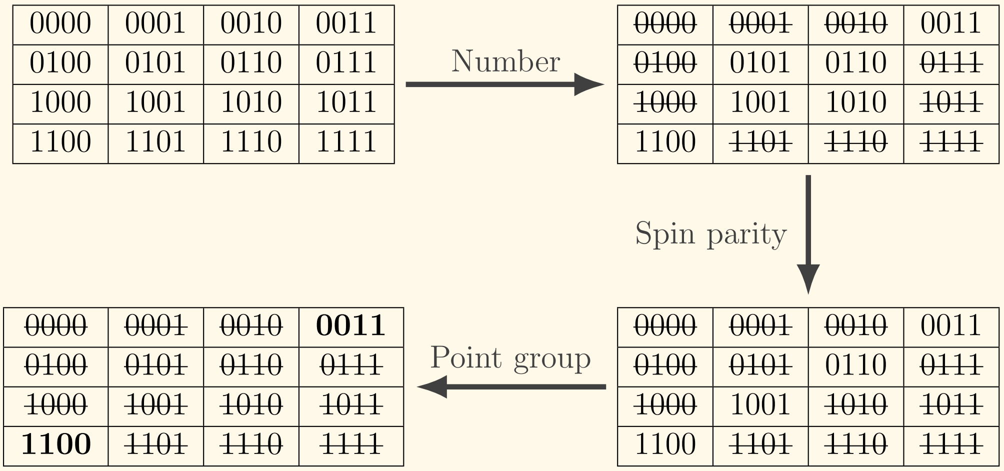


Figure 4.7: Molecular hydrogen (H_2) STO-3G equilibrium energy sweep over the ansatz parameter θ, where $E(\theta) = \left\langle 1100\right| e^{i\frac{\theta}{2}XYXX} H e^{-i\frac{\theta}{2}XYXX} \left|1100\right\rangle$. We compare the depolarized energy against postselection of the valid measurement outcomes $\left|1100\right\rangle, \left|0011\right\rangle$ as identified through Example 4.1.

and a high level of qubit connectivity, meaning it is challenging to implement at scale on near-term hardware.

Example 4.1: Postselected Measurements for H_2

We revisit molecular hydrogen, H_2, previously seen in Example 3.1. In the minimal STO-3G basis set H_2 consists of just four spin-orbitals, meaning its state-space is spanned by the $2^4 = 16$ computational basis states $|b_0 b_1 b_2 b_3\rangle$ for $b_i \in \{0,1\}$. However, we know *a priori* several molecular symmetries that we can exploit to constrain the allowed set of valid bitstrings. Firstly, we have the particlenumber operator $S_4 = 2IIII - \frac{1}{2}(ZIII + IZII + IIZI + IIIZ)$; each hydrogen atom contributes a single electron, so the allowed eigenvalue of the S_4 operator is $s = 2$ and we find 6 valid strings: $|0011\rangle, |0101\rangle, |0110\rangle, |1001\rangle, |1010\rangle, |1100\rangle$. Next we may impose the spin parity operators $ZIZI, IZIZ$, both with eigenvalue -1 since the molecule is in a singlet configuration so the electrons must be paired, i.e. one up, one down; this allows us to discard two additional strings: $|0101\rangle, |1010\rangle$. Finally, looking at the Hamiltonian terms from Example 3.1 there is a third $\mathbb{Z}_2$ symmetry $ZZII$ that arises from the molecular point group, with eigenvalue $+1$. After enforcing all these symmetries, we find just 2/16 of the available basis states, $|1100\rangle, |0011\rangle$, are actually valid for H_2. This is a particularly extreme example, as the Hamiltonian can be tapered to just a single qubit, hence the two degrees of freedom we identified here. This may further be viewed as a consequence of molecular hydrogen being noncontextual in the STO-3G basis, as highlighted in Example 3.1. In Figure 4.7 we show the effect of postselecting valid measurement outcomes.



0000	0001	0010	0011
0100	0101	0110	0111
1000	1001	1010	1011
1100	1101	1110	1111

~~0000~~	~~0001~~	~~0010~~	0011
~~0100~~	0101	0110	~~0111~~
~~1000~~	1001	1010	~~1011~~
1100	~~1101~~	~~1110~~	~~1111~~

~~0000~~	~~0001~~	~~0010~~	0011
~~0100~~	~~0101~~	0110	~~0111~~
~~1000~~	1001	~~1010~~	~~1011~~
1100	~~1101~~	~~1110~~	~~1111~~

~~0000~~	~~0001~~	~~0010~~	**0011**
~~0100~~	~~0101~~	~~0110~~	~~0111~~
~~1000~~	~~1001~~	~~1010~~	~~1011~~
1100	~~1101~~	~~1110~~	~~1111~~

In Figure 4.7 we show the effect of postselecting on these valid measurements for a sweep over the ansatz parameter from the noncontextual state circuit of Example 3.1.

4.4 Dynamical Decoupling

Dynamical decoupling (DD) is an error suppression (*not* mitigation) technique that aims to prolong the coherence of a spin system. This is achieved via stroboscopically applied control pulses which, if implemented carefully, can average over the coupling effect between a spin system and its environment. The mechanism underpinning dynamical decoupling predates quantum computing [215] but is a natural application of the technique.

To illustrate DD, one may consider the effect of qubit dephasing and how π-pulses might be integrated into the control schedule to prolong the lifetime of the qubit. In Figure 4.8 we illustrate the effect of dephasing noise on the Bloch sphere, showing that a carefully timed pulse can restore the state to its original position, although mistiming the spin echo, even if just a small discrepancy, will still result in dephasing (possibly in the opposite direction), eventually causing the state to decohere.

In our case, we apply periodic spin echos on idling qubits to suppress undesirable coupling between the system and its environment [216, 217]. We use a simple uniform sequence of π-pulses to effect the decoupling; different sequences with non-uniform spacing (such as Uhrig DD [218]) might yield improvements.

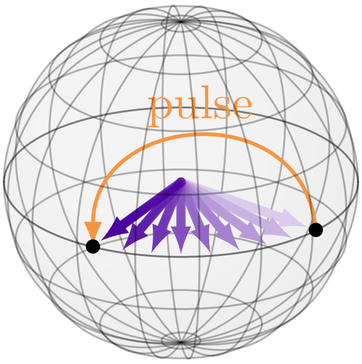


Figure 4.8: Illustration of dynamical decoupling, whereby the dephasing state is restored to its original position on the Bloch sphere.

4.5 Circuit Tiling

Noise is not uniform across the qubits of a quantum processor, hence one will observe considerable variance in the results when executing the same circuit on different parts of the chip; to mitigate this, we execute many replica circuits across the chip and average over the results, which has the added benefit of increasing the effective shot-yield. We depict our circuit parallelization scheme in Figure 4.9 for the 27-qubit *Falcon* architecture, which extends similarly to the 127-qubit *Eagle* device. One may view this as instance of ensemble averaging, often employed when computational models exhibit severe sensitivity to the initial conditions.

This noise averaging process results in improved stability of the final energy estimates, especially when used in combination with DD and MEM as introduced above. This is particularly desirable if one wishes to make inferences from the behaviour of these estimates under some noise amplification procedure, which is precisely the case for ZNE, introduced in the following Section 4.6. When performing regression, any uncertainty in the extrapolation data will propagate through to the inferred values and thus increase the variance of the final energy estimate.

This has also been observed when applying the error mitigation technique of randomized compilation (RC) [219] in combination with ZNE, where it is argued that small amounts of coherent error lead to substantial errors [220]. While RC converts coherent error into stochastic Pauli noise by implementing the target unitary in different ways, one might draw an analogy with the parallelization scheme presented here. Indeed, due to inconsistency in the noise channels for local qubit clusters, the unitary performed in each sub-circuit will differ and may explain the stable noise amplification observed in the following Section 4.6, ultimately leading to reliable extrapolation and lower variance in the final energy estimate.

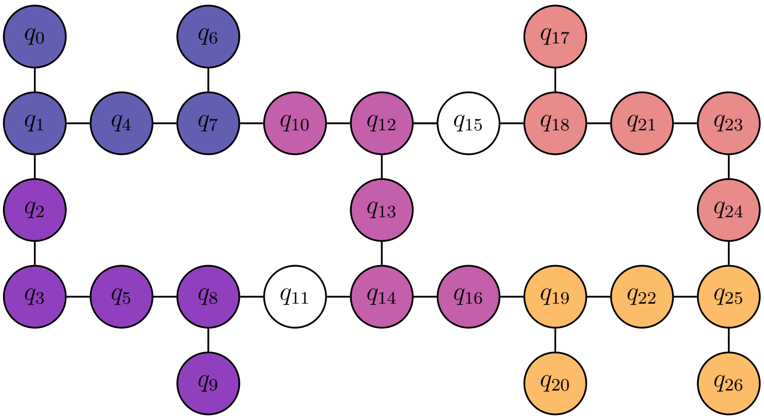


Figure 4.9: An example of circuit tiling over the IBM 27-qubit 'heavy-hex' topology found in their *Falcon* series chips. The different colours indicate replica ansatz circuits tiled across 5-qubit clusters. Not only does this increase the effective number of measurements extracted from the hardware 5-fold, but it also serves as a form of passive quantum error mitigation whereby noise is averaged over the device.

4.6 Zero Noise Extrapolation

The technique of *zero-noise extrapolation* (ZNE), also referred to in the literature as *richardson extrapolation*, operates on the principle that one may methodically amplify noise present in our quantum measurement output, obtaining a collection of increasingly noisy energy estimates before extrapolating the data and inferring the experimentally untouchable point of 'zero noise' [89, 221–227]. There are many methods of amplifying noise in our quantum circuits: some do so continuously by stretching gates temporally, requiring pulse-level control over the hardware, whereas others employ discrete approaches that either insert identity blocks of increasing complexity (e.g. unitary folding) or replace the target gate with a product over its roots.

It is the latter method we employ here. Given a quantum circuit U, some constituent native gate G and a noise parameter $\lambda \in \mathbb{N}$, we shall replace each instance of G in-circuit with the equivalent operation $\prod_{i=1}^{\lambda} \sqrt[\lambda]{G}$ to yield a noise-amplified circuit U_λ. One may note that $\lambda = 1$ corresponds with the unmodified circuit, whereas we intend to infer a value for $\lambda = 0$ by evaluating expectation values $E_\lambda = \langle\psi_{\text{ref}}| U_\lambda^\dagger P U_\lambda |\psi_{\text{ref}}\rangle$ at integer values $\lambda \in \{1, 2, 3, \ldots\}$ and extrapolating.

In particular, we take $G = \text{CNOT}$ since this is the dominant source of error by an order of magnitude, as seen in Table 5.1. In order to decompose CNOT into its roots, we define the two-qubit gate

$$\begin{aligned} \mathrm{CPhase}(\theta) &= \frac{1}{2}\big[(1+Z)\otimes I + (1-Z)\otimes P(\theta)\big] \\ &= \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & P(\theta) \end{bmatrix} \end{aligned} \tag{4.13}$$

where $P(\theta) := e^{i\theta/2}R_z(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix}$ and note that $\mathrm{CNOT} = (I \otimes H) \cdot \mathrm{CPhase}(\pi) \cdot (I \otimes H)$. In other words, the Hadamard gates applied on the target qubit diagonalize the CNOT gate and thus

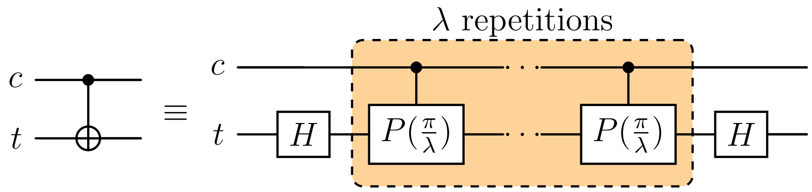
$$\sqrt[\lambda]{\mathrm{CNOT}} = (I \otimes H) \cdot \mathrm{CPhase}\Big(\frac{\pi}{\lambda}\Big) \cdot (I \otimes H). \tag{4.14}$$

The CNOT root-product decomposition is given as a circuit in Figure 4.10(a). When it comes down to implementation of zero-noise extrapolation on a quantum computer, one must be mindful of which gates are native to said device and should avoid circuit optimization routines since these may result in an unpredictable scaling of noise. For example,the CNOT is the native entangling gate on IBM Quantum *Falcon* series chips [1]; therefore, CPhase operations will be transpiled back in terms of CNOT and R_z gates at the point of execution, the decomposition of which is given in Figure 4.10(b). Such considerations can wreak havoc on zero-noise extrapolation if not controlled carefully.

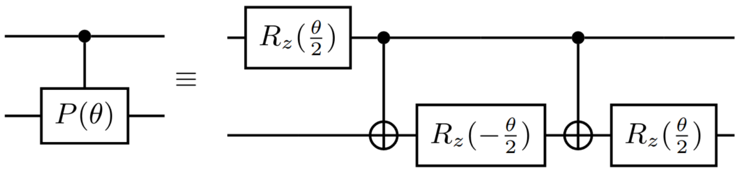
An alternative approach is a local unitary folding scheme in which one inserts CNOT pairs after each CNOT in the circuit, resulting in each being replaced with

---

[1] The CNOT was replaced by the echoed cross-resonance (ECR) gate on *Eagle r3* chips, which is equivalent to a CNOT up to single-qubit pre-rotations, and more recently the CZ gate has been adopted for the latest *Heron* architecture.



(a) Each CNOT gate may be replaced in-circuit with a product over its roots, namely $\sqrt[\lambda]{\mathrm{CNOT}_{c,t}} = H_t \mathrm{CPhase}_{c,t}(\frac{\pi}{\lambda})H_t$.



(b) Since the IBM hardware takes the CNOT as its native entangling gate, the CPhase decomposition of **(a)** is transpiled back in terms of CNOTs at the point of execution.

Figure 4.10: Noise amplification method used for zero-noise extrapolation. Given a noise amplification factor $\lambda \in \mathbb{N}$, each CNOT is replaced by 2λ CNOTs, $3\lambda + 2$ single-qubit Z-rotations and two Hadamard gates.

2λ+1 CNOT gates, compared with the 2λ encountered in our CPhase approach. In order for $\lambda \to 0$ to probe the zero-CNOT-noise regime, we would need to offset the noise amplification/gain factors in the extrapolation to account for the additional +1 CNOT of the former (as found through experimentation). By contrast, we find the CPhase decomposition to avoid the necessity of this gain offset, making for cleaner regression.

It should be noted that zero-noise extrapolation is not immune to coherent error; as stated in the introduction to Chapter 4, this is the hardest form of noise to combat since it manifests as a global bias that can be difficult to detect. Incoherent error, by contrast, is stochastic and can therefore be captured via some statistical process – such is the goal of quantum error mitigation. In Figure 4.11 we provide an example of coherent error appearing in the noise amplification of numerous 3-qubit Pauli observables, indicated by the oscillatory behaviour and thus suggestive of gate under/over rotation. This is a further consideration for ZNE when ensuring the noise amplification is conducted in a controllable manner. To produce these results we amplify to a value of $\lambda = 10$, which is far more than is typically necessary.

For our specific implementation of ZNE we shall assume that the individual noise amplified estimates have been obtained via an estimator $\mathcal{E}_\lambda$ so that $E_\lambda = \mathbb{E}(\mathcal{E}_\lambda)$, which might have previously had some other QEM strategy applied. We shall then evaluate estimates for $\lambda \in \{1, 2, 3, 4\}$ before performing weighted least squares (WLS) regression (see Section 4.1) with weights $w_\lambda = 1/\text{var}(\mathcal{E}_\lambda)$ to infer a "zero-noise" estimate $\mathcal{E}_{\text{ZNE}} = E_0$, thus penalizing highly varying points in the extrapolation. In Figure 4.12 we compare WLS against ordinary least squares

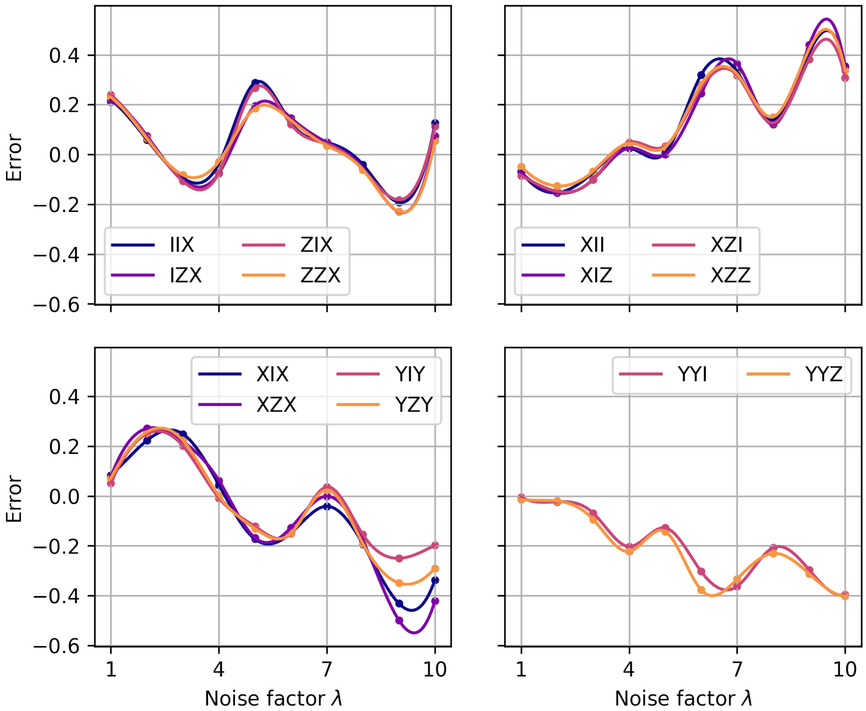


Figure 4.11: Noise amplification for 3-qubit Pauli observables on the IBM Quantum Falcon series chip *ibm_cairo*; we observe oscillatory behaviour that likely arises from coherent error caused by device miscalibration, such as over/under rotation.

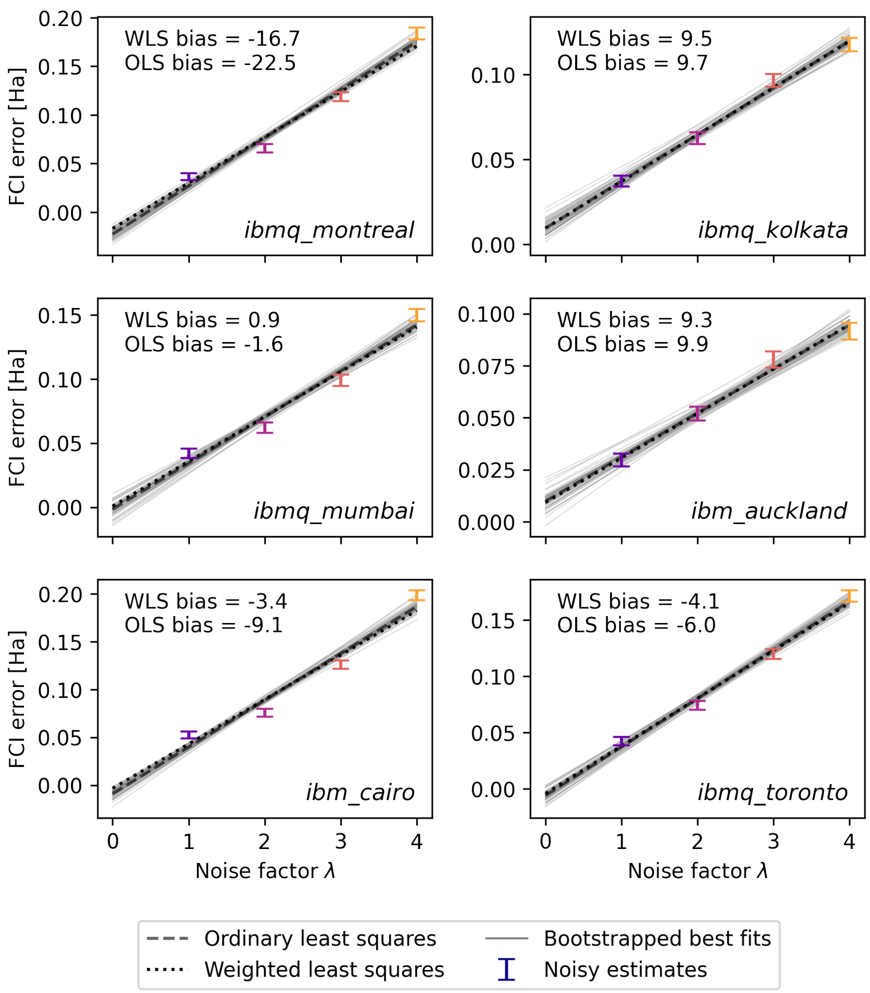


Figure 4.12: Zero-noise extrapolation for ground state preparation of the HCl STO-3G molecule at equilibrium, comparing weighted and ordinary least squares in addition to possible bootstrapped fits. Each of the noisy estimates have had measurement-error mitigation and symmetry verification applied.

(OLS) for ground state preparation of the HCl molecule (further details given in Section 5.1) and find that WLS yields the smallest bias in all cases. We note that such a regression approach allows us to quantify the success of our extrapolation via the coefficient of determination, or R^2 value, expressed as a ratio of residual and total sum of squares [228].

4.7 Echo Verification

Purification-based error mitigation techniques operate on the basis that in quantum computation we are often interested in preparing some pure state $|\psi_0\rangle = U|\mathbf{0}\rangle$, corresponding with a channel

$$\rho = \mathcal{U}(|\mathbf{0}\rangle\langle\mathbf{0}|) = U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger = |\psi_0\rangle\langle\psi_0|, \tag{4.15}$$

whereas in reality what is actually prepared on the noisy quantum hardware is some mixed state

$$\rho = \sum_{i=0}^{2^N-1} \lambda_i \rho_i \tag{4.16}$$

where $\rho_i = |\psi_i\rangle \langle\psi_i|$ and we assume $\lambda_i > \lambda_j$ for $i < j$. The central observation that purification-based methods exploit is

$$\rho^M / \operatorname{Tr}(\rho^M) \to \rho_0 \ \ (M \to \infty), \tag{4.17}$$

and the convergence is exponentially fast. This is precisely the formulation of *virtual distillation* [229], in which one prepares M copies of the mixed state ρ over disjoint quantum registers and induces their product via application of a cyclic shift operator. However, this permutation circuit is expensive and not feasible for near-term applications; the error mitigation technique we investigate here – *echo verification* (EV) [230], also referred to in the literature as *dual state purification* [231] – is closely related but may be implemented at significantly reduced cost. While the technique was first presented in the context of quantum phase estimation (QPE) [232], it was subsequently extended to the NISQ era [233]. The idea behind this method is that one prepares some quantum state, performs an intermediary readout and subsequently uncomputes the circuit before postselecting on zero measurement outcomes; this bears some resemblance to second-order virtual distillation ($M = 2$) but with the state

$$(\rho\overline{\rho} + \overline{\rho}\rho)/2 \operatorname{Tr}(\rho\overline{\rho}) \tag{4.18}$$

as opposed to form given in (4.17), with the dual state given as $\overline{\rho} = \mathcal{V}(|\mathbf{0}\rangle \langle\mathbf{0}|) := U^\dagger |\mathbf{0}\rangle \langle\mathbf{0}| U$ [231].

In EV, one prepares the pure state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(|\mathbf{0}\rangle^{\text{sys}} \otimes |+\rangle^{\text{anc}} + U^\dagger V U |\mathbf{0}\rangle^{\text{sys}} \otimes |-\rangle^{\text{anc}} \Big) \tag{4.19}$$

before postselecting on zero-measurements in the system register. Here $|\pm\rangle^{\text{anc}}$ are basis states of a single ancilla qubit. This induces the desired expectation value $\langle V\rangle = \langle\mathbf{0}| U^\dagger V U |\mathbf{0}\rangle^{\text{sys}}$ on the ancilla qubit, whose noise-free reduced state is $\rho^{\text{anc}} = |\psi\rangle \langle\psi|^{\text{anc}}$ for

$$|\psi\rangle^{\text{anc}} = \frac{1}{\sqrt{2p_{\mathbf{0}}}}\big(|+\rangle^{\text{anc}} + \langle V\rangle |-\rangle^{\text{anc}} \big) \tag{4.20}$$

where $p_{\mathbf{0}} = |(\langle\mathbf{0}| \otimes I) |\psi\rangle|^2 = \frac{1+\langle V\rangle^2}{2}$ is the probability of postselecting the zero state in the system register, depicted in Figure 4.13. Note that $p_{\mathbf{0}} \geq 1/2$. An advantage of EV over alternative error mitigation techniques is that it naturally identifies the light-cone [234, 235] of the measured observable due to a cancellation of unsupported gates owing to its dual structure, illustrated in Figure 4.14. This avoids the need to use techniques such a light-cone tracing [236] or the out-of-time-order correlator (OTOC) [237].

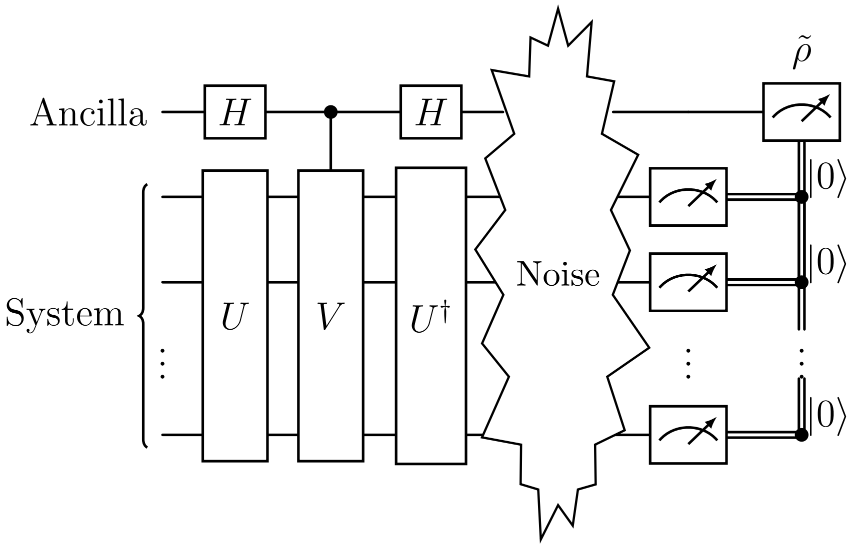


Figure 4.13: Echo Verification circuit for estimating the expectation value $\langle V\rangle = \langle\varphi| V |\varphi\rangle$ where $|\varphi\rangle = U|0\rangle$ and U, V are unitary operators; for example, V might be a Pauli operator and U some ansatz circuit. This is related to a Hadamard test for the measurement of $U^\dagger VU$ with respect to the zero state, but is distinguished through the addition of a postselection procedure. This projection onto the zero-space drives the ancilla qubit into the state (4.26) from which one may extract a noise-mitigated estimate of $\langle V\rangle$.

The quantity $\langle V\rangle$ may then be extracted from the ancilla qubit by performing X- and Z-basis measurements, which yield estimators $\mathcal{E}_X, \mathcal{E}_Z$ satisfying

$$\begin{aligned}\mathbb{E}(\mathcal{E}_X) &= \operatorname{Tr}\left(X\rho^{\text{anc}}\right) = \frac{1-\langle V\rangle^2}{1+\langle V\rangle^2}\\ \mathbb{E}(\mathcal{E}_Z) &= \operatorname{Tr}\left(Z\rho^{\text{anc}}\right) = \frac{2\langle V\rangle}{1+\langle V\rangle^2}\end{aligned} \tag{4.21}$$

and thus an unbiased estimator

$$\mathcal{E}_{\text{EV}} = \mathcal{E}_Z(1+\mathcal{E}_X)^{-1} \tag{4.22}$$

of the expectation value $\langle V\rangle = \mathbb{E}(\mathcal{E}_{\text{EV}})$. By observing that $\mathcal{E}_X^2 + \mathcal{E}_Z^2 = 1$, one may rewrite the standard EV estimator (4.22) in multiple ways [3]:

$$\begin{aligned}\mathcal{E}_{\text{EV}}^{Z-\text{bias}} &= \mathcal{E}_Z\big(1+\sqrt{1-\mathcal{E}_Z^2}\big)^{-1}\\ \mathcal{E}_{\text{EV}}^{X-\text{bias}} &= \operatorname{sign}(\mathcal{E}_Z)\sqrt{(1-\mathcal{E}_X)(1+\mathcal{E}_X)^{-1}}.\end{aligned} \tag{4.23}$$

The estimator $\mathcal{E}_{\text{EV}}^{Z-\text{bias}}$ requires only Z-basis measurements to obtain $\langle V\rangle$ when $\mathcal{E}_x \geq 0$, while $\mathcal{E}_{\text{EV}}^{X-\text{bias}}$ cannot be determined with X-basis measurements alone, since the sign information is not accessible; supplementary Z-basis readings are necessary to decide the ±1 coefficient, although significantly fewer shots suffice to indicate this sign. These alternative forms are not the same in practice due to differences in how the noise channel affects X/Z measurements on the ancilla, which is analysed formally in Section 4.7.2. As such, we may find significant disparity between $\mathcal{E}_{\text{EV}}, \mathcal{E}_{\text{EV}}^{X-\text{bias}}, \mathcal{E}_{\text{EV}}^{Z-\text{bias}}$, with one possibly displaying better performance than the others.

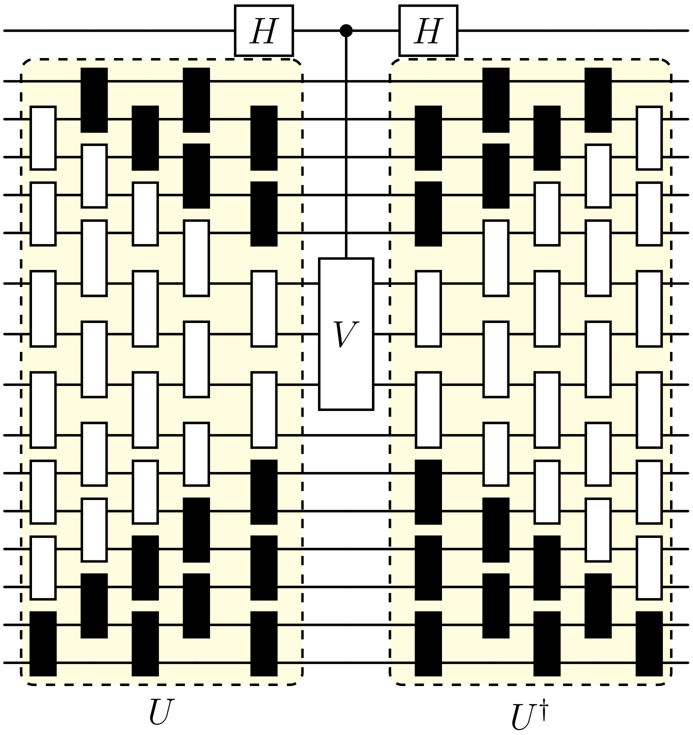


Figure 4.14: Pictorial demonstration of how the Echo Verification structure naturally identifies the light-cone for an observable. Gates lying outside the observable support (indicated here with shading) are cancelled under application of the inverse circuit and thus can lead to significant reductions in gate counts for highly localized observables. For the purposes of near-Clifford noise learning we permit a handful of rotation gates within the light-cone to assume non-Clifford rotation angles.

The only errors that are not suppressed through the EV process are those occurring in the intermediary readout phase, since errors may propagate through to the ancilla register and are not cancelled during the subsequent uncomputation. However, there is one additional trick we may employ here; if the circuit is error-free, then the state of the ancilla qubit is necessarily pure. In practice, the ancilla will be described by a mixed state

$$\rho^{\text{anc}} = (1-\epsilon)\left|\varphi_0\right\rangle\left\langle\varphi_0\right| + \epsilon\left|\varphi_1\right\rangle\left\langle\varphi_1\right| \tag{4.24}$$

where ϵ is the infidelity, which we may characterise fully via state tomography. Measuring the ancilla in the X, Y, Z bases we may reconstruct $\rho = \frac{1}{2}(I + \gamma_X X + \gamma_Y Y + \gamma_Z Z)$ where $\gamma_P = \operatorname{Tr}(P\rho^{\text{anc}})$ and identifying the largest eigenvalue with corresponding eigenvector $\left|\varphi_0\right\rangle$ we take this as an approximation to the pure state $\left|\psi\right\rangle^{\text{anc}}$ expected in the noiseless setting (Equation 4.20). Huo & Li [231] found this additional state tomography procedure to be essential in obtaining accurate results from echo verification. We later refer to this addition to EV as Tomography Purification (TP).

Furthermore, from Equation (4.21) we note that $\langle X\rangle_\phi \geq 0$, but in practice it is possible for negative value to appear from quantum experiments. In fact, the noise can be so severe that the contaminated eigenvalue of ρ dominates, resulting in spurious expectation values that can violate this non-negativity constraint maximally. This is a considerable problem when one considers the form (4.22), since this can result in division by zero, yielding a potentially infinite expectation value estimate for $\langle P\rangle_\psi$. We may combat this by always choosing the eigenvalue with positive $\langle X\rangle_\phi$, even in the case when it does not hold the greatest weight. There may also be an observed instability of the tomography purification method when $\langle Z\rangle_\phi \approx 0$, whereby the error can be increased through this procedure. Therefore, this additional step should only be executed when the raw expectation value exceeds some threshold near zero, taking the standard EV result otherwise.

A potential modification for future work would be to flip the initial state of the system register $|\mathbf{0}\rangle \to |\mathbf{1}\rangle$ via a layer of X gates and postselect on $\mathbf{1}$ measurement outcomes. While this should theoretically be no different to initializing with $|\mathbf{0}\rangle$, the effect of T_1 relaxation is for qubits to decay into the energetically favourable $|0\rangle$ state (as was observed in Figure 4.5), resulting in the erroneous postselection of invalid measurements. By flipping the initial state, we should expect to retain fewer measurements in the postselected data, but the probability of these corresponding with successful circuit runs should be improved.

In the following Sections 4.7.1 and 4.7.4 we study the EV protocol under application of several quantum channels so that we can understand how noise propagates through to the ancilla qubit and subsequently identify improvements to the technique. In Section 4.7.1 we consider the simplest setting of depolarizing noise; while it was previously argued that EV guarantees $\geq 50\%$ retention of circuit samples through the postselection procedure [231], we highlight a linear decay in the success probability $p_{\mathbf{0}}$ with the rate of depolarization to a minimum of $2/d$. This explains observations of $p_{\mathbf{0}} < 0.5$ in our previous experimental benchmarking on IBM superconducting hardware [3]. We then derive an estimator for the depolarization rate, a quantity that is ordinarily inaccessible, in terms of the postselection probability and ancilla purity, albeit with an adverse sampling overhead.

In Section 4.7.2 we study a more general noise channel consisting of arbitrary Pauli errors and discover a linear relationship between the error rates and noisy expectation values. We also note in Section 4.7.3 the limitations of the noise model used in our analyses, given that application of the noise channel is delayed to the end of the circuit as in Figure 4.13.

4.7.1 Echo Verification Under Depolarizing Noise

The standard formulation of echo verification, presented in the introduction to Section 4.7, does not take into account the effects of systematic noise; our goal is to modify the estimator (4.22) to remove any bias arising from the presence of depolarizing noise. We model this by passing the pure state $\rho = |\psi\rangle\langle\psi|$ as per Equation 4.19 through the quantum channel

$$\Phi : \rho \mapsto (1-\delta)\rho + \frac{\delta}{d} I^{\otimes N}, \tag{4.25}$$

first introduced in Equation (4.2), before application of the postselection projector $M_{\mathbf{0}} = |\mathbf{0}\rangle\langle\mathbf{0}|_{\text{sys}} \otimes I_{\text{anc}}$ with probability $p_{\mathbf{0}}(\delta) = \operatorname{Tr}(M_{\mathbf{0}}\Phi_\delta(\rho))$ and observe the propagation of noise to the ancilla qubit. Following this process, the mixed state of the ancilla qubit is

$$\begin{aligned}\rho_\delta^{\mathrm{anc}} &:= \frac{\mathrm{Tr}_{\mathrm{sys}}\left(M_{\mathbf{0}}\Phi_\delta(\rho)\right)}{\mathrm{Tr}\left(M_{\mathbf{0}}\Phi_\delta(\rho)\right)} \\ &= \frac{1-\delta}{2p_{\mathbf{0}}(\delta)}\Big[\left|+\right\rangle\left\langle+\right| + \langle V\rangle^2\left|-\right\rangle\left\langle-\right| + \langle V\rangle\big(\left|+\right\rangle\left\langle-\right| + \left|-\right\rangle\left\langle+\right|\big)\Big] + \frac{\delta}{dp_{\mathbf{0}}(\delta)}I.\end{aligned} \tag{4.26}$$

By asserting $\mathrm{Tr}\left(\rho_\delta^{\mathrm{anc}}\right) = 1$ we find there is a linear dependency between the depolarization strength and the zero postselection probability

$$p_{\mathbf{0}}(\delta) = p_{\mathbf{0}}\cdot(1-\delta) + \frac{2\delta}{d} \tag{4.27}$$

where

$$p_{\mathbf{0}} = p_{\mathbf{0}}(0) = \frac{1+\langle V\rangle^2}{2} \geq \frac{1}{2} \tag{4.28}$$

is the probability of successfully postselecting the zero state in the system register for the noiseless scenario (4.20); this has also been noted in another work [238]. Therefore, it is possible to reduce the probability of success to any value $p_{\mathbf{0}}(\delta) = p'_{\mathbf{0}}$ with $p_{\mathbf{0}} > p'_{\mathbf{0}} > \frac{2}{d}$ given a sufficiently high rate of depolarization $\delta > \frac{p_{\mathbf{0}}-p'_{\mathbf{0}}}{p_{\mathbf{0}}-2/d}$. In the worst case scenario of $\langle V\rangle = 0$ the probability of success will always be lower than 50% for non-zero depolarization, whereas for $|\langle V\rangle| = 1$ we may tolerate $\delta \leq \frac{1}{2}(\frac{1}{1-2/d}) \approx 0.5$ for large dimensions d and still retain $\geq 50\%$ of the samples.

In Figure 4.15 we observe the decay in postselection probability (4.27) for a selection of Eagle r3 chips as the circuit depth increases under time evolution of a 12-qubit Ising ring. Here, depth is taken to mean the number of echoed cross-resonance (ECR) layers, the native entangling gate on the hardware. This also serves to indicate the quality of each chip, with *ibm_sherbrooke* yielding the greatest $p_{\mathbf{0}}(\delta)$ and $\gamma(\rho^{\mathrm{anc}})$ values, whereas *ibm_cusco* is the poorest performer here; this is in agreement with IBM's error per layered gate (EPLG) metric, with stated values of 1.7% and 5.9%, respectively at the time of writing. The convergence of $p_{\mathbf{0}}(\delta)$ on the depolarizing limit of $\frac{2}{d}$ indicates that the noise channel is itself converging on the depolarizing channel, demonstrating that it describes the average effect of noise appropriately. This is problematic as it seems as though we will quickly saturate the depolarizing limit for larger systems and deep circuits, meaning the sampling overhead needs to scale inversely with the postselection probability $p_{\mathbf{0}}(\delta)$ (4.27). More precisely, for an error tolerance $\epsilon > 0$ one needs to retain $\mathcal{O}(\epsilon^{-2})$ samples through the postselection procedure and so to achieve $p_{\mathbf{0}}(\delta)\cdot n_{\mathrm{sample}} = \mathcal{O}(\epsilon^{-2})$ requires

$$n_{\mathrm{sample}} \;=\; \mathcal{O}\big(\epsilon^{-2}(1-\delta)^{-1}\big). \tag{4.29}$$

We now wish to derive a relation between the probability of success $p_{\mathbf{0}}(\delta)$ and the ancilla purity $\gamma(\rho_\delta^{\mathrm{anc}})$, with the goal of estimating the rate of depolarization δ. In order to do so, we may write the depolarized form (4.26) of the ancillary system in terms of the noiseless expression (4.20); recalling $\rho^{\mathrm{anc}} = \left|\psi\right\rangle\left\langle\psi\right|^{\mathrm{anc}}$ where $\left|\psi\right\rangle^{\mathrm{anc}} = \frac{1}{\sqrt{2p_{\mathbf{0}}}}\big(\left|+\right\rangle^{\mathrm{anc}} + \langle V\rangle\left|-\right\rangle^{\mathrm{anc}}\big)$ and by introducing the quantity $\alpha(\delta) := \frac{\delta}{dp_{\mathbf{0}}(\delta)}$, one finds

$$\rho_\delta^{\mathrm{anc}} = \big(1-2\alpha(\delta)\big)\rho^{\mathrm{anc}} + \alpha(\delta)I. \tag{4.30}$$

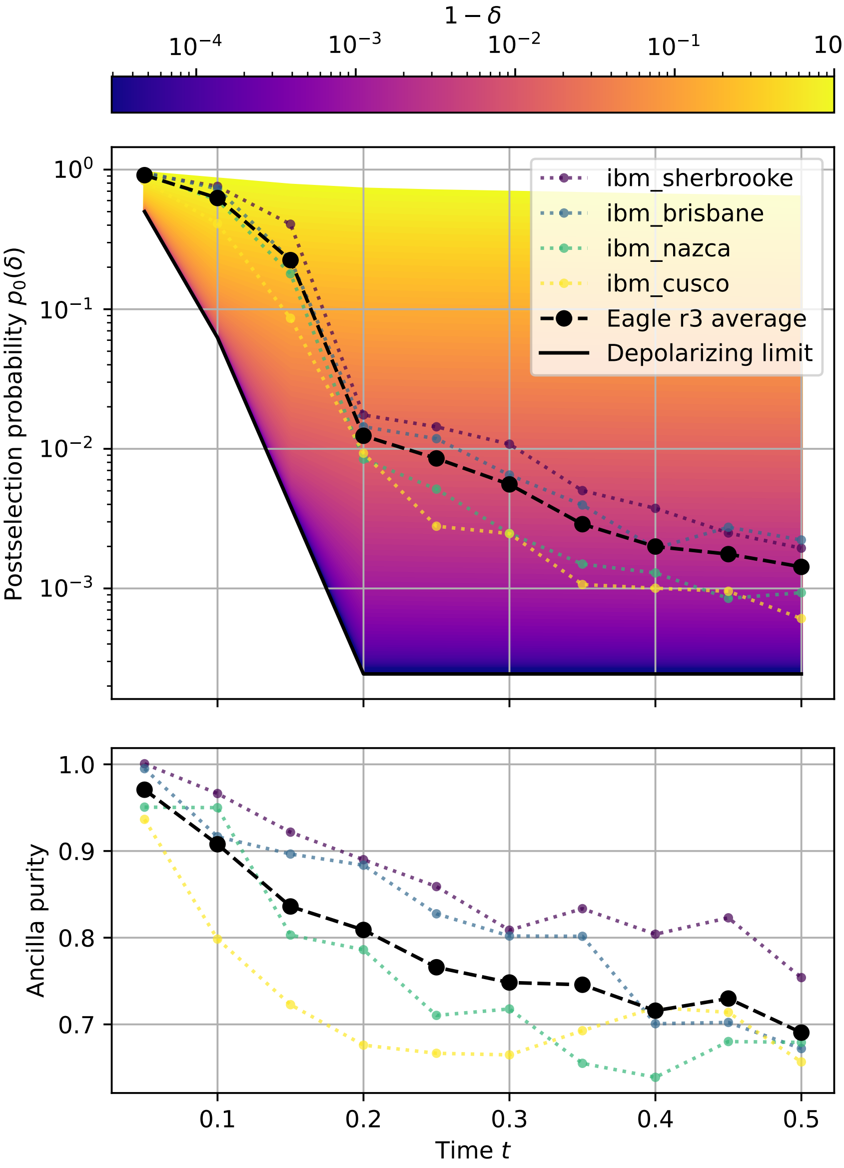


Figure 4.15: Postselection probability (4.27) and ancilla (4.26) purity over time evolution of a 12-qubit Ising ring, averaged across the Eagle r3 QPUs *ibm_cusco*, *ibm_nazca*, *ibm_brisbane* and *ibm_sherbrooke*; six circuit instances were tiled across these 127-qubit chips. The circuit depth increases linearly with the time step, thus the depolarization rate also increases and therefore we observe a decay in $p_{\mathbf{0}}(\delta)$ which will converge on the depolarizing limit of $\frac{2}{d}$. This limit is not constant as the light-cone is small for early times and is saturated from $t = 0.2$s onwards.

Using the fact that $\gamma(\rho^{\text{anc}}) = 1$, it can be seen that the purity of the ancilla state under depolarizing noise is quadratic in the newly introduced parameter $\alpha(\delta)$:

$$\gamma(\rho_\delta^{\text{anc}}) = 1 - 2\alpha(\delta) + 2\alpha(\delta)^2. \tag{4.31}$$

Solving this yields an expression for the depolarization rate in terms of the ancilla purity and postselection probability:

$$\delta = \frac{dp_{\mathbf{0}}(\delta)(1-\gamma(\rho_\delta^{\text{anc}}))}{1+\sqrt{2\gamma(\rho_\delta^{\text{anc}})-1}}. \tag{4.32}$$

The significance of this result is that δ is not ordinarily accessible. However, we have shown that it is possible to approximate the depolarization rate from a number of quantities we are able to estimate through circuit sampling; the probability of success $p_{\mathbf{0}}(\delta)$ is simply the proportion of measurements that survive the postselection procedure and the ancilla purity $\gamma(\rho_\delta^{\text{anc}})$ may be calculated by performing state tomography on the ancilla qubit, requiring just X, Y, Z measurements.

It is then possible to counteract the bias introduced by depolarizing noise, since

$$\begin{aligned} \operatorname{Tr}\left(X\rho_\delta^{\text{anc}}\right) &= \frac{1-\delta}{p_{\mathbf{0}}(\delta)} \cdot \frac{1-\langle V\rangle^2}{2} \\ \operatorname{Tr}\left(Z\rho_\delta^{\text{anc}}\right) &= \frac{1-\delta}{p_{\mathbf{0}}(\delta)} \cdot \langle V\rangle \end{aligned} \tag{4.33}$$

and therefore the standard EV estimator (4.22) yields

$$\mathbb{E}(\mathcal{E}_{\text{EV}}) = \frac{\operatorname{Tr}\left(Z\rho_\delta^{\text{anc}}\right)}{1+\operatorname{Tr}\left(Z\rho_\delta^{\text{anc}}\right)} = \langle V\rangle \cdot \frac{1-\delta}{1-\delta(1-2/d)}. \tag{4.34}$$

Finally, we obtain a depolarization-tolerant estimator

$$\mathcal{E}_{\text{EV}}^{\delta} = \mathcal{E}_{\text{EV}}\left(1+\frac{2\delta}{d(1-\delta)}\right) \tag{4.35}$$

such that $\mathbb{E}(\mathcal{E}_{\text{EV}}^{\delta}) = \langle V\rangle$ regardless of the depolarization rate δ.

The difficulty is estimating δ with sufficiently high precision, given that the dimension d grows exponentially and thus dominates this expression. The consequence is the sampling overhead scales exponentially, which is demonstrated in Figure 4.16 where we observe the estimator (4.35) error as a function of sampling budget and error rate for a 12-qubit Ising ring in an idealized model of depolarizing noise. This fact renders the above approach to EV infeasible in practice.

The next section presents analysis of EV through a Pauli channel to assess whether some other feature of the method might be exploited to derive greater suppression of estimator bias than the standard approach, while maintaining similar levels of practicality.

4.7.2 Echo Verification Under Pauli Noise

We now consider the setting of Pauli noise. We explicitly partition the qubits into the system $\mathscr{H}^{(\text{sys})}$ and ancilla $\mathscr{H}^{(\text{anc})}$ registers so that $\mathscr{H} = \mathscr{H}^{(\text{sys})} \otimes \mathscr{H}^{(\text{anc})}$ and we may write a Pauli $P \in \mathscr{B}(\mathscr{H})$ across this division $P = P^{(\text{sys})} \otimes P^{(\text{anc})}$. Since the ancillary system contains just a single qubit we have we have $\mathscr{H}^{(\text{anc})} = \mathbb{C}^2$ and $P^{(\text{anc})} \in \{I, X, Y, Z\}$, while $P^{(\text{sys})} \in \{I, X, Y, Z\}^{\otimes N}$. We note it is also possible to derive several approaches to multi-ancilla echo verification, whereby each ancillary qubit allows the calculation of a single observable expectation value, which we discuss in Section 4.7.4.

We shall associate with each Pauli operator $P_i \in \mathscr{B}(\mathscr{H})$ a probability λ_i that the error $\rho \mapsto P_i \rho P_i$ occurs; the corresponding noise channel therefore assumes the form $\Phi(\rho) := \sum_i \lambda_i P_i \rho P_i$ with $\sum_i \lambda_i = 1$. We now wish to evaluate the mixed

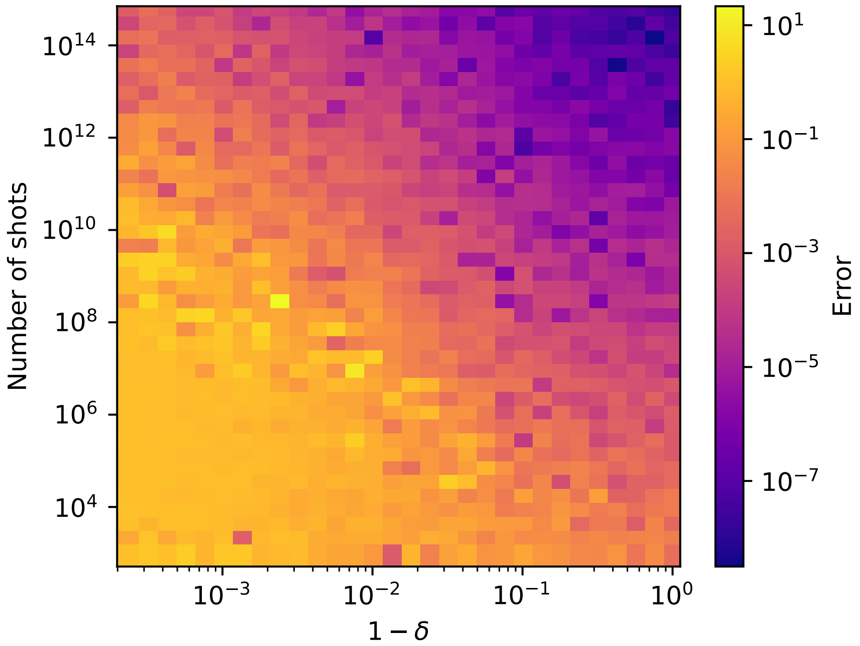


Figure 4.16: $\mathcal{E}^{\delta}_{\mathrm{EV}}$ (4.35) bias/error as a function of sampling budget and depolarization strength δ in an idealized model of depolarizing noise for an Ising ring. It may be seen in the top left of this plot that, for high levels of depolarizing noise with $\delta \approx 1$, a very large number of circuit shots is required to extract the expectation value faithfully.

state of the ancilla qubit under this channel. In echo verification we are interested in the noiseless state

$$\rho = |\psi\rangle\langle\psi| \quad \text{where} \quad |\psi\rangle = \frac{1}{\sqrt{2}}\Big(|\mathbf{0}\rangle^{(\mathrm{sys})}\otimes|+\rangle^{(\mathrm{anc})} + U^{\dagger}VU\,|\mathbf{0}\rangle^{(\mathrm{sys})}\otimes|-\rangle^{(\mathrm{anc})}\Big), \tag{4.36}$$

as per Equation (4.19). Letting $M_{\mathbf{0}} = |\mathbf{0}\rangle\langle\mathbf{0}| \otimes I$ be the projector onto the zero state of $\mathscr{H}^{(\mathrm{sys})}$, we wish to evaluate $M_{\mathbf{0}}\Phi(\rho)M_{\mathbf{0}}$ in order to understand how noise propagates from the system through to the ancilla qubit. There are four terms, labelled below ① – ④ so they may be addressed separately, that we need to consider in the expansion of $\Phi(\rho)$ for each index i:

$$\begin{aligned}
\Phi(\rho) = \sum_i \lambda_i \Big[& \\
& \textcircled{1} \leftarrow P_i^{(\mathrm{sys})}|\mathbf{0}\rangle\langle\mathbf{0}|P_i^{(\mathrm{sys})} \otimes P_i^{(\mathrm{anc})}|+\rangle\langle+|P_i^{(\mathrm{anc})} + \\
& \textcircled{2} \leftarrow P_i^{(\mathrm{sys})}U^{\dagger}VU|\mathbf{0}\rangle\langle\mathbf{0}|U^{\dagger}VUP_i^{(\mathrm{sys})} \otimes P_i^{(\mathrm{anc})}|-\rangle\langle-|P_i^{(\mathrm{anc})} + \\
& \textcircled{3} \leftarrow P_i^{(\mathrm{sys})}|\mathbf{0}\rangle\langle\mathbf{0}|U^{\dagger}VUP_i^{(\mathrm{sys})} \otimes P_i^{(\mathrm{anc})}|+\rangle\langle-|P_i^{(\mathrm{anc})} + \\
& \textcircled{4} \leftarrow P_i^{(\mathrm{sys})}U^{\dagger}VU|\mathbf{0}\rangle\langle\mathbf{0}|P_i^{(\mathrm{sys})} \otimes P_i^{(\mathrm{anc})}|-\rangle\langle+|P_i^{(\mathrm{anc})}\Big].
\end{aligned} \tag{4.37}$$

After application of the postselection operator $M_{\mathbf{0}}$, in ① we have

$$\begin{aligned}
\textcircled{1} \xrightarrow{M_{\mathbf{0}}} & |\mathbf{0}\rangle\langle\mathbf{0}|P_i^{(\mathrm{sys})}|\mathbf{0}\rangle\langle\mathbf{0}|P_i^{(\mathrm{sys})}|\mathbf{0}\rangle\langle\mathbf{0}| \otimes P_i^{(\mathrm{anc})}|+\rangle\langle+|P_i^{(\mathrm{anc})} \\
&= \delta_i^{Z}|\mathbf{0}\rangle\langle\mathbf{0}| \otimes P_i^{(\mathrm{anc})}|+\rangle\langle+|P_i^{(\mathrm{anc})}
\end{aligned} \tag{4.38}$$

where δ_i^Z is zero if $P_i^{(\text{sys})}$ contains any off diagonal Pauli errors X, Y while it is one if $P_i^{(\text{sys})}$ is of Z-type only.

Moving onto ②, we find

$$\begin{aligned} \textcircled{2} \xrightarrow{M_{\mathbf{0}}} & \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| P_i^{(\text{sys})} U^\dagger V U \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| U^\dagger V U P_i^{(\text{sys})} \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| \otimes P_i^{(\text{anc})} \left|-\right\rangle \left\langle-\right| P_i^{(\text{anc})} \\ &= |\Gamma_i|^2 \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| \otimes P_i^{(\text{anc})} \left|-\right\rangle \left\langle-\right| P_i^{(\text{anc})} \end{aligned} \tag{4.39}$$

where we have defined $\Gamma_i = \left\langle\mathbf{0}\right| U^\dagger V U P_i^{(\text{sys})} \left|\mathbf{0}\right\rangle$, which is not quite the desired expectation value $\langle V \rangle$ due to the erroneous application of a Pauli error $P_i^{(\text{sys})} \left|\mathbf{0}\right\rangle$ resulting in the ensuing overlap calculation. We do note however that $\Gamma_i = \langle V \rangle$ whenever $P_i^{(\text{sys})}$ is of Z-type error since $P_i^{(\text{sys})} \left|\mathbf{0}\right\rangle = \left|\mathbf{0}\right\rangle$ in this case, which will be useful in the following evaluation of terms ③ and ④ as we may write $\delta_i^Z \Gamma_i = \delta_i^Z \overline{\Gamma_i} = \delta_i^Z \langle V \rangle$. Explicitly,

$$\begin{aligned} \textcircled{3} \xrightarrow{M_{\mathbf{0}}} & \left|\mathbf{0}\right\rangle \underbrace{\left\langle\mathbf{0}\right| P_i^{(\text{sys})} \left|\mathbf{0}\right\rangle}_{\delta_i^Z} \underbrace{\left\langle\mathbf{0}\right| U^\dagger V U P_i^{(\text{sys})} \left|\mathbf{0}\right\rangle}_{\Gamma_i} \left\langle\mathbf{0}\right| \otimes P_i^{(\text{anc})} \left|+\right\rangle \left\langle-\right| P_i^{(\text{anc})} \\ &= \delta_i^Z \Gamma_i \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| \otimes P_i^{(\text{anc})} \left|+\right\rangle \left\langle-\right| P_i^{(\text{anc})} \\ &= \delta_i^Z \langle V \rangle \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| \otimes P_i^{(\text{anc})} \left|+\right\rangle \left\langle-\right| P_i^{(\text{anc})} \end{aligned} \tag{4.40}$$

and

$$\begin{aligned} \textcircled{4} \xrightarrow{M_{\mathbf{0}}} & \left|\mathbf{0}\right\rangle \underbrace{\left\langle\mathbf{0}\right| P_i^{(\text{sys})} \left|\mathbf{0}\right\rangle}_{\delta_i^Z} \underbrace{\left\langle\mathbf{0}\right| P_i^{(\text{sys})} U^\dagger V U \left|\mathbf{0}\right\rangle}_{\overline{\Gamma_i}} \left\langle\mathbf{0}\right| \otimes P_i^{(\text{anc})} \left|-\right\rangle \left\langle+\right| P_i^{(\text{anc})} \\ &= \delta_i^Z \overline{\Gamma_i} \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| \otimes P_i^{(\text{anc})} \left|-\right\rangle \left\langle+\right| P_i^{(\text{anc})} \\ &= \delta_i^Z \langle V \rangle \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| \otimes P_i^{(\text{anc})} \left|-\right\rangle \left\langle+\right| P_i^{(\text{anc})}. \end{aligned} \tag{4.41}$$

Finally, conjugating Equation (4.37) with the projection operator $M_{\mathbf{0}}$ yields

$$\begin{aligned} M_{\mathbf{0}} \Phi(\rho) M_{\mathbf{0}} = \; & \left|\mathbf{0}\right\rangle \left\langle\mathbf{0}\right| \otimes \sum_i \lambda_i P_i^{(\text{anc})} \Big[\delta_i^Z \left|+\right\rangle \left\langle+\right| + |\Gamma_i|^2 \left|-\right\rangle \left\langle-\right| \\ & \qquad + \delta_i^Z \langle V \rangle \Big(\left|+\right\rangle \left\langle-\right| + \left|-\right\rangle \left\langle+\right| \Big) \Big] P_i^{(\text{anc})}. \end{aligned} \tag{4.42}$$

and therefore we may read off the final state ρ^{anc} of the ancilla qubit following the postselection procedure. As a more convenient form, since $P_i^{(\text{anc})} \in \{I, X, Y, Z\}$ we may instead fix some $Q = I, X, Y, Z$ and sum internally over the 2^{2N} Pauli operators $P_i \in \{I, X, Y, Z\}^{\otimes(N+1)}$ such that the corresponding $P_i^{(\text{anc})} = Q$. This yields the alternative form for the ancilla density matrix

$$\begin{aligned} \rho^{\text{anc}} &= \sum_{Q \in \{I, X, Y, Z\}} Q \Bigg(\sum_{i : P_i^{(\text{anc})} = Q} \lambda_i \rho_i \Bigg) Q \\ &= \sum_{i=1}^{\frac{d^2}{4}} \left[\lambda_i^I \rho_i + \lambda_i^X X \rho_i X + \lambda_i^Y Y \rho_i Y + \lambda_i^Z Z \rho_i Z \right] \end{aligned} \tag{4.43}$$

where

$$\rho_i = \frac{1}{2p_{0|i}}\Big(\delta_i^Z \left|+\right\rangle \left\langle+\right| + |\Gamma_i|^2 \left|-\right\rangle \left\langle-\right| + \delta_i^Z \left\langle V \right\rangle \big[\left|+\right\rangle \left\langle-\right| + \left|-\right\rangle \left\langle+\right|\big]\Big). \tag{4.44}$$

Furthermore, by asserting $\mathrm{Tr}\left(\rho^{\mathrm{anc}}\right) = 1$ we may infer the postselection probabilities $p_{0|i} = \frac{\delta_i^Z + |\Gamma_i|^2}{2}$ conditional on a particular Pauli error P_i occurring.

Noting that $\mathrm{Tr}(PQ\rho_i Q) = (-1)^{\delta_{0,\{P,Q\}}}\,\mathrm{Tr}(P\rho_i)$, measurement of the ancilla qubit in the Z-basis under this more general noise setting yields

$$\begin{aligned}
\mathrm{Tr}\left(Z\rho^{\mathrm{anc}}\right) &= \sum_{i=1}^{\frac{d^2}{4}} \mathrm{Tr}\left(Z\rho_i\right)(\lambda_i^I - \lambda_i^X - \lambda_i^Y + \lambda_i^Z) \\
&= 2\left\langle V \right\rangle \sum_{i=1}^{\frac{d^2}{4}} \frac{\delta_i^Z}{\delta_i^Z + |\Gamma_i|^2}(\lambda_i^I - \lambda_i^X - \lambda_i^Y + \lambda_i^Z) \\
&= \frac{2\left\langle V \right\rangle}{1 + \left\langle V \right\rangle^2} \cdot \Lambda_Z
\end{aligned} \tag{4.45}$$

where

$$\Lambda_Z = \sum_{\substack{i=1 \\ P_i^{(\mathrm{sys})} \in \{I,Z\}^{\otimes N}}}^{\frac{d^2}{4}} (\lambda_i^I - \lambda_i^X - \lambda_i^Y + \lambda_i^Z). \tag{4.46}$$

The final line follows as $\Gamma_i = \left\langle V \right\rangle$ whenever $P_i^{(\mathrm{sys})}$ is strictly of Z-type error, which the Kronecker factor δ_i^Z isolates. As we can see, the Z-basis measurement yields the same as the noiseless scenario up to some damping factor Λ_Z, which comprises a sum over rates of Pauli errors that are diagonal in the system register.

Considering now the X-basis, we have

$$\begin{aligned}
\mathrm{Tr}\left(X\rho^{\mathrm{anc}}\right) &= \sum_{i=1}^{\frac{d^2}{4}} \mathrm{Tr}\left(X\rho_i\right)(\lambda_i^I + \lambda_i^X - \lambda_i^Y - \lambda_i^Z) \\
&= \sum_{i=1}^{\frac{d^2}{4}} \frac{\delta_i^Z - |\Gamma_i|^2}{\delta_i^Z + |\Gamma_i|^2}(\lambda_i^I + \lambda_i^X - \lambda_i^Y - \lambda_i^Z) \\
&= \frac{1 - \left\langle V \right\rangle^2}{1 + \left\langle V \right\rangle^2} \cdot \Lambda_X - \Omega_X
\end{aligned} \tag{4.47}$$

where

$$\begin{aligned}
\Lambda_X &= \sum_{\substack{i=1 \\ P_i^{(\mathrm{sys})} \in \{I,Z\}^{\otimes N}}}^{\frac{d^2}{4}} (\lambda_i^I + \lambda_i^X - \lambda_i^Y - \lambda_i^Z), \\
\Omega_X &= \sum_{\substack{i=1 \\ P_i^{(\mathrm{sys})} \notin \{I,Z\}^{\otimes N}}}^{\frac{d^2}{4}} (\lambda_i^I + \lambda_i^X - \lambda_i^Y - \lambda_i^Z).
\end{aligned} \tag{4.48}$$

Once again, the Pauli channel affects the expectation value linearly, however in addition to the damping factor Λ^X we find an additive error term Ω_X since $(\delta_i^Z - |\Gamma_i|^2)/(\delta_i^Z + |\Gamma_i|^2) = -1$ when $P_i^{(\mathrm{sys})} \notin \{I, Z\}^{\otimes N}$ (where $\delta_i^Z = 0$) and therefore these off-diagonal error contributions do not vanish as they did in the Z-basis measurement. For this reason, one could argue that the Z-biased variant $\mathcal{E}_{\mathrm{EV}}^{Z-\mathrm{bias}}$ (4.23) of the EV estimator should perform optimally since all the off-diagonal Pauli error is theoretically suppressed, while the standard $\mathcal{E}_{\mathrm{EV}}$ and X-biased $\mathcal{E}_{\mathrm{EV}}^{X-\mathrm{bias}}$ formulations may retain some off-diagonal noise contributions that propagate through to the ancilla qubit and only contaminate the X-basis measurements.

Finally, in the Y-basis

$$\mathrm{Tr}\,(Y\rho^{\mathrm{anc}}) = \sum_{i=1}^{\frac{d^2}{4}} \mathrm{Tr}\,(Y\rho_i)(\lambda_i^I - \lambda_i^X + \lambda_i^Y - \lambda_i^Z) = 0, \tag{4.49}$$

and thus Pauli errors alone do not lead to non-zero Y expectation value and must therefore be a consequence of sampling noise or errors outside the Pauli channel. Note that, by fixing $\lambda_i^I = \lambda_i^X = \lambda_i^Y = \lambda_i^Z = \frac{\delta}{d^2}\ \forall i \in \{1, \dots, \frac{d^2}{4}\}$ except for $\lambda_1^I = 1 - \delta + \frac{\delta}{d^2}$, we obtain $\Lambda_X = \Lambda_Z = 1 - \delta$ and $\Omega_X = 0$, thus the depolarizing channel is recovered and yields the same result as in Section 4.7.1.

Also note that the purity of the ancilla qubit satisfies

$$\gamma(\rho^{\mathrm{anc}})^2 \ \approx\ \mathrm{Tr}\,(X\rho^{\mathrm{anc}})^2 \ +\ \mathrm{Tr}\,(Y\rho^{\mathrm{anc}})^2 \ +\ \mathrm{Tr}\,(Z\rho^{\mathrm{anc}})^2 \tag{4.50}$$

where $\mathrm{Tr}\,(Y\rho^{\mathrm{anc}}) = 0$ and therefore renormalizing

$$\mathcal{E}_X \mapsto \mathcal{E}_X/\gamma(\rho^{\mathrm{anc}}), \mathcal{E}_Z \mapsto \mathcal{E}_Z/\gamma(\rho^{\mathrm{anc}}) \tag{4.51}$$

recovers the property $\mathcal{E}_X^2 + \mathcal{E}_Z^2 = 1$; this leads to a modified EV estimator that we shall refer to as *purity normalized*.

The above observation, that application of the Pauli channel in echo verification manifests as a linear relationship between the error rates and the estimation of expectation values, leads us to conclude that we may attempt to learn the noise factors $\Lambda_Z, \Lambda_X, \Omega_X$ through Clifford estimation circuits [201, 239]; we investigate this in Section 4.8.

4.7.3 Echo Verification Noise Model Limitation

We note that the noise model used for our analyses makes the assumption that all noise can be applied to the final density matrix, illustrated in Figure 4.13, rather than on a gate-by-gate basis. This really only holds true for Pauli noise applied to a Clifford circuit, therefore given an arbitrary non-Clifford circuit this simplification introduces some approximation error to the model. To illustrate this, consider a Pauli noise channel $\Phi : \rho \mapsto \sum_{i=1}^{d^2} \lambda_i P_i \rho P_i$ and a Clifford gate C. Then there exists a permutation map $\pi : \mathbb{N}_{d^2} \mapsto \mathbb{N}_{d^2}$ such that $CP_iC^\dagger = P_{\pi(i)}\ \forall i \in \mathbb{N}_{d^2}$, and therefore $CP_i = P_{\pi(i)}C$.

Applying the noise channel before application of the gate C yields

$$\begin{aligned} C\Phi(\rho)C^\dagger &= \sum_{i=1}^{d^2} \lambda_i C P_i \rho P_i C^\dagger \\ &= \sum_{i=1}^{d^2} \lambda_i P_{\pi(i)} C\rho C^\dagger P_{\pi(i)} \\ &= \sum_{i=1}^{d^2} \lambda_{\pi^{-1}(i)} P_i C\rho C^\dagger P_i \\ &= \Phi'(C\rho C^\dagger) \end{aligned} \tag{4.52}$$

where the modified channel Φ' differs from Φ by a permutation of the error rates associated with each Pauli error. Note the restricted setting of depolarizing noise, as investigated in Section 4.7.1, is the one in which we have $\Phi' \equiv \Phi$. We can apply this recursively to delay application of the noise channel to the end of the circuit.

This does not hold in general and so future work may wish to expand the EV analyses presented here to a more general noise model. In Section 4.8 we investigate (near) Clifford approximations to the target circuit, motivated by the findings of Section 4.7.2; refinement of the noise model may lead to even more effective EV methods.

4.7.4 Echo Verification with Multiple Ancilla

In this section we discuss the overhead of multi-ancilla echo verification, where each additional ancilla qubit allows a further Pauli observable to be estimated concurrently, and derive two approaches to doing so. Suppose we have a collection of Pauli operators $\{P_m\}_{m=1}^M$ whose expectation values we would like to evaluate on a state $|\psi\rangle = U|\mathbf{0}\rangle$. The state of the ancilla register produced by the circuit in Figure 4.17a has the form

$$|\varphi\rangle = \frac{1}{\sqrt{p_{\mathbf{0}}}} \sum_{\boldsymbol{b}\in\{0,1\}^M} \langle P_{\boldsymbol{b}}\rangle |\boldsymbol{b}\rangle_X \tag{4.53}$$

where $P_{\boldsymbol{b}} = \prod_{m=1:b_m=1}^M P_m$ is the product of Pauli operators P_m such that the corresponding entry in $\boldsymbol{b}$ is equal to one, i.e. $b_m = 1$; the subscript $|\cdot\rangle_X$ indicates X is the chosen computational basis. Defining by $X_{\boldsymbol{a}}$ the Pauli whose m-th qubit position is Z when $a_m = 1$ and I when $a_m = 0$, we obtain

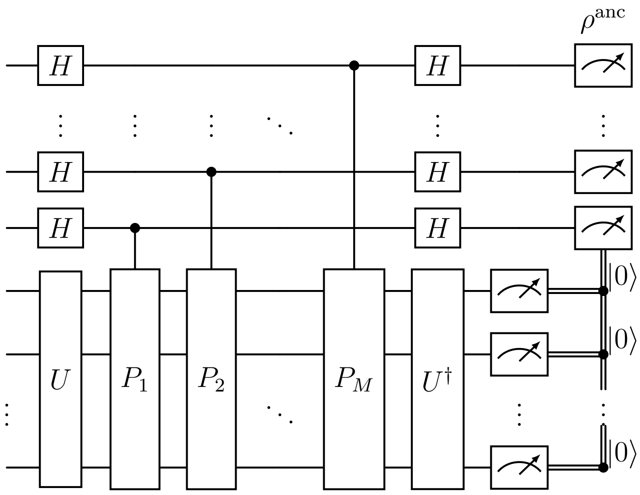
$$\langle\varphi| X_{\boldsymbol{a}} |\varphi\rangle = \frac{1}{p_{\mathbf{0}}} \sum_{\boldsymbol{b}\in\{0,1\}^M} (-1)^{\boldsymbol{a}\cdot\boldsymbol{b}} \langle P_{\boldsymbol{b}}\rangle^2 . \tag{4.54}$$

The problem is that, in order to isolate a single Pauli term P_m in the above expression, one must sum over all $X_{\boldsymbol{a}}$ such that $a_m = 1$, of which there are 2^{M-1}; we note this does not incur any additional coherent resource and is purely classical overhead. Specifically,

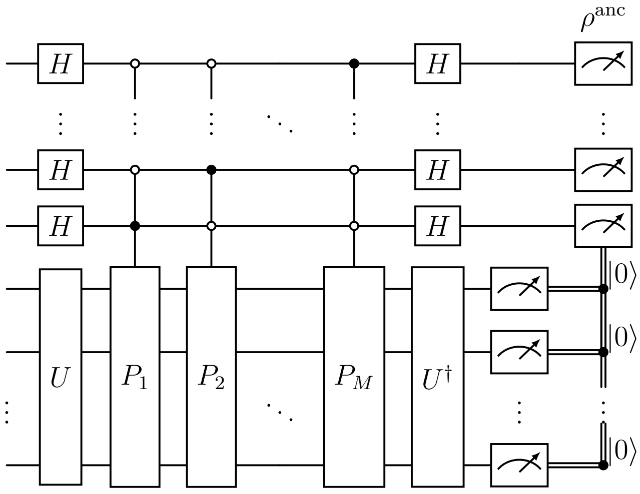
$$\frac{1 - \langle P_m\rangle^2}{p_{\mathbf{0}}} = \frac{1}{2^{M-1}} \sum_{\boldsymbol{a}\in\{0,1\}^M : a_m=1} \langle\varphi| X_{\boldsymbol{a}} |\varphi\rangle . \tag{4.55}$$

A similar expression allows us to evaluate the Z expectation values and consequently the EV estimator; the overall scaling for this approach is $\mathcal{O}(2^{2M-1})$ where M is the number of ancilla.

An alternative approach, which avoids this adverse classical postprocessing overhead, is to use multicontrols on the ancilla register as in Figure 4.17b. Now, it is possible to prepare ancilla states of the form $|\varphi\rangle = \frac{1}{\sqrt{p_0}}\Big(|0\rangle + \sum_{m=1}^{M}\langle P_m\rangle|m\rangle\Big)$ and extract the relevant expectation values with ease. However, the added gate-cost is considerable and not appropriate for near-term quantum devices; in short, the two approaches discussed in this Section offer a trade-off between easy circuit construction with exponential classical post-processing, and low-cost processing with more complex circuit structure.



(a) This circuit construction is a simple extension of standard single-ancilla EV, but results in a classical postprocessing overhead that is exponential in the number of ancilla M.



(b) This multicontrol variant avoids the adverse classical postprocessing overhead encountered for Figure 4.17a, but at the expense of greater complexity in the circuit construction.

Figure 4.17: Multi-ancilla echo verification circuits for the simultaneous estimation of multiple Pauli observables.

4.8 Echo-Verified Clifford Data Regression

The findings of Section 4.7.2 reveal the learnability of the effect of Pauli error in echo verification (EV), which one may probe using *Clifford Data Regression* (CDR) [239] to further suppress bias in the estimator through near-Clifford training circuits. The addition of CDR to the EV method produces a combined error mitigation strategy that is greater than the sum of its parts, which we shall henceforth refer to as *Echo Verified Clifford Data Regression* (EVCDR); the general workflow is outlined in Figure 4.18. We also present a practical showcase of the EVCDR method at the end of this section, in which we accurately simulate the time evolution of Ising models on heavy-hex spin-lattices consisting of 21 and 35 qubits.

The *Clifford group* consists of operators $C \in \mathscr{B}(\mathscr{H})$ satisfying $CPC^\dagger \in \mathscr{P}_N \,\forall\, P \in \mathscr{P}_N$; in other words, Clifford operations are those that map the Pauli group back onto itself, meaning it *normalizes* the Pauli group. Note also that the Pauli operators are themselves Clifford, so $\mathscr{P}_N \subset \mathscr{C}(\mathscr{H})$; additional gates that fall within the Clifford group are S, H and CNOT, which are actually sufficient to generate the full group.

Rotations $R_{x/y/z}(\theta) = \exp\left(-i\frac{\theta}{2}X/Y/Z\right)$ are generally non-Clifford, but they become Clifford at integer multiples of $\frac{\pi}{2}$, i.e. with rotation angles $\theta = \frac{k\pi}{2}$ with $k \in \mathbb{Z}$. Note that some works adopt a different rotation convention and therefore angles might differ.

An important consequence of a circuit being comprised of just Clifford gates is that it is efficiently simulable by classical means [240]; as an example, the `symmer` Python package [10] includes a basic Clifford simulator that can evaluate Hamiltonian expectation values over thousands of qubits within seconds. Furthermore, one may permit a small number of non-Clifford gates within the circuit while still maintaining classical simulability; it is this fact that we shall exploit to learn the Pauli noise present within the EV protocol.

We start with a target unitary $U(\boldsymbol{\theta})$, with $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots)$ corresponding with angles of rotation gates, for the state preparation step in EV (see Figure 4.13). We may round the parameters $\boldsymbol{\theta}$ to the nearest multiples of $\frac{\pi}{2}$ to obtain a classically simulable circuit $U(\boldsymbol{\theta}^{\text{clifford}}) \in \mathscr{C}(\mathscr{H})$ where

$$\theta_i^{\text{clifford}} = \frac{\pi}{2} \cdot \text{round}\left(\frac{2}{\pi} \cdot \theta_i\right). \tag{4.56}$$

The resulting entangling structure will match that of the target unitary and thus approximates the noise channel appropriately; this differs from the work of [201] where all single-qubit gates were dropped, leaving just that entangling structure in the Clifford estimation circuit. For a Pauli operator P whose expectation value we wish to estimate, we will have $\mathcal{C}(\boldsymbol{\theta}^{\text{clifford}}) = \langle 0| U^\dagger(\boldsymbol{\theta}^{\text{clifford}}) P U(\boldsymbol{\theta}^{\text{clifford}}) |0\rangle \in \{-1, 0, +1\}$.

We may improve the quality of our Pauli noise training data by permitting a small number $L \in \mathbb{N}$ of non-Clifford gates in the training circuits, chosen such that we maintain efficiency in evaluating the relevant expectation values. In order to do so, we select at random L parameters $\boldsymbol{\theta}^{\text{non-clifford}} = (\theta_{i_1}, \dots, \theta_{i_L})$ to leave un-rounded in the circuit, noting that this may generate up to 2^L terms in the corresponding state and thus must be set such that the available classical resource

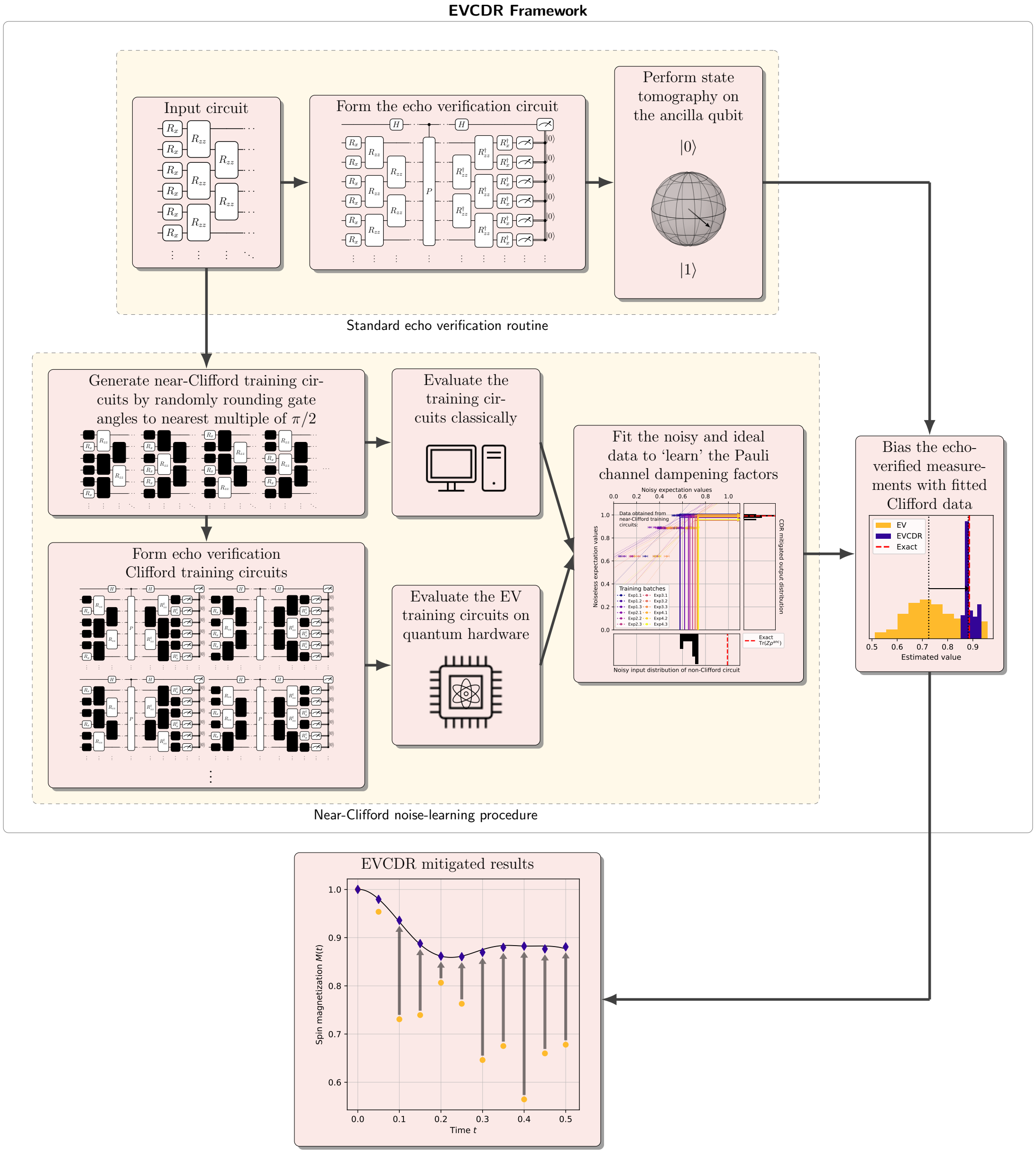


Figure 4.18: Overview of the Echo Verified Clifford Data Regression (EVCDR) framework. The results obtained from a standard Echo Verification routine are biased with a linear noise damping factor that is learned via the Clifford Data Regression procedure. This is achieved by randomly sampling near-Clifford training circuits by approximating all but a few of the rotation gates by their nearest Clifford counterpart (for example, by rounding gate angles to an integer multiple of $\frac{\pi}{2}$). By evaluating the training circuits both classically and on the quantum hardware we may learn the effect of the underlying noise channel and subsequently suppress error in the final energy estimate. We may then reflect estimates from the true non-Clifford circuit through the Clifford fitting curves to obtain CDR mitigated estimates that are then combined in the same way as the standard EV estimator.

may accommodate this overhead. We randomize over $M \in \mathbb{N}$ selections $\boldsymbol{i} = (i_1, \ldots, i_L)$, ensuring to include only choices of parameters in the light-cone circuit of $U(\boldsymbol{\theta})$ (see Figure 4.14) to avoid redundant data; we denote by $\mathcal{I}_{\text{light-cone}}$ the indexing set of such parameters. Letting $U(\boldsymbol{\theta}; \boldsymbol{i})$ be the near-Clifford circuit whose rotation angles are rounded to the nearest multiple of $\frac{\pi}{2}$, except for those having indices $\boldsymbol{i} \subset \mathcal{I}_{\text{light-cone}}$, and $\rho(\boldsymbol{\theta}, \boldsymbol{i})$ the corresponding reduced density matrix of the ancilla qubit in EV, we obtain noiseless expectation values

$$\mathcal{C} = \Big\{ \mathcal{C}(\boldsymbol{\theta}; \boldsymbol{i}) \mid \boldsymbol{i} \subset \mathcal{I}_{\text{light-cone}} \Big\} \tag{4.57}$$

and X/Z-basis noisy measurements

$$\begin{aligned} \mathcal{N}_X &= \Big\{ \operatorname{Tr}\big(X\rho(\boldsymbol{\theta}, \boldsymbol{i})\big) \mid \boldsymbol{i} \subset \mathcal{I}_{\text{light-cone}} \Big\}, \\ \mathcal{N}_Z &= \Big\{ \operatorname{Tr}\big(Z\rho(\boldsymbol{\theta}, \boldsymbol{i})\big) \mid \boldsymbol{i} \subset \mathcal{I}_{\text{light-cone}} \Big\}. \end{aligned} \tag{4.58}$$

In the previous Section 4.7.2 we derived a linear relationship between the Pauli error rates, defining learnable quantities $\Lambda_X, \Lambda_Z, \Omega_X$, and the noisy expectation values $\mathcal{N}_X, \mathcal{N}_Z$. By fitting the data

$$\left\{ \left(\frac{1 - \mathcal{C}(\boldsymbol{i})^2}{1 + \mathcal{C}(\boldsymbol{i})^2}, \mathcal{N}_X(\boldsymbol{i}) \right) \mid \boldsymbol{i} \subset \mathcal{I}_{\text{light-cone}} \right\} \tag{4.59}$$

and

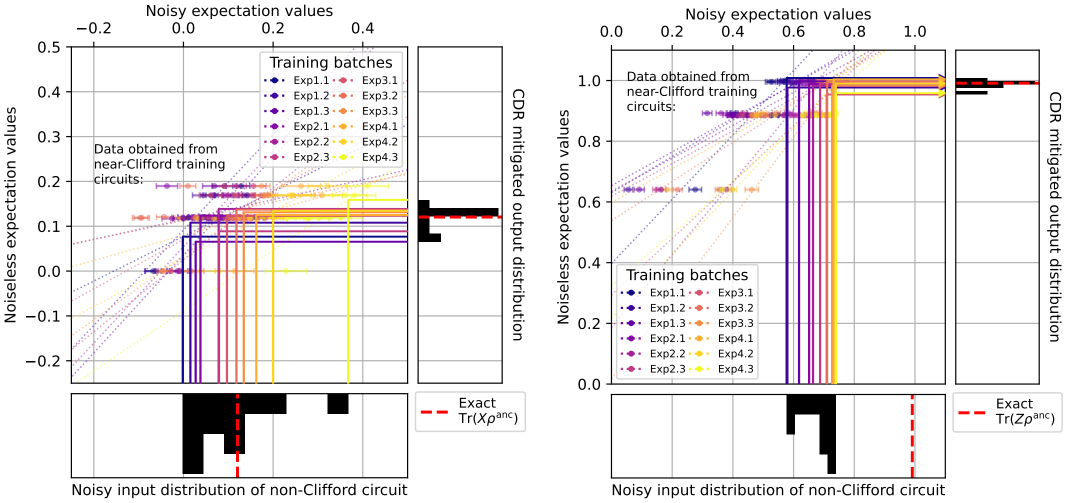
$$\left\{ \left(\frac{2\mathcal{C}(\boldsymbol{i})}{1 + \mathcal{C}(\boldsymbol{i})^2}, \mathcal{N}_Z(\boldsymbol{i}) \right) \mid \boldsymbol{i} \subset \mathcal{I}_{\text{light-cone}} \right\} \tag{4.60}$$

we obtain linear functions f_X, f_Z via standard regression techniques, in particular the bootstrapped WLS routine described in Sections 4.1.2 and 4.1.3.

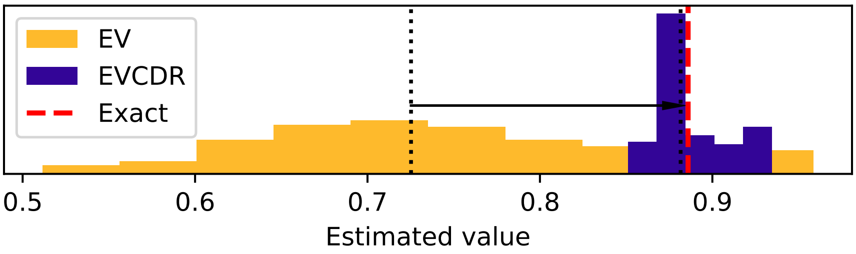
Recalling the noisy estimators $\mathcal{E}_Z, \mathcal{E}_X$ that define the standard EV estimator (4.22), we may invert the Clifford fitting curves to obtain mitigated estimators $f_Z^{-1}(\mathcal{E}_Z), f_X^{-1}(\mathcal{E}_X)$ and subsequently our Echo Verified Clifford Data Regression (EVCDR) estimator

$$\mathcal{E}_{\text{EVCDR}} = f_Z^{-1}(\mathcal{E}_Z)\Big(1 + f_X^{-1}(\mathcal{E}_X)\Big)^{-1}. \tag{4.61}$$

In Figure 4.19 we observe the intended effect of this process on the *ibm_sherbrooke* system, where we execute concurrent EVCDR procedures across numerous qubit clusters on the chip in batches. While there is significant variability in the noisy data obtained from $\mathcal{E}_X$ and $\mathcal{E}_Z$, when reflected through the Clifford fitting curves the estimates $f_X^{-1}(\mathcal{E}_X)$ and $f_Z^{-1}(\mathcal{E}_Z)$ become focused and variance is suppressed. Indeed, in performing calculations over restricted qubit subsets we find that it is possible to invert the noise channel locally.



(a) Estimating $\mathrm{Tr}(X\rho^{\mathrm{anc}})$ **(b)** Estimating $\mathrm{Tr}(Z\rho^{\mathrm{anc}})$



(c) Estimating $\langle V\rangle = \mathrm{Tr}(Z\rho^{\mathrm{anc}})/(1+\mathrm{Tr}(X\rho^{\mathrm{anc}}))$

Figure 4.19: 21-qubit example of near-Clifford circuit data used to learn **(a)** the noise damping and additive parameters Λ_X, Ω_X defined in Equation (4.48) and **(b)** the noise damping factor Λ_Z defined in Equation 4.46, corresponding with the evolution at $t = 0.15$s in Figure 5.20. Combining the ancilla expectation values provides an estimate of the desired quantity; in **(c)** we see the effects of CDR applied to EV. As shown in Figure 5.19, four 21-qubit clusters could be accommodated on the 127-qubit *ibm_sherbrooke* system, and three batches were submitted to the Runtime service, resulting in the 12 fitting curves above. A maximum of 15 non-Clifford gates were permitted in each training circuit while still allowing efficient simulation. Non-Clifford rotations were selected at random, while the remaining angles were rounded to the nearest multiple of $\frac{\pi}{2}$. The arrows define reflections through the Clifford fitting curves, yielding the final error mitigated expectation values per qubit cluster. The marginal distributions indicate the estimated expectation values for the true, non-Clifford circuit before and after CDR mitigation.

Chapter 5

Practical Implementations on Current Quantum Hardware

> Quantum phenomena do not occur in a Hilbert space, they occur in a laboratory.
>
> *Asher Peres* [241]

Much of the theoretical algorithmic speed-up from quantum computation assumes access to an ideal machine – this is far cry from the current state of quantum computation. Taking into consideration the myriad limitations facing this novel computing paradigm, such as hardware noise, short coherence times and low qubit counts with poor connectivity, any computational advantage vaporises. Before Quantum Error Correction can be realised at sufficient scale to deploy algorithms such as Quantum Phase Estimation, offering exponential speed-up over classical counterparts, there are formidable engineering challenges to overcome. Furthermore, the whole hardware ecosystem needs to be designed such that quantum computers are integrated into conventional high-performance computing platforms to meet the demand for efficient detection and subsequent correction of errors on-the-fly, requiring a highly optimized software stack. Until such challenges are resolved, we remain in the so-called Noisy intermediate-Scale Quantum (NISQ) era of quantum computation. The question of whether quantum “supremacy”, “advantage” or “utility” will be possible with NISQ computers is predominantly an open one.

In this thesis we developed a methodology for reducing the resource requirements of Hamiltonian systems, thus extending the applicability of current quantum hardware to address larger problems. In the remainder of this work we present several practical implementations in the application domains of chemistry and condensed matter, making use of the error mitigation toolkit developed throughout Chapter 4. While we make no claim of quantum advantage, we do suggest that our methodology presents a route to advantage/utility in the near-to-mid term.

5.1 Hydrogen Chloride Ground State Preparation

Broadly speaking, when applying error mitigation there is typically a trade-off between the bias of an energy estimate and our uncertainty in the result. In this section we explore this experimentally through a benchmarking of the following quantum error mitigation techniques, introduced in the previous chapter:

1. Measurement-Error Mitigation (MEM, Section 4.2);
2. Symmetry Verification (SV, Section 4.3);
3. Zero-Noise Extrapolation (ZNE, Section 4.6);
4. Echo Verification (EV, Section 4.7).
5. Echo Verification with Tomography Purification (EV+TP, Section 4.7).

In particular, we investigate combined mitigation strategies comprised of these techniques; the goal is to understand, for a fixed shot budget, the most effective approach to mitigating errors, executed across a suite of IBM quantum hardware.

The problem we take as a testbed for this QEM benchmark is preparation of the HCl molecule ground state, with the ultimate goal of measuring the corresponding energy to algorithmic accuracy (here, errors within 43 meV of full configuration interaction). Of the numerous quantum chemistry experiments performed on NISQ hardware to date, listed in Table 1.5 [3, 5, 82–112], only a select few have achieved this threshold.

The IBM Quantum 27-qubit *Falcon* series processors are equipped with the universal gate set $\{\text{CNOT}, R_z, X, \sqrt{X}\}$ and, at the time of writing, eight such devices were available to us. From the point of view of gate errors and coherence these devices were the most reliable then available through IBM Quantum [1], with the greatest Quantum Volumes (QV) [242, 243]; in Table 5.1 we provide a snapshot of the hardware specification at the point of execution of our Qiskit Runtime programs.

Taken in the minimal STO-3G basis, the full HCl problem consists of 20 qubits and therefore direct treatment is not yet feasible on current quantum computers. In order for the hardware to accommodate our problem we layered the qubit reduction techniques of tapering [136, 137] (Section 2.3) and contextual subspace [2, 5, 120, 166] (Section 3.3).

Qubit Subspace Approach

As discussed in Example 2.5, heteronuclear diatomic molecules such as HCl permit the tapering of four $\mathbb{Z}_2$ symmetries. Two of these arise from the preservation of spin up/down parities; under the Jordan-Wigner mapping [73] these manifest as $S_{\text{up/down}} = Z^{\otimes \mathcal{I}_{\text{up/down}}}$ where the sets $\mathcal{I}_{\text{up}}, \mathcal{I}_{\text{down}}$ index qubit positions encoding up (α), down (β) electron spin orbitals, respectively. These spin parity operators are still $\mathbb{Z}_2$ symmetries (i.e. single-Paulis terms) under the Bravyi-Kitaev mapping [74], however their closed form is less convenient since individual qubits do not

---

[1] The more recent *Eagle* and *Heron* chips have surpassed the decommissioned Falcon series.

Table 5.1: Breakdown of quantum hardware specification restricted to the chosen qubit cluster at the point of executing the Qiskit Runtime programs. We provide the Quantum Volume (QV), chosen 5-qubit cluster, T1/T2 times and gate duration/error for entangling (CNOT), local ($R_z, X, \sqrt{X}$) and readout operations.

			Coherence		Gate Specification		
	QV	Chosen 5q Cluster	Type	Time [μS]	Type	Time [nS]	Error $\times 10^3$
ibmq_montreal	128	{0, 1, 2, 3, 4}	**T1**:	140.92 ± 16.77	**Entangling**:	471.11 ± 78.69	7.85 ± 1.06
			T2:	82.16 ± 39.10	**Local**:	35.56 ± 0.00	0.22 ± 0.03
					Readout:	5201.78 ± 0.00	14.08 ± 2.52
ibmq_kolkata	128	{16, 19, 20, 22, 25}	**T1**:	150.92 ± 16.80	**Entangling**:	348.44 ± 177.77	5.14 ± 0.65
			T2:	135.59 ± 66.36	**Local**:	35.56 ± 0.00	0.17 ± 0.04
					Readout:	675.56 ± 0.00	10.68 ± 1.99
ibmq_mumbai	128	{0, 1, 2, 3, 4}	**T1**:	129.80 ± 28.12	**Entangling**:	556.44 ± 136.77	8.63 ± 2.20
			T2:	104.20 ± 69.62	**Local**:	35.56 ± 0.00	0.31 ± 0.17
					Readout:	3552.00 ± 0.00	18.24 ± 0.91
ibm_cairo	64	{8, 11, 13, 14, 16}	**T1**:	95.59 ± 37.85	**Entangling**:	462.22 ± 296.50	9.17 ± 4.30
			T2:	92.76 ± 69.45	**Local**:	24.89 ± 0.00	0.22 ± 0.06
					Readout:	732.44 ± 0.00	20.86 ± 11.74
ibm_auckland	64	{8, 11, 13, 14, 16}	**T1**:	162.99 ± 73.58	**Entangling**:	376.89 ± 28.61	6.28 ± 1.22
			T2:	123.67 ± 72.80	**Local**:	35.56 ± 0.00	0.23 ± 0.03
					Readout:	757.33 ± 0.00	8.30 ± 1.55
ibmq_toronto	32	{5, 8, 9, 11, 14}	**T1**:	113.96 ± 6.53	**Entangling**:	382.22 ± 61.35	7.96 ± 0.83
			T2:	171.38 ± 19.39	**Local**:	35.56 ± 0.00	0.28 ± 0.06
					Readout:	5962.67 ± 0.00	12.04 ± 4.30

represent distinct spin-orbitals. For our particular formulation of the 20-qubit HCl system with even (odd) indices encoding spin up (down) electrons we have

$$\begin{aligned} S_{\text{up}} &= ZIZIZIZIZIZIZIZIZIZI, \\ S_{\text{down}} &= IZIZIZIZIZIZIZIZIZIZ. \end{aligned} \tag{5.1}$$

We also identified two additional $\mathbb{Z}_2$ symmetries

$$\begin{aligned} S_{\sigma_h} &= IIIIIIIIZZIIIIIIZZII, \\ S_{C_2} &= IIIIIIZZZZIIIIZZZZII, \end{aligned} \tag{5.2}$$

that arise from the abelian subgroup C_{2v} of the non-abelian point group $C_{\infty v}$ (to which all heteronuclear diatomic molecules belong) generated by reflections along the molecular plane (σ_h symmetry) and rotations through an angle of 180° (C_2 symmetry). In all, with the symmetry generating set $\mathcal{S} = \{S_{\sigma_h}, S_{C_2}, S_{\text{up}}, S_{\text{down}}\}$, qubit tapering permits a reduction of 20 to 16 qubits while exactly preserving the energy spectrum.

Whereas tapering exploits physical symmetries of the Hamiltonian to remove redundant qubits, it is possible to achieve further reductions by imposing pseudo-symmetries on the system via the contextual subspace method, discussed in Section 3.3. The qubit reduction is effected by enforcing noncontextual symmetries on the contextual Hamiltonian, thus ensuring any quantum corrections are consistent with the noncontextual ground state configuration. However, the choice over which noncontextual symmetries to enforce is non-trivial. Here, we select stabilizers that preserve commutativity with the most dominant coupled-cluster amplitudes, thus maximising variational flexibility in the contextual subspace. Using this heuristic, we are able to project onto a 3-qubit contextual subspace yielding a dramatically condensed Hamiltonian

$$H = \sum_i h_i P_i \tag{5.3}$$

where we provide the explicit coefficients $h_i \in \mathbb{R}$ and Pauli terms $P_i = q_0^{(i)} \otimes q_1^{(i)} \otimes q_2^{(i)}$ in Table 5.2. The exact ground state energy of this Hamiltonian lies within 0.837 mHa (22.776 meV, close to half the target algorithmic accuracy) of the full configuration interaction (FCI) energy (-455.157Ha, calculated using PySCF [39]). Subtracting the relatively large identity term leaves a target energy of -2.066 Ha; with respect to chemical precision, this represents a challenging 0.077% error ratio that we aim to capture via QEM.

For the purposes of symmetry verification (SV, Section 4.3), we also project the particle and spin-quantum number operators of Equation (4.12) into the contextual subspace. These are not $\mathbb{Z}_2$ symmetries as they do not commute with individual terms in the Hamiltonian and are therefore nontrivial in the contextual subspace; the projection procedure respects commutation and therefore we may use the reduced operators

$$\begin{aligned} S_N &= 17 \cdot III - IIZ - \frac{1}{2}(IZI + IZZ + ZII + ZIZ), \\ S_z &= \frac{1}{4}(IZI + IZZ - ZII - ZIZ) \end{aligned} \tag{5.4}$$

for error mitigation in our HCl 3-qubit contextual subspace – as an exercise we suggest the reader confirms that these operators do indeed commute with the Hamiltonian described by the terms in Table 5.2. An interesting feature of this reduced S_N operator is the identity term; the coefficient indicates the number of particles that have been effectively enforced in the contextual subspace, in this case seventeen out of the eighteen available electrons.

Index	q_0	q_1	q_2	Coefficient	Index	q_0	q_1	q_2	Coefficient
0	I	I	I	-453.090742	17	Y	Y	X	0.035219
1	I	Z	Z	0.846721	18	I	I	X	-0.015458
2	Z	I	Z	0.846721	19	I	Z	X	0.015458
3	I	Z	I	0.620754	20	Z	I	X	0.015458
4	Z	I	I	0.620754	21	Z	Z	X	-0.015458
5	I	I	Z	0.393828	22	I	X	X	-0.009644
6	Z	Z	I	0.258369	23	I	Y	Y	-0.009644
7	Z	Z	Z	0.238049	24	Z	X	X	0.009644
8	X	Z	I	-0.061959	25	Z	Y	Y	0.009644
9	Z	X	I	0.061959	26	X	I	X	0.009644
10	Z	X	Z	-0.061959	27	X	Z	X	-0.009644
11	X	Z	Z	0.061959	28	Y	I	Y	0.009644
12	Y	Y	I	-0.055599	29	Y	Z	Y	-0.009644
13	Y	Y	Z	0.055599	30	I	X	I	0.004504
14	X	X	X	-0.035219	31	I	X	Z	-0.004504
15	X	Y	Y	-0.035219	32	X	I	I	-0.004504
16	Y	X	Y	-0.035219	33	X	I	Z	0.004504

Table 5.2: The 3-qubit contextual subspace HCl Hamiltonian, terms ordered by coefficient magnitude, we take as a testbed for the error-mitigation strategies investigated in this work.

Due to incompatibility with some of the error-mitigation techniques investigated here, we do not implement any measurement reduction strategies such as (qubit-wise) commuting decompositions or unitary partitioning [117, 119]. Instead, each Hamiltonian term is treated independently so there is zero covariance between expectation value estimates and the overall variance is therefore obtained as

$$\mathrm{Var}(H) = \sum_i h_i^2 \cdot \mathrm{Var}(P_i); \tag{5.5}$$

the statistical analysis is conducted with a bootstrapping of the raw quantum measurement data.

Before proceeding onto the quantum error mitigation benchmark, there are a few additional considerations to resolve. Firstly, one must identify a suitable ansatz circuit that is sufficiently expressible to realize the desired ground state. Secondly, we discuss the mapping of our circuits onto physical qubits, in particular for echo verification since one should be mindful of the added qubit connectivity constraints arising from parity computation stored on the ancilla qubit. Thirdly, despite not implementing any shot reduction methods in this work, we still wish to distribute the shot budget in an informed manner, preferably tailored to each device; this is the final point of discussion before moving onto the results of our benchmark.

We stress that, while eigenvalue estimation is a fundamental component of the VQE algorithm, we are not running the optimization itself on hardware, only the ground state preparation for means of benchmarking the QEM techniques presented in Chapter 4. There are additional considerations to take into account when running full VQE on quantum hardware. For example, the algorithm itself exhibits a level of robustness against noise as the optimizer can effectively combat hardware error due to the variational principle. The caveat here is that, since it is not in general possible to measure the full Hamiltonian simultaneously, the optimization may become non-variational due to state preparation discrepancies when measuring different subsets of Pauli operators.

Ansatz Construction

Initially, we tested the noncontextual projection ansatz [2] derived from the 316-term CCSD operator. The projection into the 3-qubit contextual subspace yields a 6-term excitation pool from which we identify 4 operators via qubit-ADAPT-VQE that permit chemical precision. Despite this dramatic reduction in circuit depth from the full UCCSD ansatz, the resulting noncontextual projection ansatz consists of 12 CNOT gates which we found to be prohibitive in achieving chemical precision.

To remedy this, we abandon chemical intuition in the name of hardware efficiency. It is already known that an arbitrary 3-qubit quantum state may be prepared on quantum hardware using at most 4 CNOT gates [244]. In fact, we found that only 2 CNOT gates are sufficient in constructing a 3-qubit ansatz circuit that is sufficiently expressible for our electronic structure problem, presented in Figure 5.1. In Figure 5.2 we present the outcome of a noiseless VQE simulation over this ansatz to illustrate its expressibility.

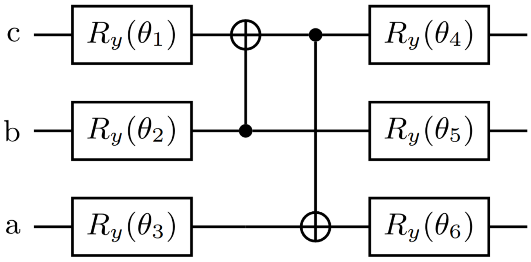


Figure 5.1: Hardware efficient HCl 3-qubit contextual subspace ansatz; the Y-rotation gates are decomposed into native gates as $R_y = \sqrt{X} R_z \sqrt{X}$. The optimal parametrization obtained from the statevector simulation in Figure 5.2 is: $\theta_1 = -0.06492667, \theta_2 = 2.89836152, \theta_3 = 0.26373807$, $\theta_4 = -0.06709062, \theta_5 = 0.01006833, \theta_6 = -0.26585046$.

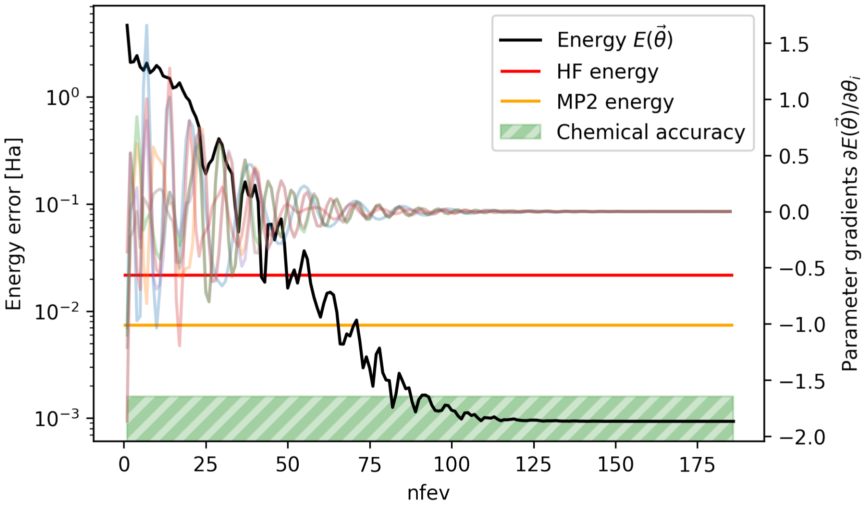


Figure 5.2: Noiseless 3-qubit CS-VQE simulation of the HCl molecule over the hardware efficient ansatz presented in Figure 5.1. The classical optimizer used is Adaptive Moment (Adam) estimation with gradients calculated using the parameter shift rule; we see that the ansatz is sufficiently expressible to achieve chemical precision.

Ancilla Readout Mapping for EV

The main bottleneck for echo verification is the ancilla readout step. Given the limited topology of the available quantum systems (Figure 4.9) and the structure of our Ansatz (Figure 5.1), it is not possible to realize every 3-qubit Pauli Z measurement basis ($IIZ, IZI, IZZ, ZII, ZIZ, ZZI, ZZZ$) without the aid of SWAP operations since at least one basis will always result in a closed loop of three CNOTs, which cannot be directly implemented on the hardware. We identified an optimal readout mapping that ensures just one measurement basis requires a SWAP operation by selecting a cluster of five qubit of the form in Figure 5.3 and implementing the readout as per Figure 5.4.

Shot Budget Distribution

To ensure a fair comparison, we define a fixed shot budget B up front and distribute according to the particular combined error-mitigation strategy. The optimal shot distribution is in proportion with $v_P := |h_P|\sqrt{\text{var}(P)}$ where $\text{var}(P) = 1 - \langle P \rangle_\psi^2$

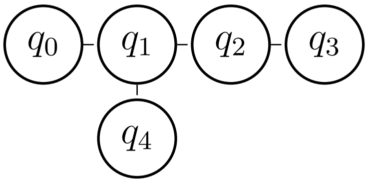
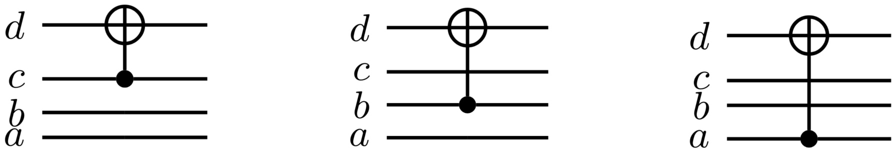


Figure 5.3: The five-qubit cluster we require for echo verification in order to facilitate readout in every possible measurement basis. There are 18 such clusters on the 27-qubit Falcon chip (see Figure 4.9) and we selected the optimal one with respect to gate and readout errors.



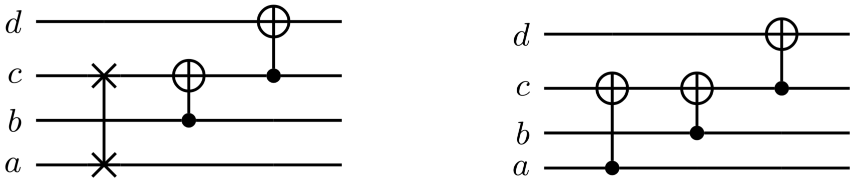
(a) IIZ
$a, b, c, d \mapsto q_0, q_2, q_1, q_4$

(b) IZI
$a, b, c, d \mapsto q_0, q_2, q_1, q_3$

(c) ZII
$a, b, c, d \mapsto q_1, q_3, q_2, q_4$

(d) IZZ
$a, b, c, d \mapsto q_0, q_2, q_1, q_4$

(e) ZIZ
$a, b, c, d \mapsto q_0, q_2, q_1, q_4$



(f) ZZI
$a, b, c, d \mapsto q_0, q_2, q_1, q_4$

(g) ZZZ
$a, b, c, d \mapsto q_0, q_2, q_1, q_4$

Figure 5.4: Ancilla readout mappings given qubit clusters of the form in Figure 5.3. Given the qubit topology of Figure 4.9 and the form of our ansatz in Figure 5.1 (where qubits a, b, c are the same as above with d the ancilla qubit), we may not entangle qubits a and b since it would result in a closed loop of three CNOT operations that is not expressible on the available quantum systems. We avoid this situation by introducing a single SWAP operation (represented in-circuit as $\times\!\!-\!\!\times$) for Hamiltonian terms of the form ZZI as in **(f)**.

[245]; however, the state-dependency means this may only be evaluated in-circuit. Therefore, we allocate $0.1\%(b = 0.001)$ of the overall budget to determine a rough estimate of the variance for each Hamiltonian term in order to rebalance the shot distribution accordingly; after this preliminary step we are left with $B' = (1-b)B$ remaining shots. For example, defining $V = \sum_P v_P$ we allow

1. ZNE: $\frac{B'v_P}{\Lambda V}$ circuit shots for each Pauli term P per noise amplification factor where Λ is the number of noisy estimates desired for the energy extrapolation procedure.

2. EV: $\frac{B'v_P}{2V}$ circuit shots for each Pauli term P, where the factor of $\frac{1}{2}$ comes

from performing both X and Z measurements over the ancilla qubit.

3. EV+ZNE: $\frac{B' v_P}{2\Lambda V}$ circuit shots for each Pauli term P per noise amplification factor.

Since the shot budget is fixed, layering multiple error-mitigation techniques may result in increased variance since fewer shots might be allocated to individual point estimates. It is the goal of this work to practically evaluate this trade-off between absolute error and uncertainty in the energy estimate, which has been noted in numerous studies [230, 246].

Simulation Results

In Table 5.3 we report the results of benchmarking our suite of error mitigation strategies for the 3-qubit HCl problem across every 27-qubit system currently available to us through IBM Quantum with a shot budget of $B = 10^6$; the order in which each QEM technique (MEM, SV, ZNE, EV, TP) appears in the combined strategy identifier indicates the order in which each method is being applied. Table 5.4 presents the average error suppression in relation to the raw estimate, calculated as

$$\left(1 - \left|\frac{\text{bias}(\mathcal{E}_{\text{QEM}})}{\text{bias}(\mathcal{E}_{\text{RAW}})}\right|\right) \times 100\%, \tag{5.6}$$

and change in standard deviation σ across our suite of systems. When $\text{bias}(\mathcal{E}_{\text{QEM}})$ is near zero, the error suppression will approach 100%, whereas values close to 0% indicate little (or no) improvement over the raw estimator; negative values of error suppression correspond with instances whereby the QEM strategy has had a detrimental effect to the energy estimate, a highly unfavourable situation.

The shot budget yields a raw standard deviation of $2 < \sigma < 3$ mHa, quantified via a bootstrapping procedure (discussed in Appendix 4.1.2). In Figure 5.5 we plot

Table 5.3: Comparison of estimator bias and standard deviation σ (given in mHa) for various error-mitigation strategies performed across a suite of IBM Quantum 27-qubit Falcon devices with a shot budget $B = 10^6$. The columns are ordered left-to-right by decreasing average error suppression (detailed in Table 5.4), with the exception of the raw estimate which is provided for reference. Note the following *classical* quantum chemistry error benchmarks: HF - 21.621mHa, MP2 - 7.360mHa, Chemical Precision - 1.6mHa.

		RAW	MEM+SV+ZNE	EV+TP	MEM+EV+TP	MEM+ZNE	MEM+SV	SV+ZNE	MEM	SV	MEM+EV+TP+ZNE	EV+TP+ZNE	ZNE	MEM+EV+ZNE	MEM+EV	EV+ZNE	EV
ibmq_montreal	**bias**	**294.7**	**16.7**	**0.8**	**6.9**	**19.0**	**30.2**	**14.9**	**85.9**	**59.2**	**61.1**	**143.2**	**240.7**	**65.6**	**72.4**	**301.9**	**208.1**
	σ	2.9	13.4	4.6	3.7	16.5	1.7	14.0	2.0	1.9	9.0	18.1	13.5	10.8	4.2	16.8	5.8
ibmq_kolkata	**bias**	**85.5**	**9.4**	**1.4**	**3.3**	**1.8**	**28.1**	**40.3**	**28.2**	**57.1**	**8.8**	**10.6**	**65.5**	**93.0**	**37.3**	**144.5**	**88.7**
	σ	2.0	2.2	7.7	6.9	7.4	1.5	1.9	1.5	1.8	17.5	18.2	4.7	17.2	6.9	18.0	7.4
ibmq_mumbai	**bias**	**235.4**	**1.4**	**37.2**	**3.3**	**36.0**	**23.8**	**18.6**	**42.9**	**44.7**	**0.6**	**54.5**	**118.1**	**146.7**	**165.1**	**332.4**	**349.8**
	σ	2.7	12.9	8.4	6.9	12.9	1.6	15.5	1.7	1.7	23.9	25.7	12.7	23.2	6.7	23.8	7.7
ibm_auckland	**bias**	**73.4**	**8.7**	**0.9**	**7.1**	**5.8**	**22.3**	**29.1**	**25.4**	**38.1**	**146.1**	**152.6**	**54.2**	**248.8**	**315.8**	**293.0**	**362.5**
	σ	2.2	2.4	6.5	5.5	2.6	1.5	0.8	1.5	1.7	15.7	15.8	1.4	16.4	6.0	16.5	6.6
ibm_cairo	**bias**	**208.9**	**4.1**	**4.8**	**4.7**	**64.4**	**7.0**	**26.5**	**6.9**	**64.0**	**19.5**	**8.9**	**99.8**	**255.2**	**200.4**	**205.4**	**242.8**
	σ	2.6	18.5	4.2	5.3	64.2	1.3	22.0	1.3	1.9	20.8	18.2	63.3	19.6	5.5	17.5	4.9
ibmq_toronto	**bias**	**125.9**	**3.6**	**24.2**	**30.8**	**18.8**	**21.4**	**11.0**	**28.6**	**37.9**	**2.3**	**125.6**	**87.7**	**8.4**	**55.0**	**123.6**	**162.4**
	σ	2.2	7.5	4.8	1.8	4.5	1.6	11.5	1.7	1.8	9.9	12.6	7.2	10.3	2.4	12.2	5.0

Table 5.4: Average error suppression and change in standard deviation under each error mitigation strategy evaluated across 27-qubit Falcon IBM Quantum devices. Ordered by decreasing mean error suppression.

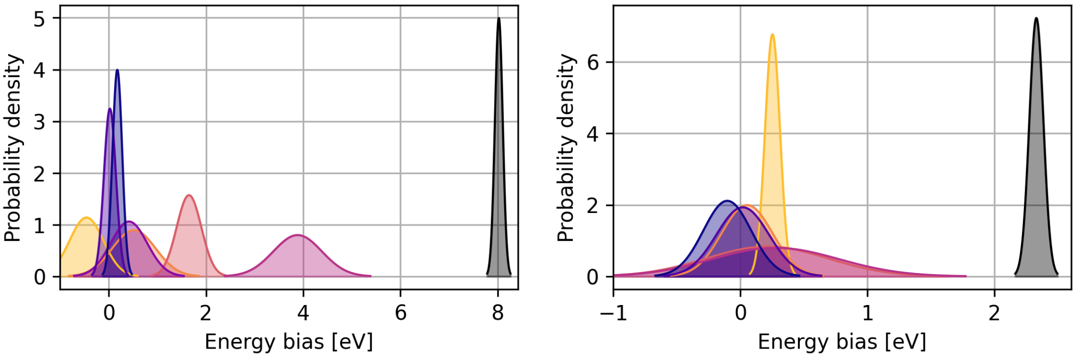
	Error Suppression [%]			Change in Std Dev		
	Mean	Best	Worst	Mean	Best	Worst
MEM+SV+ZNE	94.327	99.392	88.101	3.680	1.078	7.207
EV+TP	93.253	99.713	80.793	2.543	1.583	3.833
MEM+EV+TP	92.661	98.601	75.508	2.113	0.789	3.472
MEM+ZNE	87.094	97.877	69.185	7.069	1.202	25.063
MEM+SV	82.678	96.643	67.108	0.638	0.519	0.758
SV+ZNE	79.799	94.938	52.882	4.270	0.385	8.594
MEM	76.505	96.704	65.358	0.669	0.517	0.762
SV	63.577	80.992	33.191	0.748	0.645	0.887
MEM+EV+TP+ZNE	59.767	99.758	-98.987	6.738	3.104	8.853
EV+TP+ZNE	34.012	95.721	-107.805	7.462	5.593	9.523
ZNE	33.384	52.230	18.303	6.699	0.642	24.689
MEM+EV+ZNE	-10.180	93.343	-238.874	6.779	3.715	8.601
MEM+EV	-18.002	75.430	-330.030	2.224	1.080	3.440
EV+ZNE	-68.019	1.854	-298.966	7.235	5.453	8.989
EV	-76.967	29.366	-393.687	2.620	1.896	3.726

the bootstrap distributions for a selection of the best performing QEM strategies to illustrate the trade-off between estimator bias and variance in practice, serving as a valuable comparison with previous theoretical analyses [230].

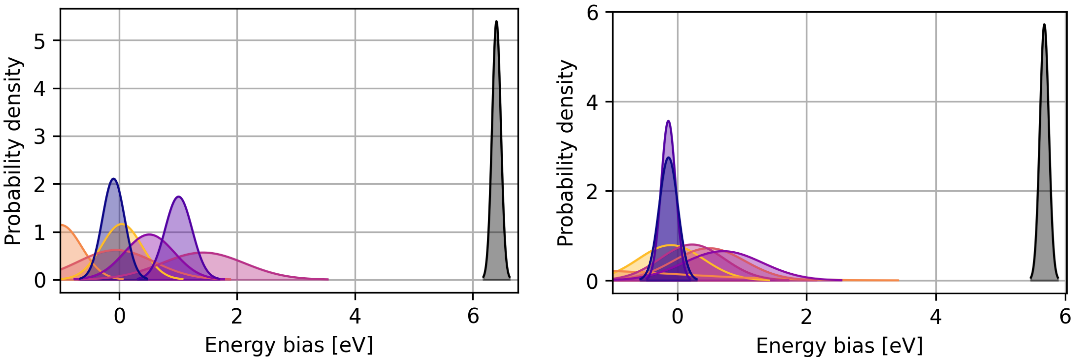
We observed that application of the MEM and SV techniques served to consistently lower both the estimator bias and standard deviation, which can be attributed to these approaches rectifying readout errors. Used in combination, the MEM+SV strategy permitted a respectable reduction in bias while also suppressing deviations with very little classical overhead.

Unlike MEM and SV, the ZNE and EV techniques necessitate modification to the quantum circuits themselves; the former, a decomposition of each CNOT gate into procedurally more complex circuit blocks, and the latter requiring a prepare-readout-invert structure with a supplementary ancilla qubit. This amounts to an increase in circuit depth for both techniques that must be taken into consideration when using them. For example, EV doubles the circuit depth and therefore one must ensure the primitive ansatz is executable within half the coherence time of the device (minus the time required for measurement). On the other hand, the increase in circuit depth arising from ZNE limits the permitted noise amplification factors, although it is not recommended to go to large factors regardless. Both of these methods can be seen to inflate the standard deviation.

By itself, EV performs very poorly (indeed, the worst four strategies were all EV-based), but when used in combination with TP we are permitted dramatic reductions in bias which exceed all other QEM strategies in the benchmark. The dependence on tomography purification for the ancilla qubit was also observed in Huo & Li [231] and is essential to obtain good results from echo verification. We stress that, although state tomography is not scalable in general, here it is applied



(a) ibmq_montreal **(b)** ibmq_kolkata



(c) ibmq_mumbai **(d)** ibm_cairo

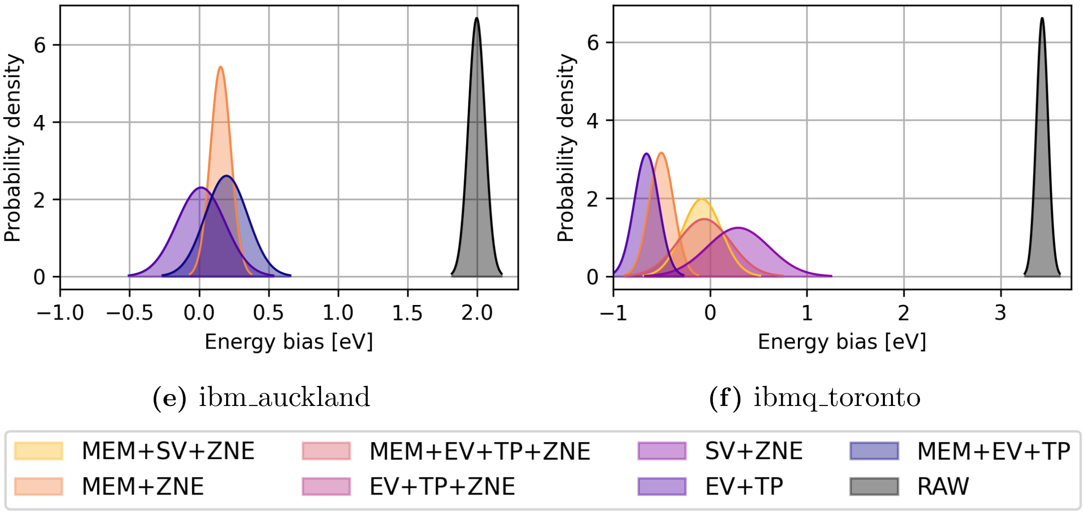


Figure 5.5: Bootstrapped distributions for some of the best QEM strategies identified through our benchmark. The mean energy of each distribution corresponds with the estimator bias.

to a single qubit and hence does not contribute a significant cost in the number of measurements required.

We found mixed success with ZNE-based strategies depending on which other QEM techniques were deployed in combination. Applied on top of MEM and SV we observed a significant improvement in error suppression. On the other hand, performing noise amplification on the ancilla qubit for the purposes of EV produced disappointing results. These observations might be attributed to coherent errors causing unpredictable noise amplification behaviour; this could have been improved by including probabilistic error cancellation [222], converting coherent error into incoherent error that may be extrapolated more confidently.

The Most Effective QEM Strategies

In this work we compared various quantum error-mitigation strategies for estimating the ground state energy of the HCl molecule on NISQ hardware. Indicated by the benchmark results of Section 5.1, we identified three hybrid strategies with the strongest performance:

- **Echo Verification with Tomography Purification (EV + TP)** yields compelling error suppression (93.253% on average) although at an increase in standard deviation (2.543 times the raw value on average); given a generous shot budget and sufficient qubit connectivity this strategy should produce reliably accurate results. Implementing echo verification requires heavy modification to the ansatz resulting in doubled circuit depth, although the errors incurred here are suppressed. Further layering measurement-error mitigation produces a similar suppression in error although the increase in standard deviation is slightly less (2.113 times the raw value on average).
- **Measurement-Error Mitigation with Symmetry Verification (MEM + SV)** comes with very low overhead yet respectable error suppression (82.678% on average) on top of a reduction in standard deviation (0.638 times the raw value on average). Furthermore, there is no required modification to the ansatz circuit since both techniques operate solely on the binary measurement output. We recommend this strategy for restrictive shot budgets or where qubit topology does not permit the readout block needed for echo verification.
- **Zero-Noise Extrapolation on top of Measurement Error Mitigation with Symmetry Verification (MEM + SV + ZNE)** is sensitive to many factors but used carefully can yield excellent results (94.327% average error suppression when we exclude the cases where extrapolation failed in Figure 4.12). There are many approaches to implementing ZNE, even extending to the pulse-level. On superconducting devices this might be preferable since it offers fine control over noise amplification. ZNE produced the largest inflation in standard deviation (3.680 times the raw value on average) and therefore a significantly greater shot budget would be necessary, due to error propagation in the extrapolation and since we evaluate several noise factors per expectation value.

As indicated by Table 5.4, each of these strategies achieved an average error suppression exceeding 80% across the suite of 27-qubit IBM Quantum chips. Given the level of noise present on these devices, reflected in the raw energy estimates, the high bar of chemical precision would necessitate a suppression of 98.783%. This was obtained for three out of eight instances of EV+TP (on the highest QV=128 systems *ibmq_montreal* and *ibmq_kolkata*, plus the QV=64 system *ibm_auckland*, with further device specifications given in Table 5.1) and a single instance of MEM+SV+ZNE (on the QV=128 system *ibmq_mumbai*), bearing in mind the standard deviation exceeds the chemically precise region and an increased shot budget would be necessary to counteract this.

From the empirical results presented in this work, it is clear that we must rely heavily on methods of quantum error mitigation if we are to obtain usable results from NISQ hardware. Through our benchmark on the IBM Quantum 27-qubit Falcon processors, we have demonstrated the most effective combined strategies which we intend to take forward in our future simulation work.

5.2 Molecular Nitrogen Dissociation Curve

For small problem instances, it is of interest to evaluate the performance of NISQ devices against various classical heuristics. A standard benchmark problem for many conventional quantum chemistry techniques is molecular nitrogen N_2 [147, 247], which is of particular interest during bond-breaking. Density matrix renormalization group (DMRG) and coupled cluster calculations were performed on N_2 in the Dunning cc-pVDZ basis set [248], and more recently using heat-bath (HCI) and quantum-selected (QSCI) configuration interaction [66]. In the dissociation limit static correlation dominates [147] and single-reference methods such as Restricted Hartree-Fock (RHF) break down; in this regime, the ground state wavefunction is not well-described by a single Slater determinant. Despite the inadequacy of the single-reference state, in the limit of all excitations post-Hartree-Fock methods such as Møller–Plesset Perturbation Theory (MP), Configuration Interaction (CI) and Coupled Cluster (CC) are still exact; however, each method requires truncation to be computationally feasible, which induces error. Furthermore, perturbation and coupled cluster approximations suffer from non-variationality [41, p. 292, 320], which is observed in the minimal STO-3G basis for the N_2 potential energy curve (PEC) in Figure 5.10.

In such scenarios, multiconfigurational methods are commonly utilized such as complete-active-space configuration interaction (CASCI) and self-consistent field (CASSCF) [249], which account for all determinants that correlate electrons in a specified number of active orbitals and thus have the flexibility to describe mixing between nearly degenerate configurations (i.e. static correlation) [250]. In Figure 5.10 we include CASCI/CASSCF calculations, in each case selecting the active space from MP2 natural orbitals; an occupation number close to zero or two indicates the corresponding spatial orbital is mostly unoccupied/occupied and can therefore be considered inactive, which naturally maximizes the correlation entropy of the wavefunction in the active space. This yields improved treatment of the bond-breaking behaviour for active spaces $(6\text{o}, 6\text{e})$ and $(7\text{o}, 8\text{e})$, while coupled cluster is more accurate around the equilibrium geometry where is is expected to perform favourably. An issue with these CAS methods is that the computational cost scales exponentially with the size of the active space and dynamical correlations outside of the active space are excluded. The missing dynamical correlation can be recovered, for example through low-order perturbations such as complete-active-space (CASPT2) or n-electron valence state (NEVPT2) second-order perturbation theory. A further problem with all these techniques is that the quality of the calculation, namely the amount of correlation energy captured, is substantially affected by the choice of active space [251], while keeping the problem computationally tractable.

Another commonly used approach to treating bond-breaking is Unrestricted Hartree-Fock (UHF), in which spin-up and spin-down orbitals are addressed separately. Sometimes, this can qualitatively describe bond dissociation; however, solutions no longer exhibit the correct spatial/spin symmetry [43], i.e. they are no longer eigenstates of the $S^2 = ||\boldsymbol{S}||^2$ operator where $\boldsymbol{S} = (S_x, S_y, S_z)$ describes the axial spin components. Since the molecular wavefunction is important to obtain observables other than energy, this represents a drawback of UHF as spin-contaminated or symmetry-broken wavefunctions are inappropriate in such cases.

Qubit Subspace Approach

In this section we invoke the Contextual Subspace approach [2, 120, 166], as developed in Section 3.3, to quantum chemistry running on superconducting devices. While we employed this technique previously for the equilibrium ground state preparation of HCl on noisy hardware [3] in Section 5.1, the variational circuit was preoptimized classically. One other work utilized the Contextual Subspace method on noisy hardware for the purposes of testing a pulse-based ansatz by calculating equilibrium energies [112]. However, only the smallest of their simulations, NH STO-3G in a 4-qubit subspace, was able to recover the Hartree-Fock energy, with correlated wavefunction methods a more challenging target. By contrast, in this work we aim to calculate the entire PEC of N_2 – not just a single point estimate – with the Contextual Subspace Variational Quantum Eigensolver (CS-VQE) running on a quantum computer; each VQE routine consists of many state preparation and gradient calculations.

We may construct contextual subspace approximations for any number of qubits between $1 - 14$, given that we first taper the molecule so a contextual subspace on 15 qubits corresponds with performing full-system VQE, as shown in Figure 5.6. Not only does the contextual subspace method allow us to reduce the number of qubits required to represent a Pauli Hamiltonian, it also has an impact on the number of terms and ℓ_1-norm of the resulting Pauli coefficients, $\Lambda = \sum_i |h_i|$, as seen in Figure 5.7. This has implications on the sampling overhead required in VQE, which scales asymptotically as $\mathcal{O}(\Lambda^2 \epsilon^{-2})$ to achieve a desired error $\epsilon > 0$ [245, 252].

It can be argued that reducing the number of qubits not only allows us to simulate larger systems on quantum hardware, but may also render such systems classically-tractable. However, the fact that Λ is additionally reduced, thus alleviating the quantum overhead further, provides a strong motivation for its use as a quantum resource reduction technique. In addition, this feature will benefit Hamiltonian simulation techniques such as qDRIFT [27], where there is an explicit quadratic scaling dependence on Λ in the resulting circuit depths.

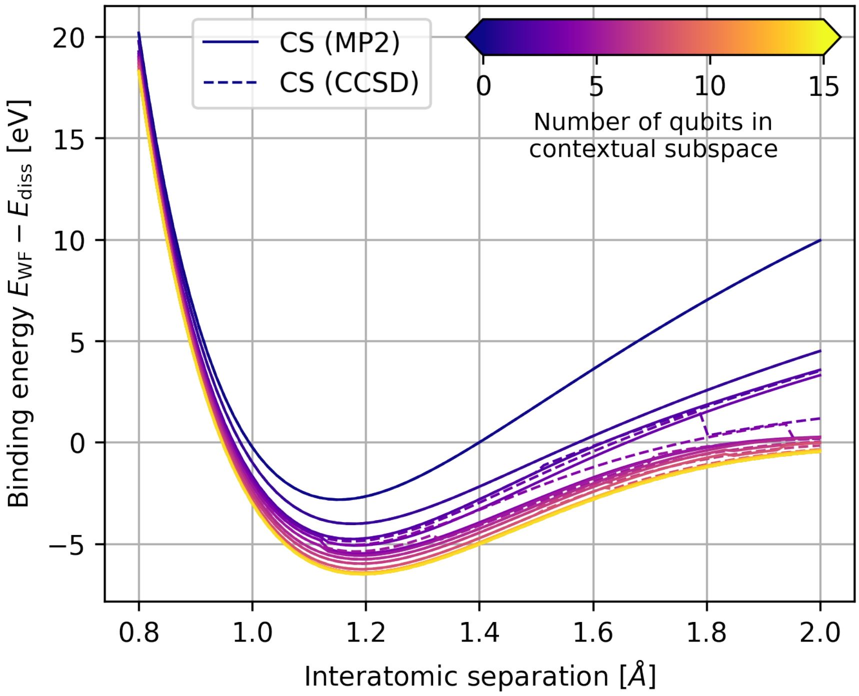


Figure 5.6: Binding potential energy curves in contextual subspaces whose corresponding stabilizers are informed either by the MP2 or CCSD wavefuntions. In the latter we encounter a large number of discontinuities that are largely absent for the MP2-informed subspaces, although the absolute error is typically lower. At 15-qubits we recover the full space and the exact FCI energy is obtained.

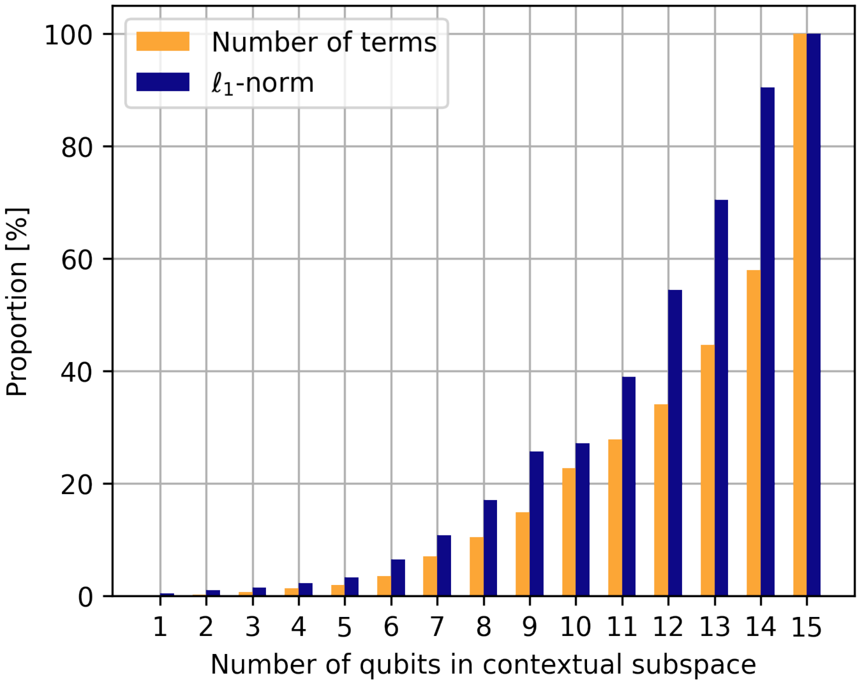


Figure 5.7: Proportion of the number of terms and ℓ_1-norm of the reduced contextual subspace Hamiltonians versus the full system at the equilibrium bond length $r = 1.192$Å.

The contextual subspace approximation is also compatible with more advanced measurement-reduction methodologies; in previous work we studied its use in combination with unitary partitioning [120], although we did not implement it for this experiment as a unitary must be applied in-circuit to realize the measurement of each anticommuting clique [117–119]. If combined with additional techniques for reducing the number of terms in the Hamiltonian, such as tensor hypercontraction [253], this could present a compelling quantum resource management framework.

Increasing the number of qubits in the contextual subspace increases the accuracy of the method. For N_2, in order to achieve algorithmic accuracy (terminology introduced in [38] and taken here to mean errors within 43 meV of FCI, with *chemical* accuracy a common misnomer when working within minimal basis sets since it implies agreement with experimental results) throughout the full PEC under the contextual subspace approximation, we need 11/12-qubits motivated by the CCSD/MP2 wavefunctions, respectively. However, increasing the number of qubits in the contextual subspace also increases the depth of the ansatz circuit and hence exposes us to the vulnerabilities of hardware noise. In Figure 1.14 of Section 1.4.5 we presented the results of running noiseless qubit-ADAPT-VQE [121, 131] over a 12-qubit subspace and observed the decay of error against the number of CNOT gates in the ansatz circuit; such circuits are too deep to obtain satisfactory results on the available hardware. There is a trade-off between a sufficiently large contextual subspace to represent the problem accurately and a sufficiently shallow ansatz circuit such that the output is not overly contaminated by noise.

We were able to achieve algorithmic accuracy on quantum hardware in Section 5.1 for the equilibrium ground state of HCl [3], where just 3-qubits were sufficient and hence a shallow ansatz was possible. A 12-qubit ansatz circuit would be too deep – and consequently too noisy – to achieve this level of accuracy on current hardware. Since algorithmic accuracy is too challenging a target for a 12-qubit simulation on real hardware, we relax this requirement. Instead we choose a contextual subspace that is sufficiently large to challenge a set of classical methods for N_2. We compare against RHF, MP2, CISD, CCSD, CCSD(T), CASCI and CASSCF for active spaces of varying size. In reproducing the PEC of N_2 we find that a 5-qubit contextual subspace, while not algorithmically accurate, yields errors that do not exceed 1 eV, as shown in Figure 5.10. It should be highlighted that the above classical techniques do not maintain algorithmic accuracy throughout the PEC either.

Simulation Results

We perform Contextual Subspace Variational Quantum Eigensolver (CS-VQE) experiments for ten points along the binding potential energy curve (PEC) of N_2 STO-3G, evenly spaced between $0.8\text{Å} - 2\text{Å}$. The extrapolation data extracted from various IBM Quantum devices used to obtain energy estimates is visible in Figure 5.8. To improve the noise extrapolation beyond the experimental results of Figure 4.12 in Section 4.6, we calibrated the noise amplification/gain factors $G(\lambda)$ using the one- and two-qubit gate error data extracted from the hardware at the time of execution; this is why the extrapolation data do not lie on integer values in Figures 5.9 and 5.8. For more reliable extrapolation, we employ inverse variance

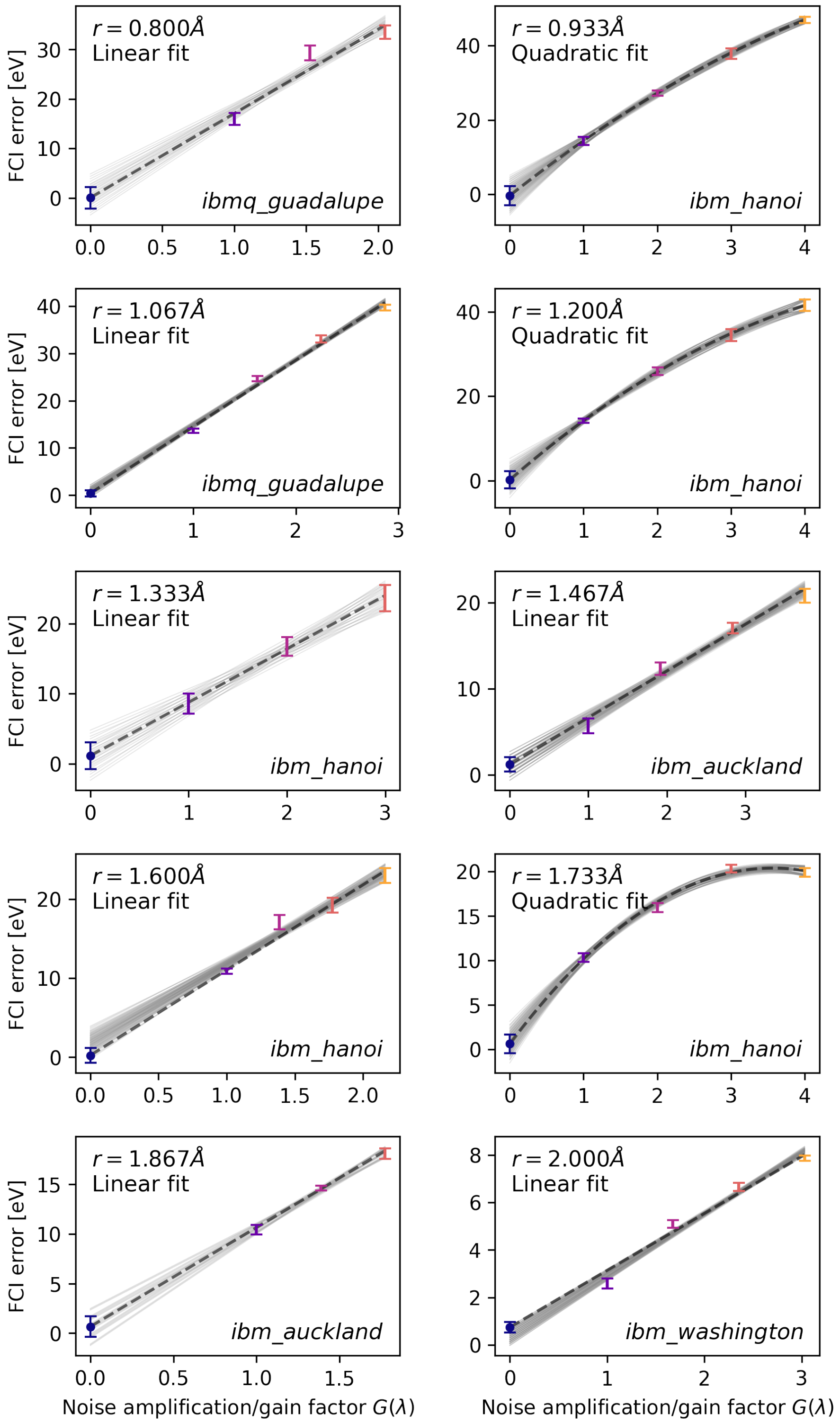
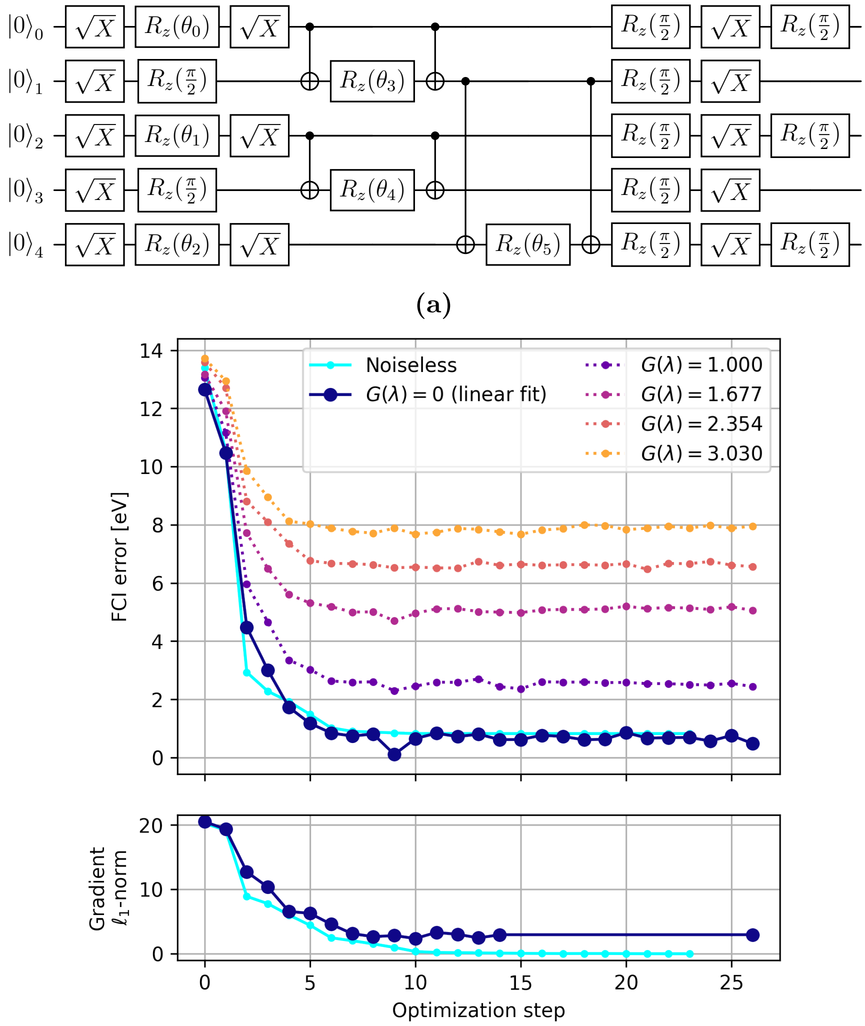


Figure 5.8: Noise fitting curves for ten evenly spaced interatomic separations of molecular nitrogen. Standard deviations were taken over the converged data in VQE in order to use weighted least squares in the linear/quadratic regression. We also plotted the spread of possible extrapolation curves given the variance of each individual noise amplified estimate. The noise amplification factors themselves were calibrated using one- and two-qubit gate error data extracted from the hardware at the time of execution.

weighted least squares regression (linear or quadratic), so that highly varying data points are penalised in the fitting procedure; the variances here are obtained from the converged VQE data, as opposed to the statistical bootstrapping procedure we used in our previous ZNE work [3]. In Figure 5.9 we present a full VQE routine executed on the *ibm_washington* system, complete with the noise amplified data that leading to the final extrapolated estimate.

The final PEC results for N_2 STO-3G may be viewed in Figure 5.10. Alongside our experimental results, we include the following classical benchmarks: RHF, MP2, CISD, CCSD, CCSD(T), CASCI and CASSCF. The active spaces of the latter two were selected using MP2 natural orbitals for fairness, since this is comparable to how the contextual subspaces are chosen as described above in Figure 5.6. We included active spaces of varying sizes, specifically (4o,2e), (5o,4e), (6o,6e)



(a)

(b)

Figure 5.9: **(a)** ansatz circuit consisting of six variational parameters and **(b)** noise amplified VQE routine for molecular nitrogen at $r = 2$Å on *ibm_washington*. We also include a noiseless routine for comparison and note the partial derivatives converge on zero in the noiseless simulation, while they are prevented from doing so in the noisy case indicated by non-zero gradient ℓ_1-norm, despite the optimizer having converged on the ground state energy.

and (7o,8e) where (Mo,Ne) denotes N electrons correlated in M spatial orbitals. A crucial point to note when comparing our CS-VQE results to the CAS methods is that, for an active subspace of M spatial orbitals, one needs $2M$ qubits to represent the problem on a quantum computer; therefore, our chosen active spaces range from 8 to 14 qubits in size, while the contextual subspace consists of just 5 qubits. This is important to bear in mind when interpreting the results.

In Figure 5.10 we see the single-reference quantum chemistry techniques - RHF, MP2, CISD, CCSD and CCSD(T) - struggling to capture the FCI energy for N_2. This holds especially true in the dissociation limit where there is no agreement between the different approaches. While the conventional techniques yield relatively low error around the equilibirum length at 1.192Å (albeit not within the target algorithmic accuracy of 43 meV), they incur large error at stretched bond lengths due to a failure of restricted Hartree-Fock to describe static correlation. Furthermore, we see instances of non-variationality, which becomes apparent at 1.140Å for MP2, 1.706Å for CCSD(T) and 1.728Å CCSD. For the CAS methods, we do not capture the bond breaking appropriately until the active space is expanded to (6o,6e) or (7o,8e), corresponding with 12 and 14 qubit subspaces.

On the other hand, our 5-qubit CS-VQE hardware experiments produce mean errors between 47 meV and 1.2 eV throughout the evaluated interatomic separations and remain below 1 eV for all but two of the bond lengths (1.333Å and 1.467Å). From direct diagonalization (the CS-DD curve in Figure 5.10) we see a true error range of 0.5 eV to 0.8 eV along the N_2 PEC. Our quantum simulations outperform all the single-reference techniques in capturing the bond dissociation behaviour; indeed, beyond an interatomic separation of 1.351Å the 5-qubit contextual subspace calculation yields lower errors than CISD and MP2, and outperforms CCSD(T) after 1.834Å, the 'gold standard' of quantum chemistry.

Turning now to the multiconfigurational approaches, our CS-VQE experiments produce lower error than CASCI/CASSCF (4o,2e) and (5o,4e) for every bond length, despite them corresponding with 8- and 10-qubit subspaces. In order for CAS to capture the dissociation satisfactorily, it needs at least the (6o,6e) space to describe the triple bond between nitrogen atoms appropriately; this is precisely what we find in Figure 5.10. While the (6o,6e) and (7o,8e) calculations do yield improved errors, particularly towards the dissociation limit, we stress that they correspond with 12- and 14-qubit subspaces. Contextual subspaces of the same size yield considerably lower error than CAS in this instance (assessed through direct diagonalization), albeit caveated with the added challenge of running hardware experiments of that scale, which could prohibit us from achieving this in practice.

Our energy advantage in the dissociation limit can be attributed to the noncontextual energy component of the CS-VQE simulations. Around the equilibrium length there is negligible difference between the Hartree-Fock and noncontextual energy, but as the bond is stretched we see the noncontextual approximation outperforming Hartree-Fock, even before the inclusion of contextual corrections obtained from VQE simulations. Interestingly, the noncontextual energy coincides exactly with the CASCI/CASSCF (4o,2e) curve between bond lengths 1.328Å and 1.909Å. This is because the noncontextual problem can accommodate a ground state that is multireference in nature, thus capturing the separated atom limit more appropriately than the single-reference RHF state here. We note that the CS-VQE optimization is still initialized in the Hartree-Fock state and therefore

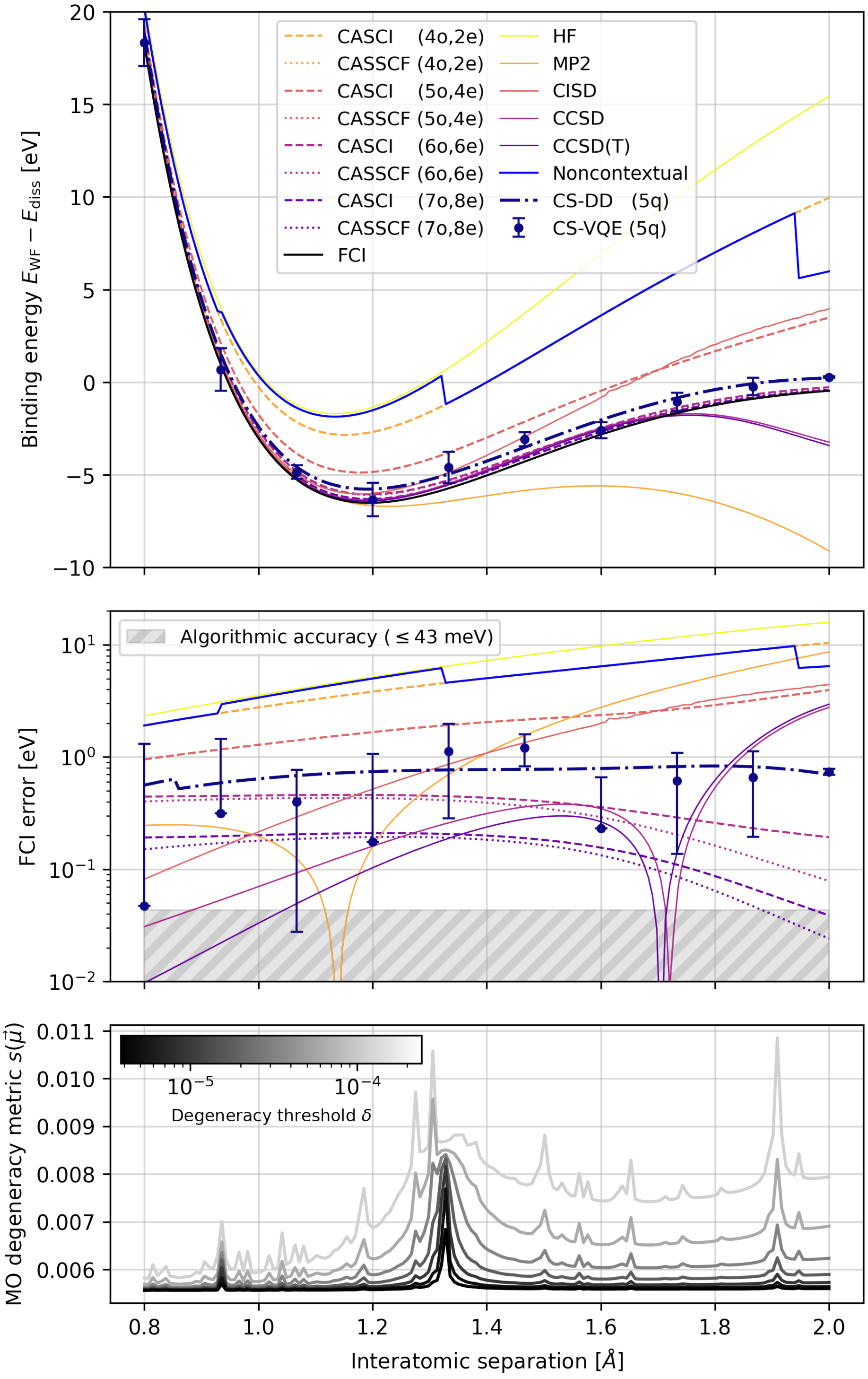


Figure 5.10: Binding potential energy curve for molecular nitrogen, N_2. The CS-VQE data points were evaluated on IBM Quantum hardware, while CS-DD corresponds with direct diagonalization of the five-qubit contextual subspaces. The quantum simulations maintain good agreement with the full configuration interaction energy throughout the entire range of interatomic separations, outperforming all the single-reference methods in the dissociation limit and remaining competitive with CASCI/CASSCF at a considerable saving of qubits. Discontinuities in the noncontextual energy coincide with peaks in our molecular-orbital degeneracy detection function (5.7). Bars indicate standard error on the mean.

does not receive an unfair advantage from this feature of the method; instead, the noncontextual contribution is included in the construction of the contextual Hamiltonian as a constant shift.

Curiously, the noncontextual PEC is not continuous and these error improvements are encountered in sharp decreases of energy, as seen in Figure 5.10 for interatomic separations $0.936\text{Å}, 1.328\text{Å}$ and 1.909Å. In order to probe this effect, we search for degeneracy in the energy levels between molecular orbitals (MO), which is known to cause issues for MP [254, 255]. This is achieved by detecting near-zero energy differences between elements of $\boldsymbol{\mu}$, a vector with length the number of orbitals M whose entries are the canonical MO energies computed through Hartree-Fock. Our candidate MO degeneracy detection function is

$$s_\delta(\boldsymbol{\mu}) := \frac{\delta\sqrt{\pi}}{2(D_{\max}-1)} \sum_{i=2}^{D_{\max}} \frac{1}{\binom{M}{i}} \left[\sum_{1 \leq j_1 < \cdots < j_i \leq M} \frac{\operatorname{erf}\left[(\mu_{j_i} - \sum_{k=1}^{i-1} \mu_{j_k})/\delta\right]}{(\mu_{j_i} - \sum_{k=1}^{i-1} \mu_{j_k})} \right] \tag{5.7}$$

where $\delta \geq 0$ acts as a filtering parameter determining the threshold of near-degeneracy between energy levels, noting $\lim_{\delta \to 0} \frac{\sqrt{\pi}\delta \operatorname{erf}(x/\delta)}{2x} = \delta_{x,0}$ and thus for $\delta = 0$ this will detect exact degeneracy. The metric satisfies $0 \leq s_\delta(\boldsymbol{\mu}) \leq 1$ and for $\delta < \delta'$ we have $s_\delta(\boldsymbol{\mu}) \leq s_{\delta'}(\boldsymbol{\mu})$. The maximum depth $D_{\max} \leq M$ allows one to truncate the outer sum since the number of inner terms increases as $\binom{M-1}{i}$; in the N_2 STO-3G case $M = 10$ so we may include all MO degeneracy contributions, but for larger systems we may truncate for ease of computation. This may be viewed in the lower subplot of Figure 5.10, where peaks indicate the presence of degenerate MOs. Encouragingly, these peaks coincide exactly with discontinuities in the noncontextual energy approximation, thus giving us confidence that the success of our Contextual Subspace techniques stems from its ability to describe static correlation in the noncontextual component.

Finally, while we identified the occurrence of noncontextual discontinuities to coincide with peaks in our MO degeneracy metric (5.7), one might consider the converse implication of our noncontextual problem as itself a test for detecting non-dynamical correlation. This is closely related to the Coulson-Fischer point, characterized by a divergence between restricted and unrestricted Hartree-Fock calculations [256], indicating a break-down of spin symmetry. The noncontextual discontinuities are reminiscent of a similar discontinuity found for H_2 in a work developing “Holomorphic Hartree-Fock Theory” [150] and warrants further investigation in this setting for the N_2 system. Furthermore, recall that we earlier demonstrated in Figure 3.5 that H_2 STO-3G is described exactly by a noncontextual theory, thus it could be instructive to investigate any possible connection to the holomorphic HF theory.

5.3 Kagome Lattice Heisenberg Model Ground State Preparation

The Heisenberg model provides an idealised quantum mechanical framework in which to study properties of magnetic systems such as critical points and phase transitions. The N-site, spin-$\frac{1}{2}$ Heisenberg Hamiltonian is defined here as

$$H = -\boldsymbol{J} \sum_{(i,j)\in\mathcal{E}} \boldsymbol{\sigma}^{(i)} \cdot \boldsymbol{\sigma}^{(j)} - h \sum_{i=0}^{N-1} \sigma_z^{(i)} \tag{5.8}$$

with coupling graph specified by edge set $\mathcal{E} \subset \mathbb{Z}_N^{\times 2}$, Pauli matrices $\boldsymbol{\sigma}^{(i)} = (\sigma_x^{(i)}, \sigma_y^{(i)}, \sigma_z^{(i)})$ acting on site i, interaction strengths $\boldsymbol{J} = (J_x, J_y, J_z) \in \mathbb{R}^3$ and external field strength $h \in \mathbb{R}$. The most general case of $J_x \neq J_y \neq J_y$ is referred to as the XYZ model, however there are restricted configurations that are of particular interest. The one we consider here is the XXX model, characterised by $J_x = J_y = J_y = J$. For $J > 0$ the XXX model exhibits ferromagnetic behaviour, whereas for $J < 0$ it may be antiferromagnetic. In this work we investigate the 12-site Kagome lattice depicted in Figure 5.11 and defined by the edge-set:

$$\begin{aligned}\mathcal{E} = \{&(0,1),(1,2),(2,3),(3,4),(4,5),(0,5),\\ &(0,6),(1,6),(1,7),(2,7),(2,8),(3,8),\\ &(3,9),(4,9),(4,10),(5,10),(5,11),(0,11)\}.\end{aligned} \tag{5.9}$$

Its geometrical structure gives rise to some exotic physics. In the antiferromagnetic setting of $J < 0$, each triangular subgraph permits two anti-parallel spins while the third is necessarily frustrated due to these competing exchange interactions [257], leading to a degeneracy of its energy spectrum. The Kagome lattice is a candidate quantum spin liquid (QSL), being predicted to exhibit entanglement, fractionalized excitations [258] and magnetic disorder, even at absolute zero temperature [259].

In Figure 5.12 we present the ground state energy E_0 and modulus of magnetization $M_z = |\sum_{i=0}^{N-1} \langle \sigma_z^{(i)} \rangle|/N$ for varying $J \in [-1,1]$ and $h \in [0,1]$. We expect a ferromagnetic system to yield $M_z = 1$ whereas, in the ideal antiferromagnetic setting, all spins should cancel to give $M_z = 0$. For the purposes of this example we restrict ourselves to the field-free setting of $h = 0$ and furthermore take $J = -1$, in which antiferromagnetic behaviour is encountered.

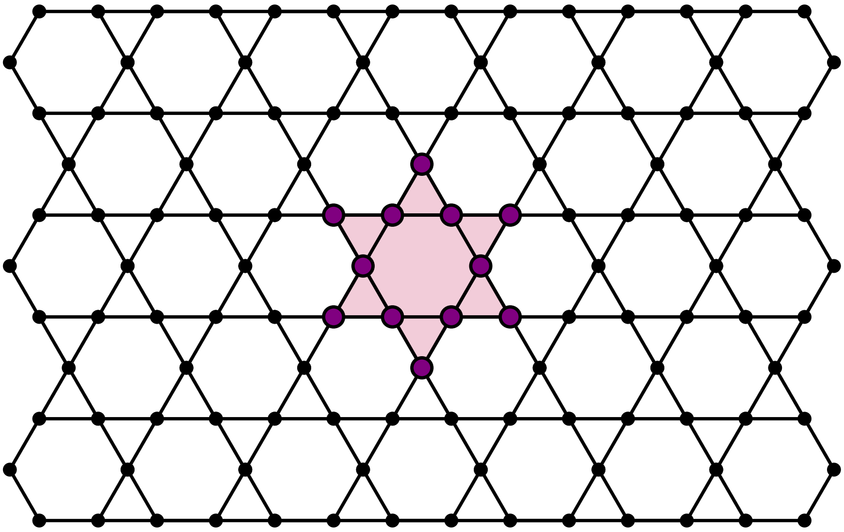


Figure 5.11: The Kagome lattice structure, consisting of hexagonal cells with joining triangular elements; the spin-frustration manifests in the latter geometric property. We isolate the central star-shaped lattice consisting of 12 spin-sites, highlighted above.

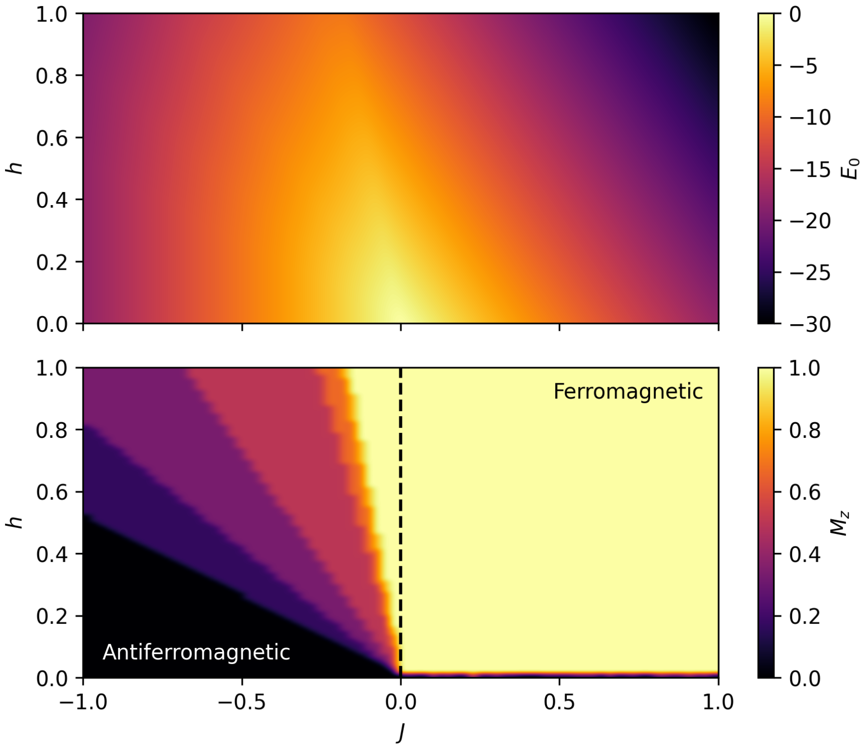


Figure 5.12: The XXX model ground state energy and modulus of magnetization for the 12-site sub-lattice highlighted in Figure 5.11 with varying coupling and field strengths J and h.

Qubit Subspace Approach

For such Heisenberg-type Hamiltonians we always find a $\mathbb{Z}_2$-symmetry of the form $\sigma_z^{\otimes N}$, which clearly commutes with each of the terms in (5.8). In the absence of an external field (that would otherwise contribute additional Hamiltonian terms $\sum_i \sigma_z^{(i)}$), which is the scenario we find ourselves in for this Kagome lattice problem, there is an additional $\mathbb{Z}_2$-symmetry $\sigma_x^{\otimes N}$; for odd-numbered lattices these symmetries anticommute, as was discussed previously in Example 2.4. Since we consider here the even case, tapering allows us to remove two degrees of freedom, reducing the full 12-qubit Hamiltonian H to a 10-qubit system that is isospectral, i.e. its energy spectrum is preserved exactly up to a potential lifting of various degeneracies. In our case, the ground eigenspace remains degenerate, indicating the presence of some non-$\mathbb{Z}_2$ symmetry.

Following Equation (2.24), the Clifford rotation

$$R = e^{-i\frac{\pi}{4}\sigma_y^{(1)}} e^{\frac{\pi}{4}\sigma_z^{(1)}\sigma_x^{\otimes N}} e^{\frac{\pi}{4}\sigma_x^{(0)}\sigma_z^{\otimes N}} \tag{5.10}$$

allows us to map the $\mathbb{Z}_2$-symmetries onto single-qubit Pauli operators $R\sigma_z^{\otimes N}R^\dagger = \sigma_x^{(0)}, R\sigma_x^{\otimes N}R^\dagger = \sigma_x^{(1)}$ acting on distinct qubit positions. Since conjugation by a unitary preserves commutation, the rotated hamiltonian $RHR^\dagger$ will necessarily consist of identity or σ_x operations in the qubit positions $0, 1$ onto which we mapped the symmetry operators. We may therefore drop these qubits, defining an effective 10-qubit Hamiltonian

$$\begin{aligned}H_{\text{taper}} =& -\sum_{\substack{(i,j)\in\mathcal{E}\\ i,j\notin\{0,1\}}} \boldsymbol{J}\cdot\boldsymbol{\sigma}^{(i')}\cdot\boldsymbol{\sigma}^{(j')}\\ &+ J_x \sum_{(0,j)\in\mathcal{E}} (I+\sigma_x^{\otimes N'})\sigma_x^{(j')} - J_z \sum_{(1,j)\in\mathcal{E}} (I+\sigma_z^{\otimes N'})\sigma_z^{(j')}\\ &- J_y \sum_{(1,j)\in\mathcal{E}} \left[\sigma_z^{\otimes j'}\sigma_x^{(j')}\sigma_z^{\otimes(N'-j'-1)} + \sigma_x^{\otimes j'}\sigma_z^{(j')}\sigma_x^{\otimes(N'-j'-1)}\right]\end{aligned} \tag{5.11}$$

where $x' = x - 2$ denotes an index shift. We reiterate that this holds only for the field free case of $h = 0$; for non-zero field strength there exists just a single $\mathbb{Z}_2$-symmetry and therefore it is only possible to remove a single qubit in this setting.

We now seek further reductions through the contextual subspace framework. This involves identifying a set of operators $\mathcal{S}$ that we wish to enforce over the Hamiltonian and will subsequently stabilize the resulting contextual subspace. We construct $\mathcal{S}$ via some reference state $|\psi\rangle$, with wave operator W such that $|\psi\rangle = W|\mathbf{0}\rangle$, that ideally has non-zero overlap with the ground eigenspace and select symmetry generators $\mathcal{S}$ that stabilize some maximal component therein, i.e. by identifying

$$\mathcal{S} = \underset{\mathcal{S}:|\mathcal{S}|=K}{\arg\max}\, \langle\psi| P_{\mathcal{S}} |\psi\rangle \tag{5.12}$$

where $P_{\mathcal{S}} := \frac{1}{2^K}\prod_{S\in\mathcal{S}}(I+S)$ projects onto the eigenspace spanned by the operators of $\mathcal{S}$. The difficulty is finding a suitable $|\psi\rangle$ in a scalable manner; while in the chemical examples of Sections 5.1 and 5.2 we used standard electronic wavefunctions such as MP2 or CCSD, for this condensed matter example we opt for a tensor network approach. For the Hamiltonian H we prepare a matrix product operator (MPO) by summing the individual terms of H represented as an MPO. During this summation, we truncate to some maximal bond dimension to maintain scalability. Once prepared, we apply a density-matrix renormalization group (DMRG) calculation (using the python package *quimb* [260]) to the MPO to prepare the reference state $|\psi\rangle$. With a limited maximum bond dimension of $D_{\max} = 17$, the reference state produces a relatively high overlap with the ground state, which we see from Figure 5.13. Then, to identify $\mathcal{S}$ we simply truncate $|\psi\rangle$ and calculate the symmetry generating set over the remaining terms, thus maximising the ℓ_1-norm of the projected state in the resulting contextual subspace.

Through application of this process to H_{taper} (5.11), we identify a reduction from 10 → 5 qubits that preserves the ground state energy *exactly*, yielding the following reduced Hamiltonian:

$$\begin{aligned}H_{\text{CS}} =& -I + 7\cdot\sigma_z^{(0)}\\ &+ (I+\sigma_z^{(0)})(\sigma_z^{(1)} + \sigma_z^{(1)}\sigma_z^{(2)} + \sigma_z^{(2)}\sigma_z^{(3)} + \sigma_z^{(3)}\sigma_z^{(4)})\\ &- (I-\sigma_z^{(0)})(\sigma_x^{(1)} + \sigma_x^{(2)} + \sigma_x^{(3)} - \sigma_x^{(4)} - \sigma_x^{(1)}\sigma_x^{(2)}\sigma_x^{(3)}\sigma_x^{(4)}).\end{aligned} \tag{5.13}$$

One might note the presence of a $\mathbb{Z}_2$-symmetry $\sigma_z^{(0)}$ and therefore it is possible to reduce the Hamiltonian to 4 qubits via a second round of tapering.

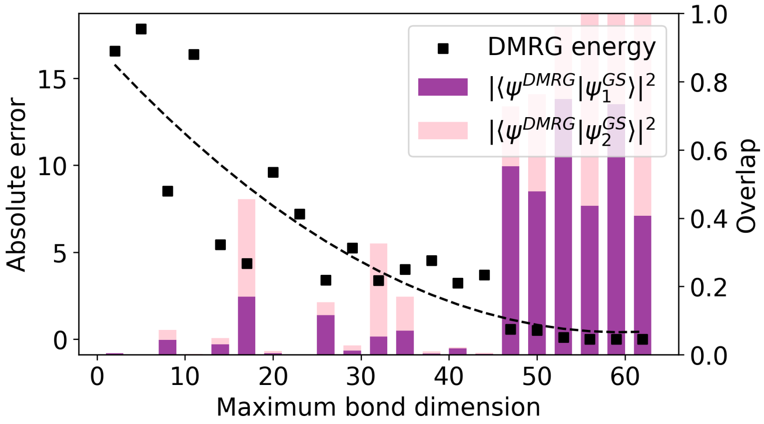


Figure 5.13: DMRG energy and ground state overlap for increasing bond dimension for the Kagome Lattice Heisenberg problem. We plot separate overlaps for each of the two eigenstates in the degenerate ground space.

Through application of both qubit tapering and the contextual subspace method we found that it is possible to reduce the full 12-qubit Kagome Lattice Heisenberg Hamiltonian to just 5-qubits while preserving the energy spectrum exactly. From a simulation point of view, this is highly beneficial as it allows us to parallelize three ansatz instances across the 16-qubits of *ibmq_guadalupe* to effectively triple the number of samples extracted from the quantum hardware in addition to averaging over noise, thus reducing variance in our energy estimate.

We also noted the presence of a $\mathbb{Z}_2$-symmetry $\sigma_z^{(0)}$. It is therefore possible to reduce the Hamiltonian to 4 qubits via tapering, although we elect to retain this symmetry for the purposes of symmetry verification, as discussed in Section 4.3. The Hamiltonian form in Equation (5.13) reveals projectors onto the ±1-eigenspaces of this symmetry; indeed, we note it is the -1 eigenvalue that will permit the lowest energy due to the dominant $7 \cdot \sigma_z^{(0)}$ term, thus we could drop the $+1$-eigenspace projector and any corresponding terms from the Hamiltonian. However, we instead opt to extract a noncontextual component from this Hamiltonian, achieved by retaining any single term corresponding with the $+1$-projector, say $(I + \sigma_z^{(0)})\sigma_z^{(1)}\sigma_z^{(2)}$, while discarding the others. It is now possible to express the contextual subspace Hamiltonian as

$$\begin{aligned} H'_{\text{CS}} = & \\ & - I + 7S_0 - S_2 + S_3 + S_0S_2 - S_0S_3 + S_1S_2S_3 - S_0S_1S_2S_3 \\ & + C_0(-I + S_0 - S_1 + S_0S_1) \\ & + C_1(I + S_0) \end{aligned} \tag{5.14}$$

where

$$\begin{aligned} & S_0 = \sigma_z^{(0)}, S_1 = \sigma_x^{(1)}\sigma_x^{(2)}, S_2 = \sigma_x^{(3)}, S_3 = \sigma_x^{(4)}, \\ & C_0 = \sigma_x^{(2)}, C_1 = \sigma_z^{(1)}\sigma_z^{(2)}. \end{aligned} \tag{5.15}$$

In particular, the Hamiltonian may be decomposed into a symmetry component plus a sum over commuting cliques such that terms across cliques pairwise anticommute. Calling C_0, C_1 the clique *representatives*, since $\{C_0, C_1\} = 0$ and the S_i commute globally amongst terms of the Hamiltonian, H'_{CS} exhibits the noncontextual structure introduced in Section 3.2. Invoking Section 3.2, the energy spectrum is therefore specified by eigenvalue assignments to the symmetry elements S_0, S_1, S_2, S_3, specifying a symmetry *sector*, and clique operator $C(r_0, r_1) = r_0 C_0 + r_1 C_1$ with $|r_0|^2 + |r_1|^2 = 1$. We may search the symmetry sectors with R_y rotations through binary relaxation of the eigenvalue assignments, since $R_y(\theta_i)\,|0\rangle = \cos(\theta_i/2)\,|0\rangle + \sin(\theta_i/2)\,|1\rangle$. Treatment of the clique operator is slightly different – in solving the noncontextual problem we optimize over clique weights r_0, r_1 which in turn yields a rotation of the form in unitary partitioning [117], the specific form of which is found by exponentiating products of clique representatives $e^{\theta/2 C_0 C_1} = R_{zy}(\theta)$ with $C_0 C_1 = -i\sigma_z^{(1)}\sigma_y^{(2)}$. Therefore, we are able to construct a simplistic ansatz

$$|\psi(\boldsymbol{\theta})\rangle = e^{-i\frac{\theta_5}{2}\sigma_z^{(1)}\sigma_y^{(2)}} \prod_{n=0}^{4} e^{-i\frac{\theta_n}{2}\sigma_y^{(n)}}\,|\mathbf{0}\rangle\,, \tag{5.16}$$

expressed as a circuit in Figure 5.14c, that should be fully-expressible, given that it searches the noncontextual energy spectrum that we have deemed to capture the ground state.

Simulation Results

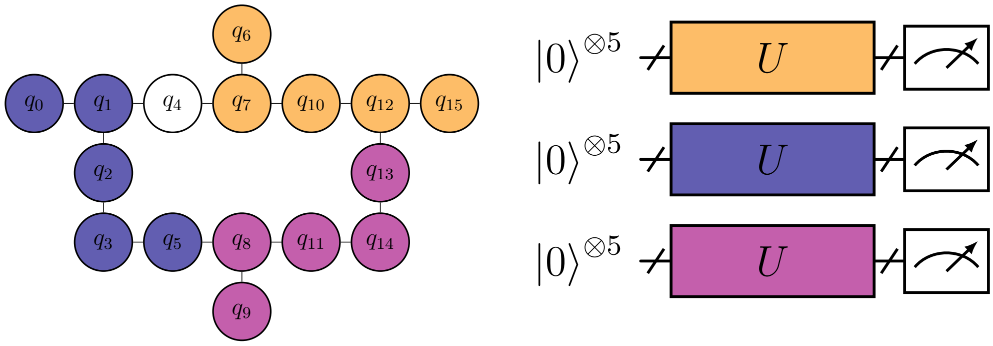
All in all, we possess a reduced 5-qubit Hamiltonian whose ground state energy coincides with that of the full 12-site Kagome lattice (Figure 5.11) field-free Heisenberg XXX Hamiltonian and a simple ansatz circuit. We are able to parallelize three instances of this ansatz across the 16-qubits available on *ibmq_guadalupe*, tripling the number of circuit shots obtained and averaging over hardware noise, as depicted in Figures 5.14a and 5.14b. In order to obtain the most accurate energy estimates possible, we deploy a quantum error mitigation (QEM) strategy consisting of measurement error mitigation (Section 4.2), Symmetry verification (Section 4.3) and zero-noise extrapolation (Section 4.6).

The next practical consideration for our simulation is the measurement overhead for our quantum experiments. Two Pauli operators P, Q (tensor products of the single-qubit Pauli matrices $I, \sigma_x, \sigma_y, \sigma_z$) are said to be *qubit-wise commuting* (QWC) if $[P^{(i)}, Q^{(i)}] = 0 \;\; \forall i \in \mathbb{Z}_N$; this is useful as the measurement outcome for a collection of QWC Pauli operators may be inferred from only a single physical measurement [116]. Conveniently, our contextual subspace Hamiltonian in Equation (5.13) is expressible as a sum of just two QWC groups, thus reducing the measurement cost significantly.

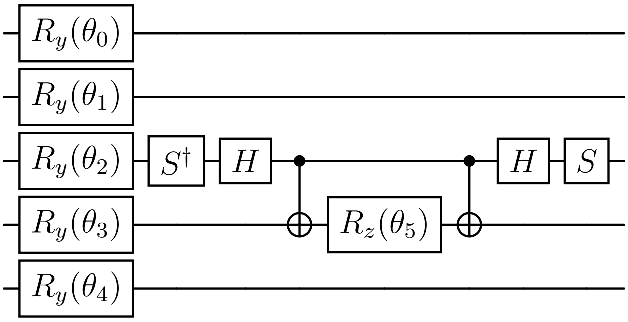
For the variational quantum eigensolver (VQE) routine itself, we choose the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [194] to perform the classical optimization over our six ansatz parameters (given explicitly in (5.16)) with objective function $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|\, H_{\mathrm{CS}}\, |\psi(\boldsymbol{\theta})\rangle$. The parameter gradients are computed in hardware using the parameter shift rule [122]:

$$\frac{\partial}{\partial\theta_i} E(\boldsymbol{\theta}) = E\Big(\boldsymbol{\theta}; \theta_i + \frac{\pi}{4}\Big) - E\Big(\boldsymbol{\theta}; \theta_i - \frac{\pi}{4}\Big). \tag{5.17}$$

In Figure 5.15 we present the results of our final VQE simulation, displaying the absolute energy error for each of the noise amplified energy estimates and the final zero-noise estimate both with/without symmetry verification. The norm of



(a) The *ibmq_guadalupe* Falcon r4P chip 'heavy-hex' topology. **(b)** Tiled ansatz circuit for the purposes of noise averaging.



(c) A 5-qubit ansatz defined over the contextual subspace that searches the noncontextual energy landscape.

Figure 5.14: Mapping our noncontextual ansatz onto the available device topology.

the parameter gradient is also plotted alongside the parameter trace itself; we observe a convergence on integer multiples of $\pi/2$, indicative of a stabilizer state solution corresponding with the discrete eigenvalue assignment one expects in noncontextual optimization.

The error ratio at convergence was computed as the mean energy estimate for points having a gradient norm less than twice the minimum value, in this case optimization step 16 onward. Before application of SV, this represents an error ratio of 0.360%, corresponding with an absolute error of 0.065, whereas post-SV yields a suppressed **error ratio of 0.078%** with absolute error 0.014.

Finally, we calculate the resource usage that went into producing these results. The simulation consisted of $N_{\text{energy}} = 38$ energy estimates and $N_{\text{grad}} = 26$ gradient evaluations over $N_{\text{param}} = 6$ parameters. Each estimate requires $N_{\text{ZNE}} = 4$ noise amplified expectation value calculations over $N_{\text{QWC}} = 2$ QWC groups. For the VQE routine presented in Figure 5.15 we set $N_{\text{shots}} = 8192$, although we also ran simulations for $N_{\text{shots}} \in \{256, 1024, 4096, 8192\}$ with both BFGS and conjugate gradient (CG) [261] classical optimizers. Altogether, the total number of circuit executions performed on *ibmq_guadalupe* is

$$\begin{aligned} N_{\text{resource}} &= N_{\text{shots}} N_{\text{QWC}} N_{\text{ZNE}} (N_{\text{energy}} + 2 N_{\text{grad}} N_{\text{param}}) \\ &= N_{\text{shots}} \times 2800 \\ &= 2.29 \times 10^7. \end{aligned} \tag{5.18}$$

The majority of this resource went into the gradient calculations; for larger systems we could apply ZNE only to the energy estimates to reduce the overhead.

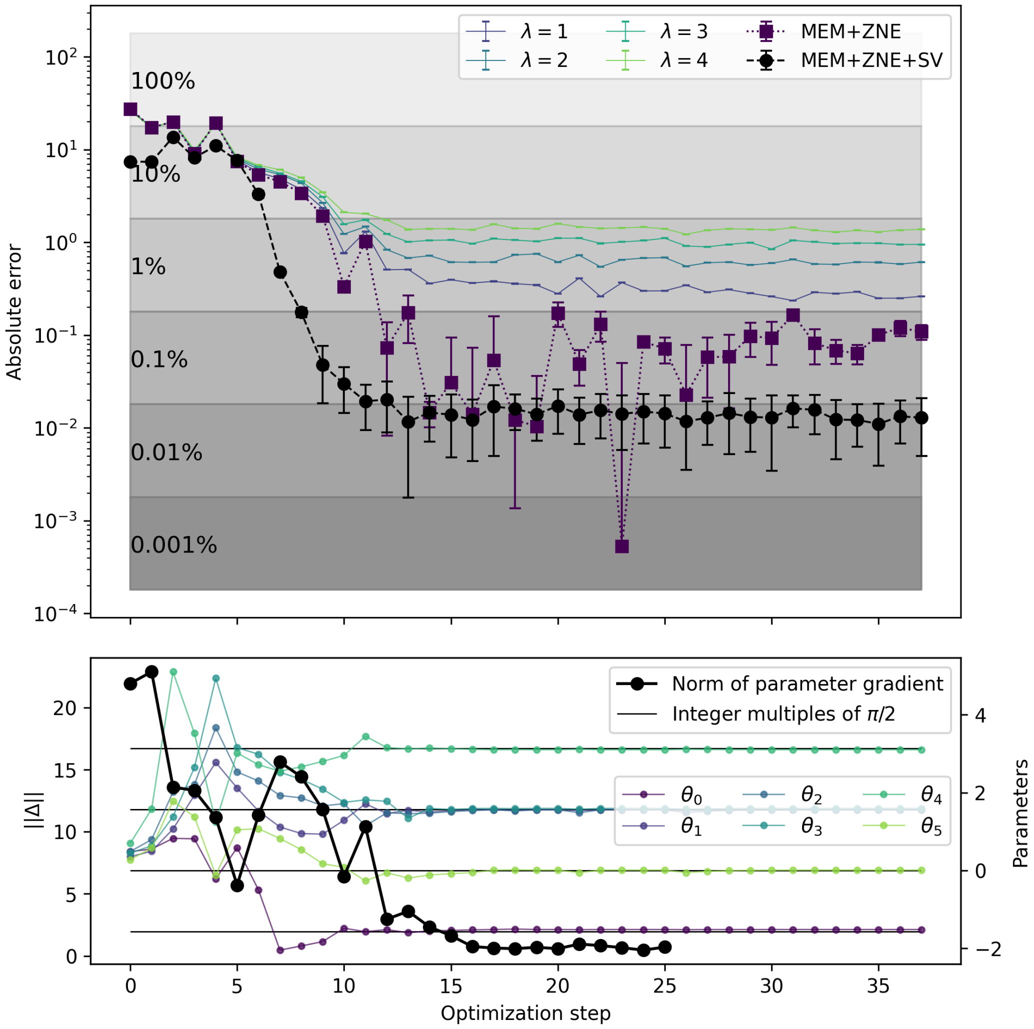


Figure 5.15: Optimizer energy trace before and after application of symmetry verification. Furthermore, the parameters converge on integer multiples of $\pi/2$, indicative of a stabilizer state.

5.4 Heavy-Hex Ising Model Time Evolution

To benchmark the Echo Verified Clifford Data Regression (EVCDR) method of Section 4.8 we choose to simulate the Ising model on 2D spin-lattices

$$H = -J \sum_{\langle a,b \rangle} Z^{(a)} Z^{(b)} - h \sum_{a} X^{(a)} \tag{5.19}$$

where $J, h \in \mathbb{R}$ parameterize strengths of the spin coupling and external field, respectively; the ratio J/h dictates the Hamiltonian time evolution. The pairwise site indices $\langle a, b \rangle$ run over nodes of the graph representing the nearest-neighbour spin couplings, which we choose such that the corresponding lattice is subgraph isomorphic with respect to the hardware topology; see the heavy-hex 4-cell in Figure 5.16.

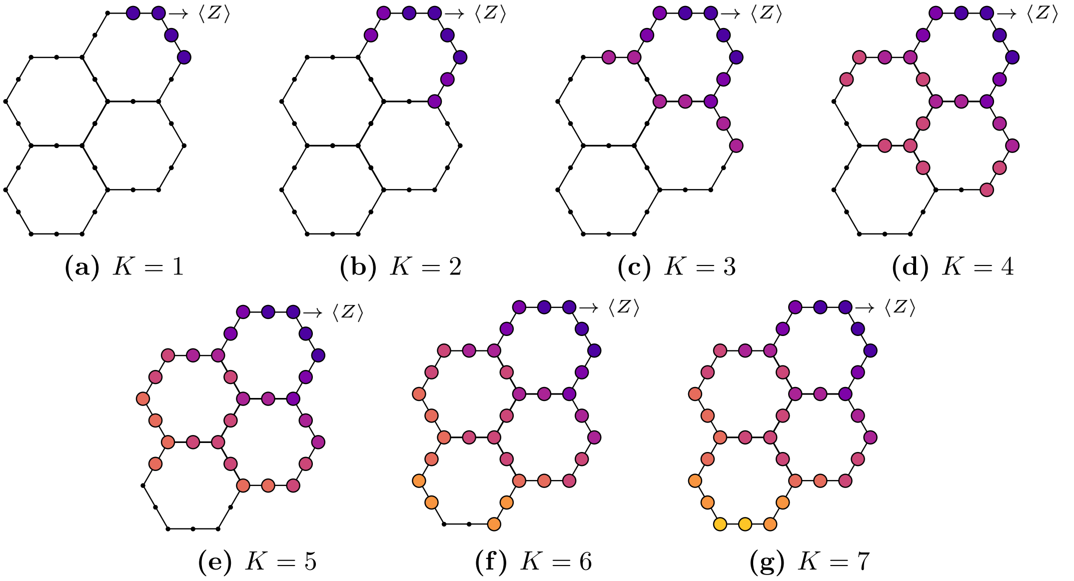
The quantity we aim to estimate is the time-dependent magnetization on a single spin-site

$$M(t) = \langle 0 | \, e^{-iHt} Z^{(0)} e^{iHt} \, | 0 \rangle \, . \tag{5.20}$$

For a number of time steps $K \in \mathbb{N}$ with step-size τ, the Hamiltonian evolution at final time $t = K\tau$ is captured by a circuit $\prod_{k=1}^{K} U(\tau)$ where

$$\begin{aligned} U(\tau) &= \prod_{\langle a,b\rangle} e^{-iJ\tau Z^{(a)}Z^{(b)}} \prod_{a} e^{-ih\tau X^{(a)}} \\ &= e^{iH\tau} + \mathcal{O}(\tau^2). \end{aligned} \tag{5.21}$$

The circuit formulation of $U(\tau)$ is depicted in Figure 5.17, which propagates the system through a time τ with each succesive application. The error here is the consequence of first-order Trotterization [21] and may be combated by taking finer time slices τ, with the trade-off being increased circuit depth, as discussed in Section 1.2.4. With each successive application of the time propagation circuit,



(a) $K = 1$ **(b)** $K = 2$ **(c)** $K = 3$ **(d)** $K = 4$

(e) $K = 5$ **(f)** $K = 6$ **(g)** $K = 7$

Figure 5.16: The heavy-hex 4-cell, consisting of 35 qubits. Figures **(a)**-**(g)** illustrate the expansion of the light-cone for successive time steps K probing time $t = K\tau$, until the spin lattice is saturated from $K = 7$ onwards.

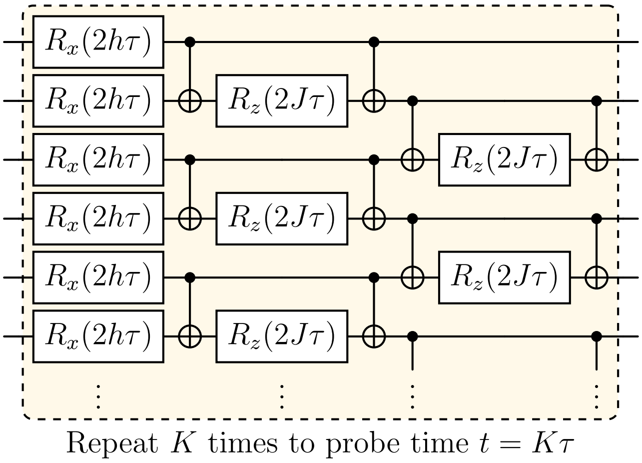
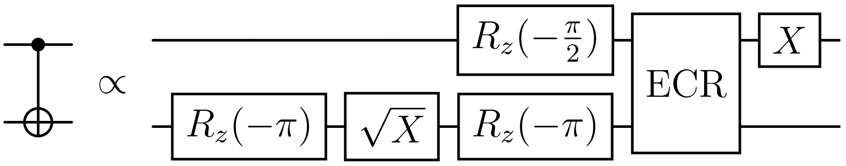
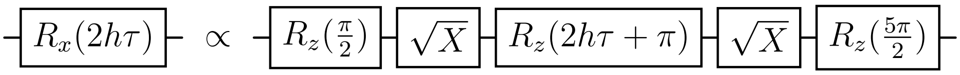


Figure 5.17: A single time step circuit for simulating the Ising Hamiltonian in Equation (5.19). Evolution for longer times is achieved via repeated application of this circuit block. The rotation convention is $R_x(\theta) = \exp\left(-i\frac{\theta}{2}X\right)$ and $R_z(\theta) = \exp\left(-i\frac{\theta}{2}Z\right)$.



(a) Global phase $\frac{\pi}{2}$



(b) Global phase $\frac{3\pi}{2}$

Figure 5.18: Transpilation of non-native gates in Figure 5.17 onto the target gateset $\{X, \sqrt{X}, R_z, \text{ECR}\}$, where ECR is the echoed cross-resonance gate.

the effective light-cone for local observables expands as illustrated in Figure 4.14. In Figure 5.16 we have indicated how this expansion grows from a single spin-site on the heavy-hex 4-cell lattice with each successive time step K for our Ising model time evolution problem. From $K = 7$ onward the lattice is saturated.

The following benchmark was deployed on the *ibm_sherbrooke* 127-qubit Eagle r3 device, with the native gateset $\{X, \sqrt{X}, R_z, \text{ECR}\}$ and a rated error per layered gate (EPLG) of 1.7% at the time of execution. For each simulation we tiled concurrent circuit instances across the chip in order to passively average hardware errors and increase the effective yield from the $80,000$ circuit shots performed at each time step; the heavy-hex 2-cell permitted four circuit tilings (example in Figure 5.19), while the 4-cell allowed just two. We performed Clifford Data Regression (CDR) within each cluster to learn the effect of the local noise channel, as opposed to pooling the data into a single CDR routine; this is visible in Figure 4.19, where each fitting line represents different clusters/batches to account for the non-uniformity of device noise. Note that each CDR circuit received the same shot allowance across three batches of five unique Clifford realizations, resulting in fifteen separate backend submissions in addition to the standard EV one.

To mitigate the problem of vanishing success probability discussed in Section 4.7.1 and Figure 4.15, we allow a small number of bit-flips in the postselection procedure; we implement this by forcing a neighbourhood of qubits around the ancilla to be zero, while those outside the selected neighbourhood may deviate by a predefined maximum Hamming distance. Although we may now be including invalid measurements that will bias our output, we argue that it is a reasonable assumption as a small number of flips will correspond with a small change to the magnetization, particularly if the flips occur far from the measured spin-site.

We developed a bespoke implementation for the circuit construction, clustering, batching and final submission to the quantum backend on top of `qiskit` functionality [187]. Pauli operator manipulation and Clifford simulation was handled using the `symmer` package [10] and the regression step in CDR made use of `statsmodels` [262]. We have provided all the code and data on an online repository [8] so that readers may reproduce the results presented here, or design their own EV-based error mitigation workflows.

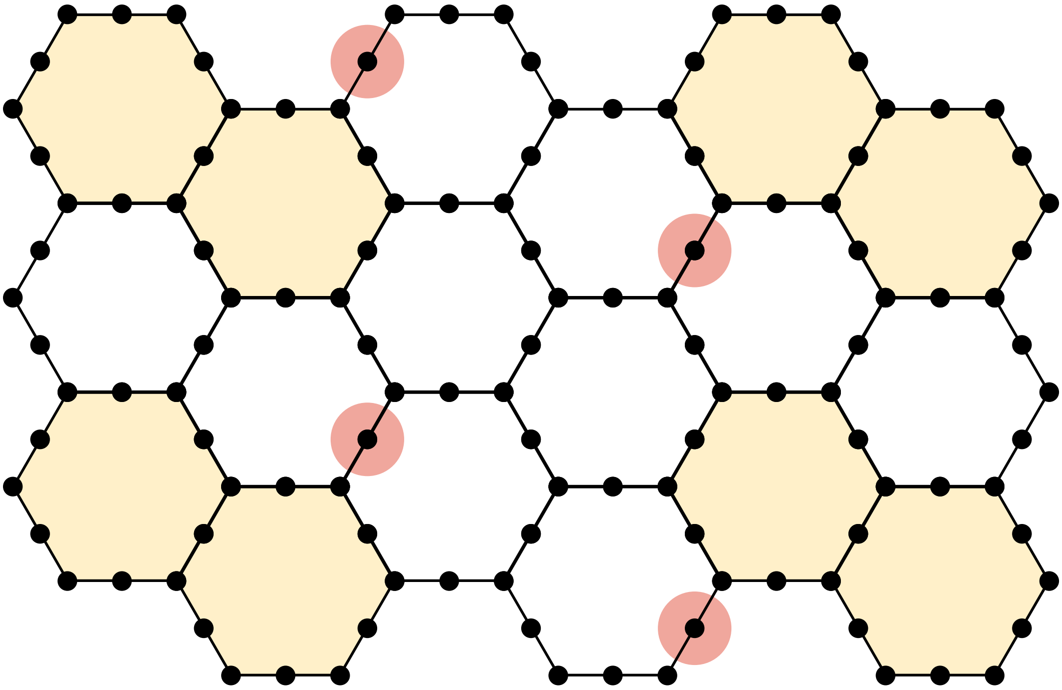


Figure 5.19: The heavy-hex qubit topology of the devices utilized for this work: *ibm_cusco*, *ibm_nazca*, *ibm_brisbane* and *ibm_sherbrooke*. The shaded areas indicate how we tile four 2-cell clusters to increase sampling throughput on the device for the results in Figure 5.20. The circular shaded areas indicate the ancilla qubits selected to probe the spin magnetization of the adjacent site in the lattice region. For the larger 4-cell, the results of which are presented in Figure 5.21, only two clusters may be packed onto the device.

For this practical demonstration we compare these echo-verified estimators:

- **Standard**: the approach outlined in the introductory Section 4.7.
- **Spectral Purification**: described previously as tomography purification, this is a modification to the standard estimator, whereby one decomposes the state of the ancilla qubit $\rho^{\text{anc}} = \lambda_0 |\phi_0\rangle \langle\phi_0| + \lambda_1 |\phi_1\rangle \langle\phi_1|$ and selects the dominant pure component, say $\lambda_0 > \lambda_1$ (discussed in [3, 231]).
- **Purity Normalization**: through the observation $\gamma(\rho^{\text{anc}})^2 \approx \text{Tr}\left(X\rho^{\text{anc}}\right)^2 + \text{Tr}\left(Y\rho^{\text{anc}}\right)^2 + \text{Tr}\left(Z\rho^{\text{anc}}\right)^2$ where $\gamma(\rho^{\text{anc}})$ is the purity, one may re-normalize the ancillary estimators $\mathcal{E}_X \mapsto \mathcal{E}_X/\gamma(\rho^{\text{anc}}), \mathcal{E}_Z \mapsto \mathcal{E}_Z/\gamma(\rho^{\text{anc}})$ to recover $\mathcal{E}_X^2 + \mathcal{E}_Z^2 = 1$, the property used to form alternative EV estimators in Equation (4.23).
- **Clifford Data Regression**: approximate the target unitary with near-Clifford circuits that may be evaluated classically to learn the effect of noise on the EV estimate, as detailed in Section 4.8 and pictured in Figure 4.18. The noise manifests as a linear relationship between the noisy and ideal expectation values and can therefore be analysed via regression techniques.

Heavy-Hex 2-cell simulation (21 qubits)

We first present the heavy-hex 2-cell results of Figure 5.20, consisting of 21-qubits in total. With respect to the Ising model (5.19) with $J = 4, h = 2$, we observe the evolution of the magnetization of a single spin over 10 steps to a final time $T = 1/2$ in increments of $\tau = 1/20$. Due to the light-cone reduction depicted in Figure 4.14, we do not saturate the spin lattice until time step $K = 5$, while the number of entangling ECR layers continues to increase linearly past this point to a maximum of 113 at $K = 10$. We find that the standard EV approach yields poor errors in excess of 0.1, while the purity normalized and spectral purification modifications suppress the absolute error to within 0.1 and 0.01. EVCDR, on the other hand, consistently produces the lowest errors in the range $0.01 - 0.0001$.

Heavy-Hex 4-cell simulation (35 qubits)

Following the 21-qubit case, we now attempt the EVCDR technique on the heavy-hex 4-cell defined over 35-qubits visible in Figure 5.21. In the Ising model we take $J = 1, h = 8$ and evaluate the evolution to a final time of $T = 1$ in 15 steps of size $\tau = 1/15$. Here, the light-cone does not saturate the lattice until time step $K = 7$, with the number of entangling ECR layers reaching a maximum of 173 at the final step $K = 15$. Once again, EVCDR offers the most consistent results with absolute errors never exceeding 0.1 throughout the time evolution, with many points dropping to order 0.01 in error. Visually comparing the results in Figures 5.20 and 5.21, it might appear that all the EV methods investigated here perform better for the larger 35-qubit simulation; however, we draw attention to the scale on the Y-axis where the 4-cell evolves between $[-1, +1]$ while the 2-cell is contained in $[0.85, 1]$. The error plot is more instructive, indicating approximately one order of magnitude increase in error going from the 2-cell to the larger 4-cell.

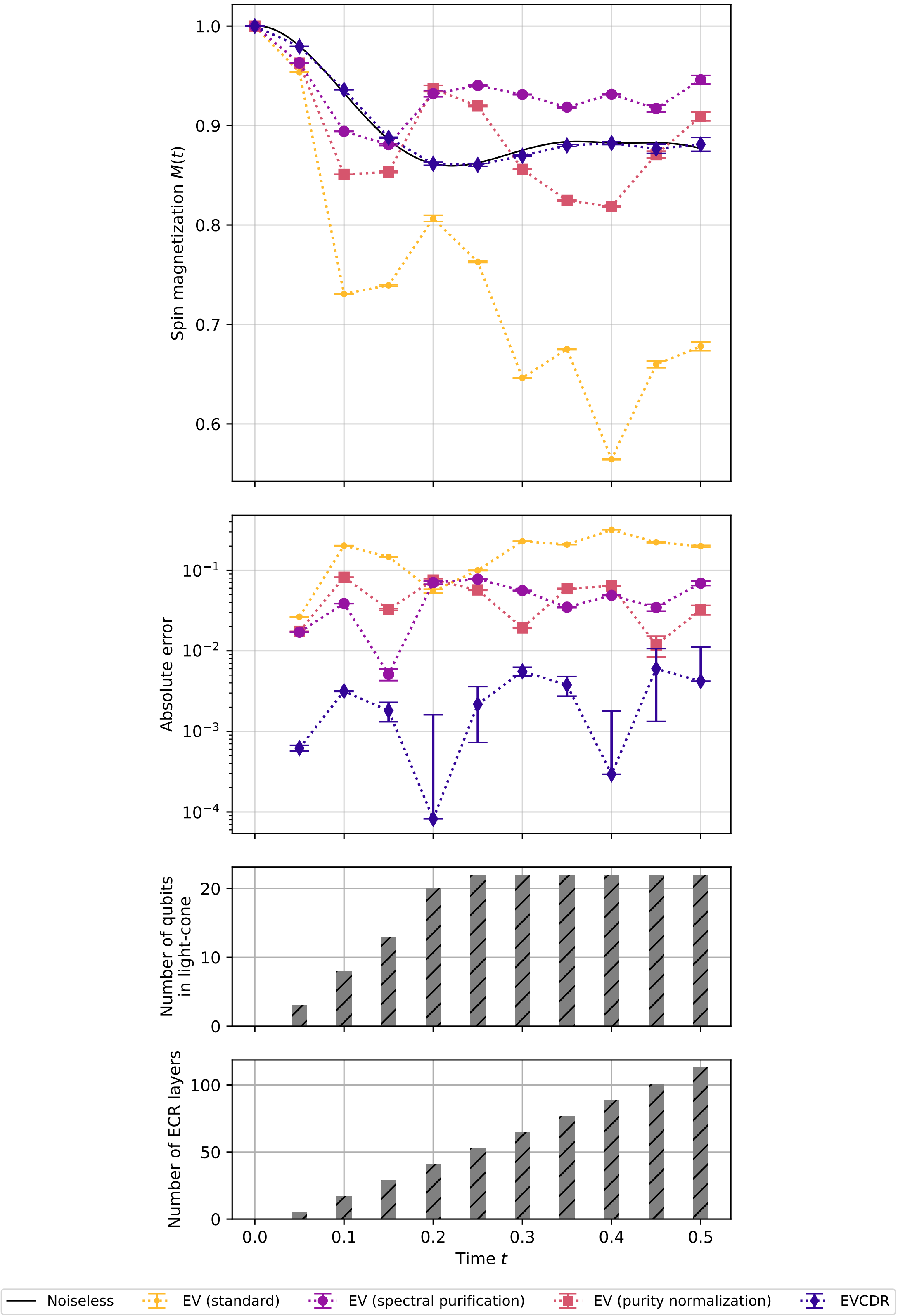


Figure 5.20: Evolution of spin magnetization for a single site in the 21-qubit heavy-hex 2-cell Ising model of Equation (5.19) with $J = 4, h = 2$, comparing various echo verification estimators. We evolve to a final time $T = 1/2$ over 10 steps, corresponding with a step size of $\tau = 1/20$. Error bars show standard error on the mean.

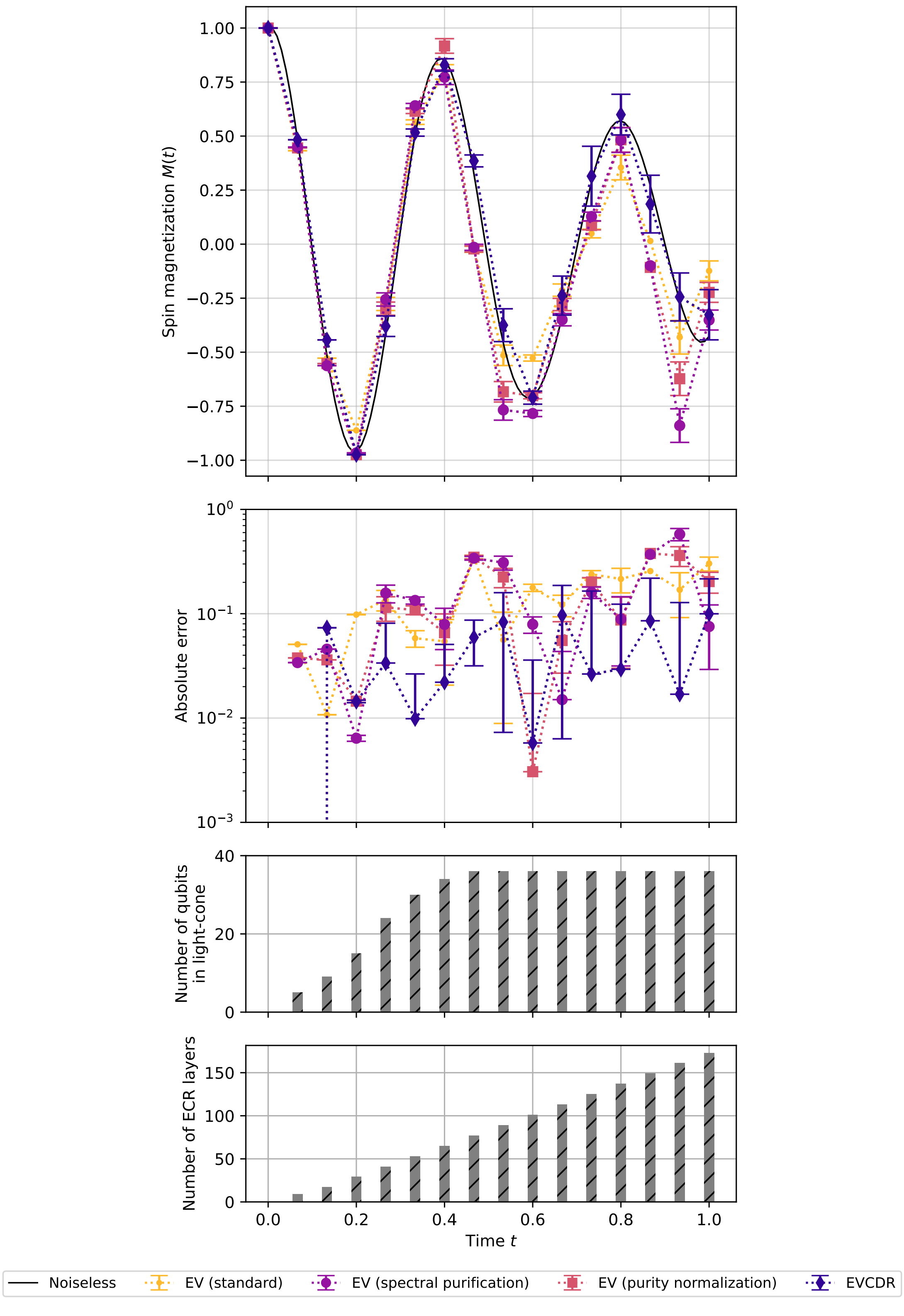


Figure 5.21: Evolution of spin magnetization for a single site in the 35-qubit heavy-hex 4-cell Ising model of Equation (5.19) with $J = 1, h = 8$, comparing various echo verification estimators. We evolve to a final time $T = 1$ over 15 steps, corresponding with a step size of $\tau = 1/15$. Error bars show standard error on the mean.

Chapter 6

Conclusion and Outlook

In this thesis we developed a scalable framework for balancing quantum and classical workloads in the simulation of Pauli operator systems. Implementation efficiency comes from the utilization of the symplectic representation of Pauli operators, whereby many operations over the Pauli group reduce to simple binary vector manipulation, as demonstrated through the `symmer` software package [10] that was developed to support this work. This quantum compute resource management system is based on the quantum foundational concept of strong measurement contextuality and thus distinguishes itself from alternative qubit subspace, embedding and active space techniques. The Contextual Subspace (CS) framework [2, 3, 5, 120, 166] has been designed with noisy Intermediate-Scale Quantum applications in mind, with the goal of augmenting the computational reach of current and near-term quantum hardware to address larger physical systems. To complement the theoretical results towards advancing the Contextual Subspace methodology presented in Chapters 2 and 3, we also provided several practical demonstrations of the Variational Quantum Eigensolver (VQE) on IBM Quantum superconducting hardware which serves as a proof-of-concept and validation of the technology. Experimental realization of CS-VQE required the development of effective Quantum Error Mitigation (QEM) strategies, as discussed in Chapter 4 and implemented for several applications in Chapter 5.

In Section 5.1 we studied the problem of ground state preparation for hydrogen chloride, HCl, which provided an ideal testbed to assess the efficacy of numerous contemporary QEM techniques: Measurement-Error Mitigation (Section 4.2), Symmetry Verification (Section 4.3), Zero-Noise Extrapolation (Section 4.6) and Echo Verification (Section 4.7). After identification of effective QEM strategies, we then applied this in Section 5.2 to the more challenging molecular nitrogen system, N_2, whose triple bond causes problems for single-reference wavefunction methods and is a commonly used benchmark problem [147, 247]. We simulated the full bond dissociation of N_2 on superconducting hardware and compared errors against common single-reference methods as well as more competitive Complete Active Space (CAS) approaches. We found that CS-VQE was able to achieve comparable results with CAS techniques, but with a considerably reduced qubit resource overhead. Then, in Section 5.3, we applied the same CS-VQE implementation used for N_2 to a field-free Heisenberg model over the Kagome lattice, whose geometrical properties cause spin frustration and other exotic physical effects. Our robust simulation framework achieved an error threshold of 0.078%

and claimed first place globally in the IBM Quantum Open Science Prize 2023 [1], for which the Kagome problem had been posed. Outside of VQE, in Section 5.4 we also simulated the Hamiltonian time evolution for an Ising model system over the heavy-hex lattice that is native to IBM Quantum hardware. This problem was used to demonstrate the efficacy of a novel QEM technique combining Echo Verification and Clifford Data Regression, as described in Section 4.8.

Despite our successes in applying VQE experimentally, there are some inherent scaling limitations associated with the algorithm (and Variational Quantum Algorithms in general) that mean future work will explore alternative approaches. First of all, the measurement overhead is prohibitively high and scales with the number of terms in the Pauli Hamiltonian; in chemistry, this scales as $\mathcal{O}(N^4)$. While there have been advances in measurement reduction schemes that aim to combat this overhead [115–120], there are considerable concerns over whether this is sufficient for VQE to have any hope of achieving the essential milestone of "advantage" in the development of quantum computing. Moreover, an additional downside to VQE is that hardware noise enters the energy estimation procedure and has a severe impact on error; this necessitates the heavy application of QEM, which itself brings additional overhead in both coherent resource and classical postprocessing. The final nail in VQE's coffin comes from the difficulty in optimizing the variational circuits to minimize Hamiltonian energy. While the circuit depths in VQE are much shallower than, say, an algorithm like Quantum Phase Estimation (QPE), chemically-motivated ansätze still require circuits at a depth that is impractical for near-term applications. By contrast, shallower "hardware efficient" circuits are hampered by the insidious issue of barren plateaus [123–125, 127]; note also that recent research suggests this is an issue for problem-inspired circuit as well, although to a lesser degree [130].

Taking into account the above limitations of VQE, it is unlikely to provide any form of algorithmic advantage in the chemical domain. As such, future research will pursue alternative routes. One promising methodology is Quantum-Selected Configuration Interaction [59, 66], which treats the quantum hardware as an electron configuration sampling machine and subsequently the Hamiltonian is projected into this subspace and solved via conventional High-Performance Computing (HPC). Since HPC handles the diagonalization step, any potential source of quantum advantage will arise from the circuit implementation being able to preferentially sample "important" configurations more effectively than classical approaches such as Heatbath Configuration Interaction [56–58]. State sampling is a classically-hard problem, while it is a natural workload for a quantum computer and thus a more hopeful application of the technology. Furthermore, since Contextual Subspace is a highly flexible framework that is not tied to any particular solver such as VQE, it is fully compatible with QSCI and therefore we may endeavour to build a CS-QSCI technique that will facilitate the near-term simulation of large chemical systems at such scale to be considered of scientific and industrial significance.

Looking further to the future, even within the fault-tolerant regime of quantum computing that would arise from the realization of large-scale Quantum Error Correction, the CS methodology will still be widely applicable in allowing us to exceed whatever the logical qubit count of the day might be. One may imagine a CS-QPE algorithm that fully utilizes the capabilities of a fault-tolerant system.

To reiterate, the power of the Contextual Subspace methodology lies in its adaptability, being agnostic to the underlying solver (whether VQE, QSCI, QPE or other), and the flexibility to balance quantum and classical workloads, wherein the partitioning is motivated by the quantum foundational concept of strong measurement contextuality. This will be especially important going forwards, as we are already seeing the integration of Quantum Processing Units (QPUs) into HPC platforms, where such load-balancing tasks are a common requirement in achieving optimal performance from increasingly heterogeneous computing platforms to minimize slack in the scheduling of compute processes. In short, the Contextual Subspace methodology, the theory for which has been advanced within this thesis and validated experimentally through several practical implementations, provides a framework that will scale naturally with the maturation of quantum hardware in the coming years/decades and provide a viable route to achieving quantum advantage.

Bibliography

[1] IBM. *Here are the 2023 IBM Quantum Open Science Prize winners*. 2023. URL: `https://www.ibm.com/quantum/blog/ibm-quantum-open-science-winners-2023`.

[2] Tim Weaving, Alexis Ralli, William M. Kirby, Andrew Tranter, Peter J. Love, and Peter V. Coveney. "A Stabilizer Framework for the Contextual Subspace Variational Quantum Eigensolver and the Noncontextual Projection Ansatz". In: *Journal of Chemical Theory and Computation* 19 (2023), pp. 808–821. DOI: `10.1021/acs.jctc.2c00910`.

[3] Tim Weaving, Alexis Ralli, William M Kirby, Peter J Love, Sauro Succi, and Peter V Coveney. "Benchmarking noisy intermediate scale quantum error mitigation strategies for ground state preparation of the HCl molecule". In: *Phys. Rev. Res.* 5 (2023), p. 43054. DOI: `10.1103/PhysRevResearch.5.043054`.

[4] Tim Weaving. *Supporting code and data repository for article "Benchmarking noisy intermediate scale quantum error mitigation strategies for ground state preparation of the HCl molecule"*. 2023. URL: `https://github.com/TimWeaving/quantum-error-mitigation`.

[5] Tim Weaving, Alexis Ralli, Peter J Love, Sauro Succi, and Peter V Coveney. "Contextual subspace variational quantum eigensolver calculation of the dissociation curve of molecular nitrogen on a superconducting quantum computer". In: *npj Quantum Information* 11.1 (2025), p. 25. DOI: `10.1038/s41534-025-01019-8`.

[6] Tim Weaving. *Supporting code and data repository for article "Contextual subspace variational quantum eigensolver calculation of the dissociation curve of molecular nitrogen on a superconducting quantum computer"*. 2024. URL: `https://github.com/TimWeaving/N2-CS-VQE`.

[7] Tim Weaving, Alexis Ralli, Peter J Love, Sauro Succi, and Peter V Coveney. "Accurately Simulating the Time Evolution of an Ising Model with Echo Verified Clifford Data Regression on a Superconducting Quantum Computer". In: *Quantum* 9 (2025), p. 1732. DOI: `10.22331/q-2025-05-05-1732`.

[8] Tim Weaving. *Supporting code and data repository for article "Accurately Simulating the Time Evolution of an Ising Model with Echo Verified Clifford Data Regression on a Superconducting Quantum Computer"*. 2025. URL: `https://github.com/TimWeaving/EVTools`.

[9] Tim Weaving, Alexis Ralli, Vinul Wimalaweera, Peter J Love, and Peter V Coveney. “Simulating the Antiferromagnetic Heisenberg Model on a Spin-Frustrated Kagome Lattice with the Contextual Subspace Variational Quantum Eigensolver”. In: *arXiv preprint* (2025). arXiv: `2506.12391`.

[10] Alexis Ralli and Tim Weaving. *Symmer*. 2024. URL: `https://github.com/UCL-CCS/symmer`.

[11] Richard P. Feynman. “Simulating physics with computers”. In: *International Journal of Theoretical Physics* 21 (1982), pp. 467–488. DOI: `10.1007/BF02650179`.

[12] Man-Duen Choi. “Completely positive linear maps on complex matrices”. In: *Linear algebra and its applications* 10 (1975), pp. 285–290. DOI: `10.1016/0024-3795(75)90075-0`.

[13] B Camino, J Buckeridge, PA Warburton, V Kendon, and SM Woodley. “Quantum computing and materials science: A practical guide to applying quantum annealing to the configurational analysis of materials”. In: *Journal of Applied Physics* 133.22 (2023). DOI: `10.1063/5.0151346`.

[14] Yaoyun Shi. “Both Toffoli and Controlled-NOT need little help to do universal quantum computation”. In: *arXiv preprint quant-ph/0205115* (2002). arXiv: `quant-ph/0205115`.

[15] Dorit Aharonov. “A Simple Proof that Toffoli and Hadamard are Quantum Universal”. In: *arXiv preprint* (2003). arXiv: `quant-ph/0301040`.

[16] W. Heisenberg. “Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik”. In: *Zeitschrift für Physik* 43 (1927), pp. 172–198. DOI: `10.1007/BF01397280`.

[17] Niels Bohr. “The Quantum Postulate and the Recent Development of Atomic Theory”. In: *Nature* 121.3050 (1928), pp. 580–590. DOI: `10.1038/121580a0`.

[18] Hugh Everett III. “”Relative state” formulation of quantum mechanics”. In: *Reviews of Modern Physics* 29.3 (1957), p. 454. DOI: `10.1103/RevModPhys.29.454`.

[19] Bryce Seligman Dewitt and Neill Graham. *The Many-Worlds Interpretation of Quantum Mechanics: A Fundamental Exposition by Hugh Everett, III, with Papers by JA Wheeler, BS DeWitt, LN Cooper and D. Van Vechten, and N. Graham.* Princeton University Press, 2025.

[20] Rolando D. Somma. “A Trotter-Suzuki approximation for Lie groups with applications to Hamiltonian simulation”. In: *Journal of Mathematical Physics* 57 (2016). DOI: `10.1063/1.4952761`.

[21] Andrew M. Childs, Yuan Su, Minh C. Tran, Nathan Wiebe, and Shuchen Zhu. “Theory of Trotter Error with Commutator Scaling”. In: *Physical Review X* 11 (2021), p. 011020. DOI: `10.1103/PhysRevX.11.011020`.

[22] Jonathan Romero, Ryan Babbush, Jarrod R McClean, Cornelius Hempel, Peter J Love, and Alán Aspuru-Guzik. “Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz”. In: *Quantum Science and Technology* 4 (2018), p. 014008. DOI: `10.1088/2058-9565/aad3e4`.

[23] Andrew Tranter, Peter J. Love, Florian Mintert, Nathan Wiebe, and Peter V. Coveney. “Ordering of Trotterization: Impact on Errors in Quantum Simulation of Electronic Structure”. In: *Entropy* 21 (2019), p. 1218. DOI: `10.3390/e21121218`.

[24] Harper R. Grimsley, Daniel Claudino, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall. “Is the Trotterized UCCSD Ansatz Chemically Well-Defined?” In: *Journal of Chemical Theory and Computation* 16 (2020), pp. 1–6. DOI: `10.1021/acs.jctc.9b01083`.

[25] Masuo Suzuki. “Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations”. In: *Physics Letters A* 146 (1990), pp. 319–323. DOI: `10.1016/0375-9601(90)90962-N`.

[26] Dominic W. Berry, Graeme Ahokas, Richard Cleve, and Barry C. Sanders. “Efficient Quantum Algorithms for Simulating Sparse Hamiltonians”. In: *Communications in Mathematical Physics* 270 (2007), pp. 359–371. DOI: `10.1007/s00220-006-0150-x`.

[27] Earl Campbell. “Random Compiler for Fast Hamiltonian Simulation”. In: *Physical Review Letters* 123 (2019), p. 070503. DOI: `10.1103/PhysRevLett.123.070503`.

[28] Yingkai Ouyang, David R. White, and Earl T. Campbell. “Compilation by stochastic Hamiltonian sparsification”. In: *Quantum* 4 (2020), p. 235. DOI: `10.22331/q-2020-02-27-235`.

[29] Kouhei Nakaji, Mohsen Bagherimehrab, and Alán Aspuru-Guzik. “High-Order Randomized Compiler for Hamiltonian Simulation”. In: *PRX Quantum* 5 (2024), p. 020330. DOI: `10.1103/PRXQuantum.5.020330`.

[30] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. “Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics”. In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. ACM, 2019, pp. 193–204. ISBN: 9781450367059. DOI: `10.1145/3313276.3316366`.

[31] Alexis Ralli, Gabriel Greene-Diniz, David Muñoz Ramo, and Nathan Fitzpatrick. “Calculating the Single-Particle Many-body Green’s Functions via the Quantum Singular Value Transform Algorithm”. In: *arXiv preprint* (2023). arXiv: `2307.13583`.

[32] Guang Hao Low and Isaac L. Chuang. “Hamiltonian Simulation by Qubitization”. In: *Quantum* 3 (2019), p. 163. DOI: `10.22331/q-2019-07-12-163`.

[33] John M. Martyn, Zane M. Rossi, Andrew K. Tan, and Isaac L. Chuang. “Grand Unification of Quantum Algorithms”. In: *PRX Quantum* 2 (2021), p. 040203. DOI: `10.1103/PRXQuantum.2.040203`.

[34] Gushu Li, Yufei Ding, and Yuan Xie. “Tackling the Qubit Mapping Problem for NISQ-Era Quantum Devices”. In: *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*. ACM, 2019, pp. 1001–1014. ISBN: 9781450362405. DOI: `10.1145/3297858.3304023`.

[35] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. "Simulated Quantum Computation of Molecular Energies". In: *Science* 309 (2005), pp. 1704–1707. DOI: `10.1126/science.1113479`.

[36] Seunghoon Lee, Joonho Lee, Huanchen Zhai, Yu Tong, Alexander M. Dalzell, Ashutosh Kumar, Phillip Helms, Johnnie Gray, Zhi-Hao Cui, Wenyuan Liu, Michael Kastoryano, Ryan Babbush, John Preskill, David R. Reichman, Earl T. Campbell, Edward F. Valeev, Lin Lin, and Garnet Kin-Lic Chan. "Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry". In: *Nature Communications* 14 (2023), p. 1952. DOI: `10.1038/s41467-023-37587-6`.

[37] Garnet Kin-Lic Chan. "Spiers Memorial Lecture: Quantum chemistry, classical heuristics, and quantum advantage". In: *Faraday Discussions* 254 (2024), pp. 11–52. DOI: `10.1039/D4FD00141A`.

[38] Mario Motta and Julia E. Rice. "Emerging quantum computing algorithms for quantum chemistry". In: *WIREs Computational Molecular Science* 12 (2022), e1580. DOI: `10.1002/wcms.1580`.

[39] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. "PySCF: the Python-based simulations of chemistry framework". In: *WIREs Computational Molecular Science* 8 (2018), e1340. DOI: `10.1002/wcms.1340`.

[40] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory.* Wiley, 2000. ISBN: 9780471967552. DOI: `10.1002/9781119019572`.

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

[42] Alexis Ralli, Tim Weaving, Peter V. Coveney, and Peter J. Love. "Bridging Quantum Chemistry and MaxCut: Classical Performance Guarantees and Quantum Algorithms for the Hartree-Fock Method". In: *arXiv preprint* (2025). arXiv: `2506.04223`.

[43] Anna I. Krylov, C. David Sherrill, Edward F. C. Byrd, and Martin Head-Gordon. "Size-consistent wave functions for nondynamical correlation energy: The valence active space optimized orbital coupled-cluster doubles model". In: *The Journal of Chemical Physics* 109 (1998), pp. 10669–10678. DOI: `10.1063/1.477764`.

[44] Péter Pulay. "Convergence acceleration of iterative sequences. The case of SCF iteration". In: *Chemical Physics Letters* 73 (1980), pp. 393–398. DOI: `10.1016/0009-2614(80)80396-4`.

[45] P. Pulay. "Improved SCF convergence acceleration". In: *Journal of Computational Chemistry* 3 (1982), pp. 556–560. DOI: `10.1002/jcc.540030413`.

[46] Konstantin N. Kudin, Gustavo E. Scuseria, and Eric Cancès. "A black-box self-consistent field convergence algorithm: One step closer". In: *The Journal of Chemical Physics* 116 (2002), pp. 8255–8261. DOI: `10.1063/1.1470195`.

[47] Xiangqian Hu and Weitao Yang. “Accelerating self-consistent field convergence with the augmented Roothaan–Hall energy function”. In: *The Journal of Chemical Physics* 132 (2010). DOI: `10.1063/1.3304922`.

[48] Norbert Schuch and Frank Verstraete. “Computational complexity of interacting electrons and fundamental limitations of density functional theory”. In: *Nature Physics* 5 (2009), pp. 732–735. DOI: `10.1038/nphys1370`.

[49] Rodney J. Bartlett and Monika Musiał. “Coupled-cluster theory in quantum chemistry”. In: *Reviews of Modern Physics* 79 (2007), pp. 291–352. DOI: `10.1103/RevModPhys.79.291`.

[50] T. Daniel Crawford and Henry F. Schaefer. “An Introduction to Coupled Cluster Theory for Computational Chemists”. In: vol. 14. 2000, pp. 33–136. DOI: `10.1002/9780470125915.ch2`.

[51] Andrew G. Taube and Rodney J. Bartlett. “New perspectives on unitary coupled-cluster theory”. In: *International Journal of Quantum Chemistry* 106 (2006), pp. 3393–3401. DOI: `10.1002/qua.21198`.

[52] Yangchao Shen, Xiang Zhang, Shuaining Zhang, Jing-Ning Zhang, Man-Hong Yung, and Kihwan Kim. “Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure”. In: *Physical Review A* 95 (2017), p. 020501. DOI: `10.1103/PhysRevA.95.020501`.

[53] Joonho Lee, William J. Huggins, Martin Head-Gordon, and K. Birgitta Whaley. “Generalized Unitary Coupled Cluster Wave functions for Quantum Computation”. In: *Journal of Chemical Theory and Computation* 15 (2019), pp. 311–324. DOI: `10.1021/acs.jctc.8b01004`.

[54] B. Huron, J. P. Malrieu, and P. Rancurel. “Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions”. In: *The Journal of Chemical Physics* 58 (1973), pp. 5745–5759. DOI: `10.1063/1.1679199`.

[55] Stefano Evangelisti, Jean-Pierre Daudey, and Jean-Paul Malrieu. “Convergence of an improved CIPSI algorithm”. In: *Chemical Physics* 75 (1983), pp. 91–102. DOI: `10.1016/0301-0104(83)85011-3`.

[56] Adam A. Holmes, Norm M. Tubman, and C. J. Umrigar. “Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling”. In: *Journal of Chemical Theory and Computation* 12 (2016), pp. 3674–3680. DOI: `10.1021/acs.jctc.6b00407`.

[57] Sandeep Sharma, Adam A. Holmes, Guillaume Jeanmairet, Ali Alavi, and C. J. Umrigar. “Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory”. In: *Journal of Chemical Theory and Computation* 13 (2017), pp. 1595–1604. DOI: `10.1021/acs.jctc.6b01028`.

[58] Junhao Li, Matthew Otten, Adam A. Holmes, Sandeep Sharma, and C. J. Umrigar. “Fast semistochastic heat-bath configuration interaction”. In: *The Journal of Chemical Physics* 149 (2018). DOI: `10.1063/1.5055390`.

[59] Keita Kanno, Masaya Kohda, Ryosuke Imai, Sho Koh, Kosuke Mitarai, Wataru Mizukami, and Yuya O. Nakagawa. “Quantum-Selected Configuration Interaction: classical diagonalization of Hamiltonians in subspaces selected by quantum computers”. In: *arXiv preprint* (2023). arXiv: `2302.11320`.

[60] Yuya O. Nakagawa, Masahiko Kamoshita, Wataru Mizukami, Shotaro Sudo, and Yu-ya Ohnishi. “ADAPT-QSCI: Adaptive Construction of an Input State for Quantum-Selected Configuration Interaction”. In: *Journal of Chemical Theory and Computation* (2024). DOI: `10.1021/acs.jctc.4c00846`.

[61] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. “Quantum supremacy using a programmable superconducting processor”. In: *Nature* 574 (2019), pp. 505–510. DOI: `10.1038/s41586-019-1666-5`.

[62] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, Peng Hu, Xiao-Yan Yang, Wei-Jun Zhang, Hao Li, Yuxuan Li, Xiao Jiang, Lin Gan, Guangwen Yang, Lixing You, Zhen Wang, Li Li, Nai-Le Liu, Chao-Yang Lu, and Jian-Wei Pan. “Quantum computational advantage using photons”. In: *Science* 370 (2020), pp. 1460–1463. DOI: `10.1126/science.abe8770`.

[63] Yulin Wu, Wan-Su Bao, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, Ming Gong, Cheng Guo, Chu Guo, Shaojun Guo, Lianchen Han, Linyin Hong, He-Liang Huang, Yong-Heng Huo, Liping Li, Na Li, Shaowei Li, Yuan Li, Futian Liang, Chun Lin, Jin Lin, Haoran Qian, Dan Qiao, Hao Rong, Hong Su, Lihua Sun, Liangyuan Wang, Shiyu Wang, Dachao Wu, Yu Xu, Kai Yan, Weifeng Yang, Yang Yang, Yangsen Ye, Jianghan Yin, Chong Ying, Jiale Yu, Chen Zha, Cha Zhang, Haibin Zhang, Kaili Zhang, Yiming Zhang, Han Zhao, Youwei Zhao, Liang Zhou, Qingling Zhu, Chao-Yang Lu, Cheng-Zhi Peng, Xiaobo Zhu, and Jian-Wei Pan. “Strong Quantum Computational Advantage Using a Superconducting Quantum Processor”. In: *Phys-*

ical Review Letters 127 (2021), p. 180501. DOI: `10.1103/PhysRevLett.127.180501`.

[64] Lars S. Madsen, Fabian Laudenbach, Mohsen Falamarzi. Askarani, Fabien Rortais, Trevor Vincent, Jacob F. F. Bulmer, Filippo M. Miatto, Leonhard Neuhaus, Lukas G. Helt, Matthew J. Collins, Adriana E. Lita, Thomas Gerrits, Sae Woo Nam, Varun D. Vaidya, Matteo Menotti, Ish Dhand, Zachary Vernon, Nicolás Quesada, and Jonathan Lavoie. “Quantum computational advantage with a programmable photonic processor”. In: *Nature* 606 (2022), pp. 75–81. DOI: `10.1038/s41586-022-04725-x`.

[65] Ayoub Hafid, Hokuto Iwakiri, Kento Tsubouchi, Nobuyuki Yoshioka, and Masaya Kohda. “Hardness of classically sampling quantum chemistry circuits”. In: *arXiv preprint* (2025). arXiv: `2504.12893`.

[66] Javier Robledo-Moreno, Mario Motta, Holger Haas, Ali Javadi-Abhari, Petar Jurcevic, William Kirby, Simon Martiel, Kunal Sharma, Sandeep Sharma, Tomonori Shirakawa, Iskandar Sitdikov, Rong-Yang Sun, Kevin J. Sung, Maika Takita, Minh C. Tran, Seiji Yunoki, and Antonio Mezzacapo. “Chemistry Beyond Exact Solutions on a Quantum-Centric Supercomputer”. In: *arXiv preprint* (2024). arXiv: `2405.05068`.

[67] Danil Kaliakin, Akhil Shajan, Javier Robledo Moreno, Zhen Li, Abhishek Mitra, Mario Motta, Caleb Johnson, Abdullah Ash Saki, Susanta Das, Iskandar Sitdikov, Antonio Mezzacapo, and Kenneth M. Merz. “Accurate quantum-centric simulations of supramolecular interactions”. In: *arXiv preprint* (2024). arXiv: `2410.09209`.

[68] Ieva Liepuoniute, Kirstin D. Doney, Javier Robledo-Moreno, Joshua A. Job, Will S. Friend, and Gavin O. Jones. “Quantum-Centric Study of Methylene Singlet and Triplet States”. In: *arXiv preprint* (2024). arXiv: `2411.04827`.

[69] Stefano Barison, Javier Robledo Moreno, and Mario Motta. “Quantum-centric computation of molecular excited states with extended sample-based quantum diagonalization”. In: *arXiv preprint* (2024). arXiv: `2411.00468`.

[70] Akhil Shajan, Danil Kaliakin, Abhishek Mitra, Javier Robledo Moreno, Zhen Li, Mario Motta, Caleb Johnson, Abdullah Ash Saki, Susanta Das, Iskandar Sitdikov, Antonio Mezzacapo, and Kenneth M. Merz. “Towards quantum-centric simulations of extended molecules: sample-based quantum diagonalization enhanced with density matrix embedding theory”. In: *arXiv preprint* (2024). arXiv: `2411.09861`.

[71] Thomas M Bickley, Angus Mingare, Tim Weaving, Michael Williams de la Bastida, Shunzhou Wan, Martina Nibbi, Philipp Seitz, Alexis Ralli, Peter J Love, Minh Chung, et al. “Extending Quantum Computing through Subspace, Embedding and Classical Molecular Dynamics Techniques”. In: *arXiv preprint* (2025). arXiv: `2505.16796`.

[72] Yuri Alexeev et al. “Quantum-centric supercomputing for materials science: A perspective on challenges and future directions”. In: *Future Generation Computer Systems* 160 (2024), pp. 666–710. DOI: `10.1016/j.future.2024.04.060`.

[73] P. Jordan and E. Wigner. "Über das Paulische Äquivalenzverbot". In: *Zeitschrift für Physik* 47 (1928), pp. 631–651. DOI: `10.1007/BF01331938`.

[74] Sergey B. Bravyi and Alexei Yu. Kitaev. "Fermionic Quantum Computation". In: *Annals of Physics* 298 (2002), pp. 210–226. DOI: `10.1006/aphy.2002.6254`.

[75] Jacob T. Seeley, Martin J. Richard, and Peter J. Love. "The Bravyi-Kitaev transformation for quantum computation of electronic structure". In: *The Journal of Chemical Physics* 137 (2012). DOI: `10.1063/1.4768229`.

[76] Andrew Tranter, Peter J. Love, Florian Mintert, and Peter V. Coveney. "A Comparison of the Bravyi–Kitaev and Jordan–Wigner Transformations for the Quantum Simulation of Quantum Chemistry". In: *Journal of Chemical Theory and Computation* 14 (2018), pp. 5617–5630. DOI: `10.1021/acs.jctc.8b00450`.

[77] Aaron Miller, Zoltán Zimborás, Stefan Knecht, Sabrina Maniscalco, and Guillermo García-Pérez. "Bonsai Algorithm: Grow Your Own Fermion-to-Qubit Mappings". In: *PRX Quantum* 4 (2023), p. 030314. DOI: `10.1103/PRXQuantum.4.030314`.

[78] Charles Derby, Joel Klassen, Johannes Bausch, and Toby Cubitt. "Compact fermion to qubit mappings". In: *Physical Review B* 104 (2021), p. 035118. DOI: `10.1103/PhysRevB.104.035118`.

[79] William M. Kirby, Sultana Hadi, Michael Kreshchuk, and Peter J. Love. "Quantum simulation of second-quantized Hamiltonians in compact encoding". In: *Physical Review A* 104 (2021), p. 042607. DOI: `10.1103/PhysRevA.104.042607`.

[80] William Kirby, Bryce Fuller, Charles Hadfield, and Antonio Mezzacapo. "Second-Quantized Fermionic Operators with Polylogarithmic Qubit and Gate Complexity". In: *PRX Quantum* 3 (2022), p. 020351. DOI: `10.1103/PRXQuantum.3.020351`.

[81] Kentaro Yamamoto, Samuel Duffield, Yuta Kikuchi, and David Muñoz Ramo. "Demonstrating Bayesian quantum phase estimation with quantum error detection". In: *Physical Review Research* 6 (2024), p. 013221. DOI: `10.1103/PhysRevResearch.6.013221`.

[82] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. "A variational eigenvalue solver on a photonic quantum processor". In: *Nature Communications* 5 (2014), p. 4213. DOI: `10.1038/ncomms5213`.

[83] Yangchao Shen, Xiang Zhang, Shuaining Zhang, Jing-Ning Zhang, Man-Hong Yung, and Kihwan Kim. "Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure". In: *Physical Review A* 95 (2017), p. 020501. DOI: `10.1103/PhysRevA.95.020501`.

[84] 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. "Scalable Quantum Simulation of Molecular Energies". In: *Physical Review X* 6 (2016), p. 031007. DOI: `10.1103/PhysRevX.6.031007`.

[85] Raffaele Santagati, Jianwei Wang, Antonio A. Gentile, Stefano Paesani, Nathan Wiebe, Jarrod R. McClean, Sam Morley-Short, Peter J. Shadbolt, Damien Bonneau, Joshua W. Silverstone, David P. Tew, Xiaoqi Zhou, Jeremy L. O'Brien, and Mark G. Thompson. "Witnessing eigenstates for quantum simulation of Hamiltonian spectra". In: *Science Advances* 4 (2018), pp. 1–12. DOI: `10.1126/sciadv.aap9646`.

[86] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets". In: *Nature* 549 (2017), pp. 242–246. DOI: `10.1038/nature23879`.

[87] 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. "Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm". In: *Physical Review X* 8 (2018), p. 011021. DOI: `10.1103/PhysRevX.8.011021`.

[88] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P. Lanyon, Peter Love, Ryan Babbush, Alán Aspuru-Guzik, Rainer Blatt, and Christian F. Roos. "Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator". In: *Physical Review X* 8 (2018), p. 031022. DOI: `10.1103/PhysRevX.8.031022`.

[89] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. "Error mitigation extends the computational reach of a noisy quantum processor". In: *Nature* 567 (2019), pp. 491–495. DOI: `10.1038/s41586-019-1040-7`.

[90] Yunseong Nam, Jwo-Sy Chen, Neal C. Pisenti, Kenneth Wright, Conor Delaney, Dmitri Maslov, Kenneth R. Brown, Stewart Allen, Jason M. Amini, Joel Apisdorf, Kristin M. Beck, Aleksey Blinov, Vandiver Chaplin, Mika Chmielewski, Coleman Collins, Shantanu Debnath, Kai M. Hudek, Andrew M. Ducore, Matthew Keesan, Sarah M. Kreikemeier, Jonathan Mizrahi, Phil Solomon, Mike Williams, Jaime David Wong-Campos, David Moehring, Christopher Monroe, and Jungsang Kim. "Ground-state energy estimation of the water molecule on a trapped-ion quantum computer". In: *npj Quantum Information* 6 (2020), p. 33. DOI: `10.1038/s41534-020-0259-3`.

[91] Scott E. Smart and David A. Mazziotti. "Quantum-classical hybrid algorithm using an error-mitigating N-representability condition to compute the Mott metal-insulator transition". In: *Physical Review A* 100 (2019), p. 022517. DOI: `10.1103/PhysRevA.100.022517`.

[92] Alexander J. McCaskey, Zachary P. Parks, Jacek Jakowski, Shirley V. Moore, Titus D. Morris, Travis S. Humble, and Raphael C. Pooser. "Quantum chemistry as a benchmark for near-term quantum computers". In: *npj Quantum Information* 5 (2019), p. 99. DOI: `10.1038/s41534-019-0209-0`.

[93] Julia E. Rice, Tanvi P. Gujarati, Mario Motta, Tyler Y. Takeshita, Eunseok Lee, Joseph A. Latone, and Jeannette M. Garcia. "Quantum computation of dominant products in lithium–sulfur batteries". In: *The Journal of Chemical Physics* 154 (2021), p. 134115. DOI: `10.1063/5.0044068`.

[94] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Edward Farhi, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Matthew P. Harrigan, Alan Ho, Sabrina Hong, Trent Huang, William J. Huggins, Lev Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Erik Lucero, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Hartmut Neven, Murphy Yuezhen Niu, Thomas E. O'Brien, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Doug Strain, Kevin J. Sung, Marco Szalay, Tyler Y. Takeshita, Amit Vainsencher, Theodore White, Nathan Wiebe, Z. Jamie Yao, Ping Yeh, and Adam Zalcman. "Hartree-Fock on a superconducting qubit quantum computer". In: *Science* 369 (2020), pp. 1084–1089. DOI: `10.1126/science.abb9811`.

[95] Qi Gao, Gavin O. Jones, Mario Motta, Michihiko Sugawara, Hiroshi C. Watanabe, Takao Kobayashi, Eriko Watanabe, Yu-ya Ohnishi, Hajime Nakamura, and Naoki Yamamoto. "Applications of quantum computing for investigations of electronic transitions in phenylsulfonyl-carbazole TADF emitters". In: *npj Computational Materials* 7 (2021), p. 70. DOI: `10.1038/s41524-021-00540-6`.

[96] Yukio Kawashima, Erika Lloyd, Marc P. Coons, Yunseong Nam, Shunji Matsuura, Alejandro J. Garza, Sonika Johri, Lee Huntington, Valentin Senicourt, Andrii O. Maksymov, Jason H. V. Nguyen, Jungsang Kim, Nima Alidoust, Arman Zaribafiyan, and Takeshi Yamazaki. "Optimizing electronic structure simulations on a trapped-ion quantum computer using problem decomposition". In: *Communications Physics* 4 (2021), p. 245. DOI: `10.1038/s42005-021-00751-9`.

[97] Andrew Eddins, Mario Motta, Tanvi P. Gujarati, Sergey Bravyi, Antonio Mezzacapo, Charles Hadfield, and Sarah Sheldon. "Doubling the Size of Quantum Simulators by Entanglement Forging". In: *PRX Quantum* 3 (2022), p. 010309. DOI: `10.1103/PRXQuantum.3.010309`.

[98] Kentaro Yamamoto, David Zsolt Manrique, Irfan T Khan, Hideaki Sawada, and David Muñoz Ramo. "Quantum hardware calculations of periodic systems with partition-measurement symmetry verification: Simplified models of hydrogen chain and iron crystals". In: *Physical Review Research* 4 (2022), p. 33110. DOI: `10.1103/PhysRevResearch.4.033110`.

[99] Josh J M Kirsopp, Cono Di Paola, David Zsolt Manrique, Michal Krompiec, Gabriel Greene-Diniz, Wolfgang Guba, Agnes Meyder, Detlef Wolf, Martin Strahm, and David Muñoz Ramo. "Quantum computational quantification of protein–ligand interactions". In: *International Journal of Quantum Chemistry* 122 (2022), pp. 1–16. DOI: `10.1002/qua.26975`.

[100] Kaixuan Huang, Xiaoxia Cai, Hao Li, Zi-Yong Ge, Ruijuan Hou, Hekang Li, Tong Liu, Yunhao Shi, Chitong Chen, Dongning Zheng, et al. "Variational quantum computation of molecular linear response properties on a superconducting quantum processor". In: *The Journal of Physical Chemistry Letters* 13 (2022), pp. 9114–9121. DOI: `10.1021/acs.jpclett.2c02381`.

[101] Phalgun Lolur, Mårten Skogh, Werner Dobrautz, Christopher Warren, Janka Biznárová, Amr Osman, Giovanna Tancredi, Göran Wendin, Jonas Bylander, and Martin Rahm. "Reference-State Error Mitigation: A Strategy for High Accuracy Quantum Computation of Chemistry". In: *Journal of Chemical Theory and Computation* 19 (2023), pp. 783–789. DOI: `10.1021/acs.jctc.2c00807`.

[102] Vicente Leyton-Ortega, Swarnadeep Majumder, and Raphael C Pooser. "Quantum error mitigation by hidden inverses protocol in superconducting quantum devices". In: *Quantum Science and Technology* 8 (2022), p. 14008. DOI: `10.1088/2058-9565/aca92d`.

[103] Zhiding Liang, Jinglei Cheng, Hang Ren, Hanrui Wang, Fei Hua, Zhixin Song, Yongshan Ding, Fred Chong, Song Han, Xuehai Qian, and Yiyu Shi. "NAPA: Intermediate-level Variational Native-pulse Ansatz for Variational Quantum Algorithms". In: *arXiv preprint* (2022). arXiv: `2208.01215`.

[104] Mario Motta, Gavin O. Jones, Julia E. Rice, Tanvi P. Gujarati, Rei Sakuma, Ieva Liepuoniute, Jeannette M. Garcia, and Yu-ya Ohnishi. "Quantum chemistry simulation of ground- and excited-state properties of the sulfonium cation on a superconducting quantum processor". In: *Chemical Science* 14 (2023), pp. 2915–2927. DOI: `10.1039/D2SC06019A`.

[105] T. E. O'Brien et al. "Purification-based quantum error mitigation of pair-correlated electron simulations". In: *Nature Physics* 19 (2023), pp. 1787–1792. DOI: `10.1038/s41567-023-02240-y`.

[106] I T Khan, M Tudorovskaya, J J M Kirsopp, D Muñoz Ramo, P Warrier, D K Papanastasiou, and R Singh. "Chemically aware unitary coupled cluster with ab initio calculations on an ion trap quantum computer: A refrigerant chemicals' application". In: *The Journal of Chemical Physics* 158 (2023). DOI: `10.1063/5.0144680`.

[107] Luning Zhao, Joshua Goings, Kyujin Shin, Woomin Kyoung, Johanna I Fuks, June-Koo Kevin Rhee, Young Min Rhee, Kenneth Wright, Jason Nguyen, Jungsang Kim, et al. "Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers". In: *npj Quantum Information* 9 (2023), p. 60. DOI: `10.1038/s41534-023-00730-8`.

[108] Shaojun Guo, Jinzhao Sun, Haoran Qian, Ming Gong, Yukun Zhang, Fusheng Chen, Yangsen Ye, Yulin Wu, Sirui Cao, Kun Liu, Chen Zha, Chong Ying, Qingling Zhu, He-Liang Huang, Youwei Zhao, Shaowei Li, Shiyu Wang, Jiale Yu, Daojin Fan, Dachao Wu, Hong Su, Hui Deng, Hao Rong, Yuan Li, Kaili Zhang, Tung-Hsun Chung, Futian Liang, Jin Lin, Yu Xu, Lihua Sun, Cheng Guo, Na Li, Yong-Heng Huo, Cheng-Zhi Peng, Chao-Yang Lu, Xiao Yuan, Xiaobo Zhu, and Jian-Wei Pan. "Experimental quantum computational chemistry with optimized unitary coupled cluster ansatz". In: *Nature Physics* 20 (2024), pp. 1240–1246. DOI: `10.1038/s41567-024-02530-z`.

[109] Pei Liu, Ruixia Wang, Jing-Ning Zhang, Yingshan Zhang, Xiaoxia Cai, Huikai Xu, Zhiyuan Li, Jiaxiu Han, Xuegang Li, Guangming Xue, et al. "Performing SU(d) Operations and Rudimentary Algorithms in a Superconducting Transmon Qudit for d=3 and d=4". In: *Physical Review X* 13 (2023), p. 21028. DOI: `10.1103/PhysRevX.13.021028`.

[110] Emil Dimitrov, Goar Sanchez-Sanz, James Nelson, Lee O'Riordan, Myles Doyle, Sean Courtney, Venkatesh Kannan, Hassan Naseri, Alberto Garcia Garcia, James Tricker, Marisa Faraggi, Joshua Goings, and Luning Zhao. "Pushing the Limits of Quantum Computing for Simulating PFAS Chemistry". In: *arXiv preprint* (2023). arXiv: `2311.01242`.

[111] Michael A. Jones, Harish J. Vallury, and Lloyd C.L. Hollenberg. "Ground-state-energy calculation for the water molecule on a superconducting quantum processor". In: *Physical Review Applied* 21 (2024), p. 064017. DOI: `10.1103/PhysRevApplied.21.064017`.

[112] Zhiding Liang, Zhixin Song, Jinglei Cheng, Hang Ren, Tianyi Hao, Rui Yang, Yiyu Shi, and Tongyang Li. "SpacePulse: Combining Parameterized Pulses and Contextual Subspace for More Practical VQE". In: *arXiv preprint* (2023). arXiv: `2311.17423`.

[113] Oscar Higgott, Daochen Wang, and Stephen Brierley. "Variational quantum computation of excited states". In: *Quantum* 3 (2019), p. 156. DOI: `10.22331/q-2019-07-01-156`.

[114] Lila Cadi Tazi and Alex JW Thom. "Folded spectrum vqe: A quantum computing method for the calculation of molecular excited states". In: *Journal of Chemical Theory and Computation* 20.6 (2024), pp. 2491–2504. DOI: `10.1021/acs.jctc.3c01378`.

[115] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F. Izmaylov. "Measurement optimization in the variational quantum eigensolver using a minimum clique cover". In: *The Journal of Chemical Physics* 152 (2020). DOI: `10.1063/1.5141458`.

[116] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F. Izmaylov. "Measuring All Compatible Operators in One Series of Single-Qubit Measurements Using Unitary Transformations". In: *Journal of Chemical Theory and Computation* 16 (2020), pp. 2400–2409. DOI: `10.1021/acs.jctc.0c00008`.

[117] Artur F. Izmaylov, Tzu-Ching Yen, Robert A. Lang, and Vladyslav Verteletskyi. "Unitary Partitioning Approach to the Measurement Problem in the Variational Quantum Eigensolver Method". In: *Journal of Chemical Theory and Computation* 16 (2020), pp. 190–195. DOI: `10.1021/acs.jctc.9b00791`.

[118] Andrew Zhao, Andrew Tranter, William M. Kirby, Shu Fay Ung, Akimasa Miyake, and Peter J. Love. "Measurement reduction in variational quantum algorithms". In: *Physical Review A* 101 (2020), p. 062322. DOI: `10.1103/PhysRevA.101.062322`.

[119] Alexis Ralli, Peter J. Love, Andrew Tranter, and Peter V. Coveney. "Implementation of measurement reduction for the variational quantum eigensolver". In: *Physical Review Research* 3 (2021), p. 033195. DOI: `10.1103/PhysRevResearch.3.033195`.

[120] Alexis Ralli, Tim Weaving, Andrew Tranter, William M. Kirby, Peter J. Love, and Peter V. Coveney. "Unitary partitioning and the contextual subspace variational quantum eigensolver". In: *Physical Review Research* 5 (2023), p. 013095. DOI: `10.1103/PhysRevResearch.5.013095`.

[121] Harper R. Grimsley, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall. "An adaptive variational algorithm for exact molecular simulations on a quantum computer". In: *Nature Communications* 10 (2019), p. 3007. DOI: `10.1038/s41467-019-10988-2`.

[122] Robert M. Parrish, Edward G. Hohenstein, Peter L. McMahon, and Todd J. Martinez. "Hybrid Quantum/Classical Derivative Theory: Analytical Gradients and Excited-State Dynamics for the Multistate Contracted Variational Quantum Eigensolver". In: *arXiv preprint* (2019). arXiv: `1906.08728`.

[123] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. "Barren plateaus in quantum neural network training landscapes". In: *Nature Communications* 9 (2018), p. 4812. DOI: `10.1038/s41467-018-07090-4`.

[124] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. "Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus". In: *PRX Quantum* 3 (2022), p. 010313. DOI: `10.1103/PRXQuantum.3.010313`.

[125] M. Cerezo, Martin Larocca, Diego García-Martín, N. L. Diaz, Paolo Braccia, Enrico Fontana, Manuel S. Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, Eric R. Anschuetz, and Zoë Holmes. "Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing". In: *arXiv preprint* (2023). arXiv: `2312.09121`.

[126] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. “Noise-induced barren plateaus in variational quantum algorithms”. In: *Nature Communications* 12 (2021), p. 6961. DOI: `10.1038/s41467-021-27045-6`.

[127] Lorenzo Leone, Salvatore F.E. Oliviero, Lukasz Cincio, and M. Cerezo. “On the practical usefulness of the Hardware Efficient Ansatz”. In: *Quantum* 8 (2024), p. 1395. DOI: `10.22331/q-2024-07-03-1395`.

[128] Mario Motta, Kevin J. Sung, K. Birgitta Whaley, Martin Head-Gordon, and James Shee. “Bridging physical intuition and hardware efficiency for correlated electronic states: the local unitary cluster Jastrow ansatz for electronic structure”. In: *Chemical Science* 14 (2023), pp. 11213–11227. DOI: `10.1039/D3SC02516K`.

[129] Bryan T. Gard, Linghua Zhu, George S. Barron, Nicholas J. Mayhall, Sophia E. Economou, and Edwin Barnes. “Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm”. In: *npj Quantum Information* 6 (2020), p. 10. DOI: `10.1038/s41534-019-0240-1`.

[130] Rui Mao, Guojing Tian, and Xiaoming Sun. “Towards determining the presence of barren plateaus in some chemically inspired variational quantum algorithms”. In: *Communications Physics* 7 (2024), p. 342. DOI: `10.5281/zenodo.13359192`.

[131] Ho Lun Tang, V.O. Shkolnikov, George S. Barron, Harper R. Grimsley, Nicholas J. Mayhall, Edwin Barnes, and Sophia E. Economou. “Qubit-ADAPT-VQE: An Adaptive Algorithm for Constructing Hardware-Efficient Ansätze on a Quantum Processor”. In: *PRX Quantum* 2 (2021), p. 020310. DOI: `10.1103/PRXQuantum.2.020310`.

[132] Harper R. Grimsley, George S. Barron, Edwin Barnes, Sophia E. Economou, and Nicholas J. Mayhall. “Adaptive, problem-tailored variational quantum eigensolver mitigates rough parameter landscapes and barren plateaus”. In: *npj Quantum Information* 9 (2023), p. 19. DOI: `10.1038/s41534-023-00681-0`.

[133] Alpár Jüttner and Péter Madarasi. “VF2++—An improved subgraph isomorphism algorithm”. In: *Discrete Applied Mathematics* 242 (2018), pp. 69–81. DOI: `10.1016/j.dam.2018.02.018`.

[134] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. “Exploring Network Structure, Dynamics, and Function using NetworkX”. In: 2008, pp. 11–15. DOI: `10.25080/TCWV9851`.

[135] P. W. Anderson. “More Is Different”. In: *Science* 177 (1972), pp. 393–396. DOI: `10.1126/science.177.4047.393`.

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

[137] Kanav Setia, Richard Chen, Julia E. Rice, Antonio Mezzacapo, Marco Pistoia, and James D. Whitfield. "Reducing Qubit Requirements for Quantum Simulations Using Molecular Point Group Symmetries". In: *Journal of Chemical Theory and Computation* 16 (2020), pp. 6091–6097. DOI: `10.1021/acs.jctc.0c00113`.

[138] Daniel Gottesman. "An introduction to quantum error correction and fault-tolerant quantum computation". In: vol. 68. 2010, pp. 13–58. DOI: `10.1090/psapm/068/2762145`.

[139] Jeroen Dehaene and Bart De Moor. "Clifford group, stabilizer states, and linear and quadratic operations over GF(2)". In: *Physical Review A* 68 (2003), p. 042318. DOI: `10.1103/PhysRevA.68.042318`.

[140] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2012. ISBN: 9781107002173. DOI: `10.1017/CBO9780511976667`.

[141] Karl Kraus, Arno Böhm, John D Dollard, and W H Wootters. *States, Effects, and Operations Fundamental Notions of Quantum Theory.* Ed. by Karl Kraus, A. Böhm, J. D. Dollard, and W. H. Wootters. Vol. 190. Springer Berlin Heidelberg, 1983. ISBN: 978-3-540-12732-1. DOI: `10.1007/3-540-12732-1`.

[142] Julia Kempe, Alexei Kitaev, and Oded Regev. "The Complexity of the Local Hamiltonian Problem". In: vol. 35. SIAM, 2004, pp. 372–383. DOI: `10.1007/978-3-540-30538-5_31`.

[143] William M. Kirby and Peter J. Love. "Classical simulation of noncontextual Pauli Hamiltonians". In: *Physical Review A* 102 (2020), p. 032418. DOI: `10.1103/PhysRevA.102.032418`.

[144] Chr. Møller and M. S. Plesset. "Note on an Approximation Treatment for Many-Electron Systems". In: *Physical Review* 46 (1934), pp. 618–622. DOI: `10.1103/PhysRev.46.618`.

[145] Dieter Cremer. "Møller–Plesset perturbation theory: from small molecule methods to methods for thousands of atoms". In: *WIREs Computational Molecular Science* 1 (2011), pp. 509–530. DOI: `10.1002/wcms.58`.

[146] ISAIAH SHAVITT. "The history and evolution of configuration interaction". In: *Molecular Physics* 94 (1998), pp. 3–17. DOI: `10.1080/002689798168303`.

[147] Troy Van Voorhis and Martin Head-Gordon. "Benchmark variational coupled cluster doubles results". In: *The Journal of Chemical Physics* 113 (2000), pp. 8873–8879. DOI: `10.1063/1.1319643`.

[148] C. Genovese, A. Meninno, and S. Sorella. "Assessing the accuracy of the Jastrow antisymmetrized geminal power in the H4 model system". In: *The Journal of Chemical Physics* 150 (2019). DOI: `10.1063/1.5081933`.

[149] Igor O. Sokolov, Panagiotis Kl. Barkoutsos, Pauline J. Ollitrault, Donny Greenberg, Julia Rice, Marco Pistoia, and Ivano Tavernelli. "Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents?" In: *The Journal of Chemical Physics* 152 (2020). DOI: `10.1063/1.5141835`.

[150] Hugh GA Burton and Alex JW Thom. "Holomorphic Hartree-Fock theory: an inherently multireference approach". In: *Journal of chemical theory and computation* 12.1 (2016), pp. 167–173. DOI: `10.1021/acs.jctc.5b01005`.

[151] A. Yu. Kitaev. "Quantum measurements and the Abelian Stabilizer Problem". In: *arXiv preprint* (1995). arXiv: `quant-ph/9511026`.

[152] Werner Heisenberg. *Physics and philosophy: The revolution in modern science.* Vladimir Djambov, 1958.

[153] Pierre Simon Laplace. *A Philosophical Essay on Probabilities, translated into English from the original French 6th ed. by Truscott, F.W. and Emory, F.L.* Dover Publications, 1951.

[154] James Clerk Maxwell. *Theory of Heat.* Cambridge University Press, 2011. ISBN: 9781108032018. DOI: `10.1017/CBO9781139057943`.

[155] Paul Ehrenfest. "Welche Züge der Lichtquantenhypothese spielen in der Theorie der Wärmestrahlung eine wesentliche Rolle?" In: *Annalen der Physik* 341 (1911), pp. 91–118.

[156] Max Planck. *The theory of heat radiation.* Blakiston, 1914.

[157] Werner Heisenberg. *Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen.* Springer, 1985. DOI: `10.1007/BF01328377`.

[158] E. Schrödinger. "An Undulatory Theory of the Mechanics of Atoms and Molecules". In: *Physical Review* 28 (1926), pp. 1049–1070. DOI: `10.1103/PhysRev.28.1049`.

[159] Max Born. "Quantenmechanik der stoßvorgänge". In: *Zeitschrift für physik* 38 (1926), pp. 803–827.

[160] A Einstein, B Podolsky, and N Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In: *Phys. Rev.* 47 (1935), pp. 777–780. DOI: `10.1103/PhysRev.47.777`.

[161] J. S. Bell. "On the Einstein Podolsky Rosen paradox". In: *Physics Physique Fizika* 1 (1964), pp. 195–200. DOI: `10.1103/PhysicsPhysiqueFizika.1.195`.

[162] Vladimir B. Braginsky, Farid Ya Khalili, and Kip S. Thorne. *Quantum Measurement.* Cambridge University Press, 1992. ISBN: 9780521419284. DOI: `10.1017/CBO9780511622748`.

[163] James L. Park and Henry Margenau. "Simultaneous measurability in quantum theory". In: *International Journal of Theoretical Physics* 1 (1968), pp. 211–283. DOI: `10.1007/BF00668668`.

[164] Samson Abramsky and Adam Brandenburger. "The sheaf-theoretic structure of non-locality and contextuality". In: *New Journal of Physics* 13 (2011), p. 113036. DOI: `10.1088/1367-2630/13/11/113036`.

[165] Robert Raussendorf, Juani Bermejo-Vega, Emily Tyhurst, Cihan Okay, and Michael Zurel. "Phase-space-simulation method for quantum computation with magic states on qubits". In: *Physical Review A* 101 (2020), p. 012350. DOI: `10.1103/PhysRevA.101.012350`.

[166] William M. Kirby, Andrew Tranter, and Peter J. Love. “Contextual Subspace Variational Quantum Eigensolver”. In: *Quantum* 5 (2021), p. 456. DOI: `10.22331/q-2021-05-14-456`.

[167] JOHN S. BELL. “On the Problem of Hidden Variables in Quantum Mechanics”. In: *Reviews of Modern Physics* 38 (1966), pp. 447–452. DOI: `10.1103/RevModPhys.38.447`.

[168] Asher Peres. “Incompatible results of quantum measurements”. In: *Physics Letters A* 151 (1990), pp. 107–108. DOI: `10.1016/0375-9601(90)90172-K`.

[169] N. David Mermin. “Simple unified form for the major no-hidden-variables theorems”. In: *Physical Review Letters* 65 (1990), pp. 3373–3376. DOI: `10.1103/PhysRevLett.65.3373`.

[170] William M. Kirby and Peter J. Love. “Contextuality Test of the Nonclassicality of Variational Quantum Eigensolvers”. In: *Physical Review Letters* 123 (2019), p. 200501. DOI: `10.1103/PhysRevLett.123.200501`.

[171] Jarrod R McClean, Nicholas C Rubin, Kevin J Sung, Ian D Kivlichan, Xavier Bonet-Monroig, Yudong Cao, Chengyu Dai, E Schuyler Fried, Craig Gidney, Brendan Gimby, Pranav Gokhale, Thomas Häner, Tarini Hardikar, Vojtěch Havlíček, Oscar Higgott, Cupjin Huang, Josh Izaac, Zhang Jiang, Xinle Liu, Sam McArdle, Matthew Neeley, Thomas O’Brien, Bryan O’Gorman, Isil Ozfidan, Maxwell D Radin, Jhonathan Romero, Nicolas P D Sawaya, Bruno Senjean, Kanav Setia, Sukin Sim, Damian S Steiger, Mark Steudtner, Qiming Sun, Wei Sun, Daochen Wang, Fang Zhang, and Ryan Babbush. “OpenFermion: the electronic structure package for quantum computers”. In: *Quantum Science and Technology* 5 (2020), p. 034014. DOI: `10.1088/2058-9565/ab8ebc`.

[172] Alexis Ralli, Tim Weaving, and Peter J Love. “Noncontextual Pauli Hamiltonians”. In: *arXiv preprint* (2025). arXiv: `2506.19778`.

[173] Robert W Spekkens. “Contextuality for preparations, transformations, and unsharp measurements”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 71.5 (2005), p. 052108. DOI: `10.1103/PhysRevA.71.052108`.

[174] Robert W. Spekkens. “Evidence for the epistemic view of quantum states: A toy theory”. In: *Physical Review A* 75 (2007), p. 032110. DOI: `10.1103/PhysRevA.75.032110`.

[175] Robert W. Spekkens. “Quasi-Quantization: Classical Statistical Theories with an Epistemic Restriction”. In: Springer Netherlands, 2016, pp. 83–135. ISBN: 978-94-017-7303-4. DOI: `10.1007/978-94-017-7303-4_4`.

[176] Robert W. Spekkens. “Negativity and Contextuality are Equivalent Notions of Nonclassicality”. In: *Physical Review Letters* 101 (2008), p. 020401. DOI: `10.1103/PhysRevLett.101.020401`.

[177] Simon Kochen and E. P. Specker. “The Problem of Hidden Variables in Quantum Mechanics”. In: Springer Netherlands, 1975, pp. 293–328. DOI: `10.1007/978-94-010-1795-4_17`.

[178] Soran Jahangiri, Diego Guala, Utkarsh Azad, and Juan Miguel Arrazola. *Pennylane Blog: Top 20 molecules for Quantum Computing.* 2024. URL: `https://pennylane.ai/blog/2024/01/top-20-molecules-for-quantum-computing#1-hydrogen-hsub2sub`.

[179] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien, and M. G. Thompson. "Experimental Bayesian Quantum Phase Estimation on a Silicon Photonic Chip". In: *Physical Review Letters* 118 (2017), p. 100503. DOI: `10.1103/PhysRevLett.118.100503`.

[180] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O'Brien. "Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States". In: *Physical Review X* 10 (2020), p. 031064. DOI: `10.1103/PhysRevX.10.031064`.

[181] Peter Schwerdtfeger. "The Pseudopotential Approximation in Electronic Structure Theory". In: *ChemPhysChem* 12 (2011), pp. 3143–3155. DOI: `10.1002/cphc.201100387`.

[182] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer. "Gate-count estimates for performing quantum chemistry on small quantum computers". In: *Physical Review A* 90 (2014), p. 022305. DOI: `10.1103/PhysRevA.90.022305`.

[183] David Poulin, M. B. Hastings, Dave Wecker, Nathan Wiebe, Andrew C. Doherty, and Matthias Troyer. "The Trotter Step Size Required for Accurate Quantum Simulation of Quantum Chemistry". In: *arXiv preprint* (2014). arXiv: `1406.4920`.

[184] Karol Kowalski. "Properties of coupled-cluster equations originating in excitation sub-algebras". In: *The Journal of Chemical Physics* 148 (2018), p. 94104. DOI: `10.1063/1.5010693`.

[185] Nicholas P. Bauman, Eric J. Bylaska, Sriram Krishnamoorthy, Guang Hao Low, Nathan Wiebe, Christopher E. Granade, Martin Roetteler, Matthias Troyer, and Karol Kowalski. "Downfolding of many-body Hamiltonians using active-space models: Extension of the sub-system embedding sub-algebras approach to unitary coupled cluster formalisms". In: *The Journal of Chemical Physics* 151 (2019), p. 14107. DOI: `10.1063/1.5094643`.

[186] Russell D Johnson III. *NIST Computational Chemistry Comparison and Benchmark Database.* 2020. DOI: `10.18434/T47C7Z`.

[187] M D SAJID ANIS et al. *Qiskit: An Open-source Framework for Quantum Computing.* 2021. DOI: `10.5281/zenodo.2573505`.

[188] V. O. Shkolnikov, Nicholas J. Mayhall, Sophia E. Economou, and Edwin Barnes. "Avoiding symmetry roadblocks and minimizing the measurement overhead of adaptive variational quantum eigensolvers". In: *Quantum* 7 (2023), p. 1040. DOI: `10.22331/q-2023-06-12-1040`.

[189] Dmitry A. Fedorov, Yuri Alexeev, Stephen K. Gray, and Matthew Otten. "Unitary Selective Coupled-Cluster Method". In: *Quantum* 6 (2022), p. 703. DOI: `10.22331/q-2022-05-02-703`.

[190] Ilya G. Ryabinkin, Robert A. Lang, Scott N. Genin, and Artur F. Izmaylov. "Iterative Qubit Coupled Cluster Approach with Efficient Screening of Generators". In: *Journal of Chemical Theory and Computation* 16 (2020), pp. 1055–1063. DOI: `10.1021/acs.jctc.9b01084`.

[191] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint* (2014). arXiv: `1412.6980`.

[192] David Wierichs, Christian Gogolin, and Michael Kastoryano. "Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer". In: *Physical Review Research* 2 (2020), p. 043246. DOI: `10.1103/PhysRevResearch.2.043246`.

[193] Aidan Pellow-Jarman, Ilya Sinayskiy, Anban Pillay, and Francesco Petruccione. "A comparison of various classical optimizers for a variational quantum linear solver". In: *Quantum Information Processing* 20 (2021), p. 202. DOI: `10.1007/s11128-021-03140-x`.

[194] R. Fletcher. *Practical Methods of Optimization.* Wiley, 2000. ISBN: 9780471915478. DOI: `10.1002/9781118723203`.

[195] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. "Quantum Natural Gradient". In: *Quantum* 4 (2020), p. 269. DOI: `10.22331/q-2020-05-25-269`.

[196] Aric Hagberg, Pieter Swart, and Daniel S Chult. "Exploring network structure, dynamics, and function using networkx". In: *Proceedings of the 7th Python in Science Conference* (2008). URL: `https://www.osti.gov/biblio/960616`.

[197] Andrew Arrasmith, Lukasz Cincio, Rolando D. Somma, and Patrick J. Coles. "Operator Sampling for Shot-frugal Optimization in Variational Algorithms". In: *arXiv preprint* (2020). arXiv: `2004.06252`.

[198] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". In: *Nature Physics* 16 (2020), pp. 1050–1057. DOI: `10.1038/s41567-020-0932-7`.

[199] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. "Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows". In: *arXiv preprint* (2020). arXiv: `2006.15788`.

[200] Kristan Temme, Ewout van den Berg, Abhinav Kandala, and Jay Gambetta. "With fault tolerance the ultimate goal, error mitigation is the path that gets quantum computing to usefulness". In: *IBMQ Blog* (2022). URL: `https://www.ibm.com/quantum/blog/gammabar-for-quantum-advantage`.

[201] Miroslav Urbanek, Benjamin Nachman, Vincent R. Pascuzzi, Andre He, Christian W. Bauer, and Wibe A. de Jong. "Mitigating Depolarizing Noise on Quantum Computers with Noise-Estimation Circuits". In: *Physical Review Letters* 127 (2021), p. 270502. DOI: `10.1103/PhysRevLett.127.270502`.

[202] Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout van den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, and Abhinav Kandala. "Evidence for the utility of quantum computing before fault tolerance". In: *Nature* 618 (2023), pp. 500–505. DOI: `10.1038/s41586-023-06096-3`.

[203] Bradley Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1994. ISBN: 9780429246593. DOI: `10.1201/9780429246593`.

[204] Pauli Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: `10.1038/s41592-019-0686-2`.

[205] Ralph D'Agostino and E. S. Pearson. "Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{\{b_1\}}$". In: *Biometrika* 60 (1973), p. 613. DOI: `10.2307/2335012`.

[206] A. C. Aitken. "IV.—On Least Squares and Linear Combination of Observations". In: *Proceedings of the Royal Society of Edinburgh* 55 (1936), pp. 42–48. DOI: `10.1017/S0370164600014346`.

[207] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C. Mckay, and Jay M. Gambetta. "Mitigating measurement errors in multiqubit experiments". In: *Physical Review A* 103 (2021), p. 042605. DOI: `10.1103/PhysRevA.103.042605`.

[208] Paul D. Nation, Hwajung Kang, Neereja Sundaresan, and Jay M. Gambetta. "Scalable Mitigation of Measurement Errors on Quantum Computers". In: *PRX Quantum* 2 (2021), p. 040326. DOI: `10.1103/PRXQuantum.2.040326`.

[209] Ewout van den Berg, Zlatko K. Minev, and Kristan Temme. "Model-free readout-error mitigation for quantum expectation values". In: *Physical Review A* 105 (2022), p. 032620. DOI: `10.1103/PhysRevA.105.032620`.

[210] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, 2003. ISBN: 978-0-89871-534-7. DOI: `10.1137/1.9780898718003`.

[211] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien. "Low-cost error mitigation by symmetry verification". In: *Physical Review A* 98 (2018), p. 062339. DOI: `10.1103/PhysRevA.98.062339`.

[212] Sam McArdle, Xiao Yuan, and Simon Benjamin. "Error-Mitigated Digital Quantum Simulation". In: *Physical Review Letters* 122 (2019), p. 180501. DOI: `10.1103/PhysRevLett.122.180501`.

[213] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O'Brien, and L. DiCarlo. "Experimental error mitigation via symmetry verification in a variational quantum eigensolver". In: *Physical Review A* 100 (2019), p. 010302. DOI: `10.1103/PhysRevA.100.010302`.

[214] Zhenyu Cai. "Quantum Error Mitigation using Symmetry Expansion". In: *Quantum* 5 (2021), p. 548. DOI: `10.22331/q-2021-09-21-548`.

[215] E. L. Hahn. “Spin Echoes”. In: *Physical Review* 80 (1950), pp. 580–594. DOI: `10.1103/PhysRev.80.580`.

[216] Lorenza Viola and Seth Lloyd. “Dynamical suppression of decoherence in two-state quantum systems”. In: *Physical Review A* 58 (1998), pp. 2733–2744. DOI: `10.1103/PhysRevA.58.2733`.

[217] P. Facchi, S. Tasaki, S. Pascazio, H. Nakazato, A. Tokuse, and D. A. Lidar. “Control of decoherence: Analysis and comparison of three different strategies”. In: *Physical Review A* 71 (2005), p. 022302. DOI: `10.1103/PhysRevA.71.022302`.

[218] Götz S. Uhrig. “Keeping a Quantum Bit Alive by Optimized π-pulse Sequences”. In: *Physical Review Letters* 98 (2007), p. 100504. DOI: `10.1103/PhysRevLett.98.100504`.

[219] Joel J. Wallman and Joseph Emerson. “Noise tailoring for scalable quantum computation via randomized compiling”. In: *Physical Review A* 94 (2016), p. 052325. DOI: `10.1103/PhysRevA.94.052325`.

[220] Tomochika Kurita, Hammam Qassim, Masatoshi Ishii, Hirotaka Oshima, Shintaro Sato, and Joseph Emerson. “Synergetic quantum error mitigation by randomized compiling and zero-noise extrapolation for the variational quantum eigensolver”. In: *Quantum* 7 (2023), p. 1184. DOI: `10.22331/q-2023-11-20-1184`.

[221] Ying Li and Simon C. Benjamin. “Efficient Variational Quantum Simulator Incorporating Active Error Minimization”. In: *Physical Review X* 7 (2017), p. 021050. DOI: `10.1103/PhysRevX.7.021050`.

[222] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. “Error Mitigation for Short-Depth Quantum Circuits”. In: *Physical Review Letters* 119 (2017), p. 180509. DOI: `10.1103/PhysRevLett.119.180509`.

[223] Suguru Endo, Simon C. Benjamin, and Ying Li. “Practical Quantum Error Mitigation for Near-Future Applications”. In: *Physical Review X* 8 (2018), p. 031027. DOI: `10.1103/PhysRevX.8.031027`.

[224] Tudor Giurgica-Tiron, Yousef Hindy, Ryan LaRose, Andrea Mari, and William J. Zeng. “Digital zero noise extrapolation for quantum error mitigation”. In: *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2020, pp. 306–316. ISBN: 978-1-7281-8969-7. DOI: `10.1109/QCE49297.2020.00045`.

[225] Andre He, Benjamin Nachman, Wibe A. de Jong, and Christian W. Bauer. “Zero-noise extrapolation for quantum-gate error mitigation with identity insertions”. In: *Physical Review A* 102 (2020), p. 012426. DOI: `10.1103/PhysRevA.102.012426`.

[226] Andrea Mari, Nathan Shammah, and William J. Zeng. “Extending quantum probabilistic error cancellation by noise scaling”. In: *Physical Review A* 104 (2021), p. 052607. DOI: `10.1103/PhysRevA.104.052607`.

[227] Oliver G. Maupin, Ashlyn D. Burch, Brandon Ruzic, Christopher G. Yale, Antonio Russo, Daniel S. Lobser, Melissa C. Revelle, Matthew N. Chow, Susan M. Clark, Andrew J. Landahl, and Peter J. Love. "Error mitigation, optimization, and extrapolation on a trapped-ion testbed". In: *Physical Review A* 110 (2024), p. 032416. DOI: `10.1103/PhysRevA.110.032416`.

[228] Norman R. Draper and Harry Smith. *Applied Regression Analysis*. Vol. 326. Wiley, 1998. ISBN: 9780471170822. DOI: `10.1002/9781118625590`.

[229] William J. Huggins, Sam McArdle, Thomas E. O'Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. "Virtual Distillation for Quantum Error Mitigation". In: *Physical Review X* 11 (2021), p. 041036. DOI: `10.1103/PhysRevX.11.041036`.

[230] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. "Quantum error mitigation". In: *Reviews of Modern Physics* 95 (2023), p. 045005. DOI: `10.1103/RevModPhys.95.045005`.

[231] Mingxia Huo and Ying Li. "Dual-state purification for practical quantum error mitigation". In: *Physical Review A* 105 (2022), p. 022427. DOI: `10.1103/PhysRevA.105.022427`.

[232] Thomas E. O'Brien, Stefano Polla, Nicholas C. Rubin, William J. Huggins, Sam McArdle, Sergio Boixo, Jarrod R. McClean, and Ryan Babbush. "Error Mitigation via Verified Phase Estimation". In: *PRX Quantum* 2 (2021), p. 020317. DOI: `10.1103/PRXQuantum.2.020317`.

[233] Zhenyu Cai. "Resource-efficient Purification-based Quantum Error Mitigation". In: *arXiv preprint* (2021). arXiv: `2107.07279`.

[234] Minh C. Tran, Andrew Y. Guo, Yuan Su, James R. Garrison, Zachary Eldredge, Michael Foss-Feig, Andrew M. Childs, and Alexey V. Gorshkov. "Locality and Digital Quantum Simulation of Power-Law Interactions". In: *Physical Review X* 9 (2019), p. 031006. DOI: `10.1103/PhysRevX.9.031006`.

[235] Lorenzo Leone, Salvatore F.E. Oliviero, Lukasz Cincio, and M. Cerezo. "On the practical usefulness of the Hardware Efficient Ansatz". In: *Quantum* 8 (2024), p. 1395. DOI: `10.22331/q-2024-07-03-1395`.

[236] Minh C Tran, Kunal Sharma, and Kristan Temme. "Locality and error mitigation of quantum circuits". In: *arXiv preprint* (2023). arXiv: `2303.06496`.

[237] Xiao Mi, Pedram Roushan, Chris Quintana, Salvatore Mandra, Jeffrey Marshall, Charles Neill, Frank Arute, Kunal Arya, Juan Atalaya, Ryan Babbush, et al. "Information scrambling in quantum circuits". In: *Science* 374.6574 (2021), pp. 1479–1483. DOI: `10.1126/science.abg5029`.

[238] Oriel Kiss, Michele Grossi, and Alessandro Roggero. "Quantum error mitigation for Fourier moment computation". In: *Physical Review D* 111.3 (2025), p. 034504. DOI: `10.1103/PhysRevD.111.034504`.

[239] Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio. "Error mitigation with Clifford quantum-circuit data". In: *Quantum* 5 (2021), p. 592. DOI: `10.22331/q-2021-11-26-592`.

[240] Daniel Gottesman. “Stabilizer Codes and Quantum Error Correction”. In: *arXiv preprint* (1997). arXiv: `quant-ph/9705052`.

[241] Asher Peres and Leslie E. Ballentine. “Quantum Theory: Concepts and Methods”. In: *American Journal of Physics* 63 (1995), pp. 285–286. DOI: `10.1119/1.17946`.

[242] Andrew W. Cross, Lev S. Bishop, Sarah Sheldon, Paul D. Nation, and Jay M. Gambetta. “Validating quantum computers using randomized model circuits”. In: *Physical Review A* 100 (2019), p. 032328. DOI: `10.1103/PhysRevA.100.032328`.

[243] Salonik Resch and Ulya R. Karpuzcu. “Benchmarking Quantum Computers and the Impact of Quantum Noise”. In: *ACM Computing Surveys* 54 (2022), pp. 1–35. DOI: `10.1145/3464420`.

[244] Marko Žnidarič, Olivier Giraud, and Bertrand Georgeot. “Optimal number of controlled-NOT gates to generate a three-qubit state”. In: *Physical Review A* 77 (2008), p. 032320. DOI: `10.1103/PhysRevA.77.032320`.

[245] Nicholas C Rubin, Ryan Babbush, and Jarrod McClean. “Application of fermionic marginal constraints to hybrid quantum algorithms”. In: *New Journal of Physics* 20 (2018), p. 053020. DOI: `10.1088/1367-2630/aab919`.

[246] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. “Fundamental limits of quantum error mitigation”. In: *npj Quantum Information* 8 (2022), p. 114. DOI: `10.1038/s41534-022-00618-z`.

[247] Aaron M. Lee and Nicholas C. Handy. “Dissociation of hydrogen and nitrogen molecules studied using density functional theory”. In: *Journal of the Chemical Society, Faraday Transactions* 89 (1993), p. 3999. DOI: `10.1039/ft9938903999`.

[248] Garnet Kin-Lic Chan, Mihály Kállay, and Jürgen Gauss. “State-of-the-art density matrix renormalization group and coupled cluster theory studies of the nitrogen binding curve”. In: *The Journal of Chemical Physics* 121 (2004), pp. 6110–6116. DOI: `10.1063/1.1783212`.

[249] Björn O. Roos, Peter R. Taylor, and Per E.M. Sigbahn. “A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach”. In: *Chemical Physics* 48 (1980). DOI: `10.1016/0301-0104(80)80045-0`.

[250] Benjamin G. Levine, Andrew S. Durden, Michael P. Esch, Fangchun Liang, and Yinan Shu. “CAS without SCF—Why to use CASCI and where to get the orbitals”. In: *The Journal of Chemical Physics* 154 (2021). DOI: `10.1063/5.0042147`.

[251] Christopher J. Stein and Markus Reiher. “Automated Selection of Active Orbital Spaces”. In: *Journal of Chemical Theory and Computation* 12 (2016), pp. 1760–1771. DOI: `10.1021/acs.jctc.6b00156`.

[252] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. “Progress towards practical quantum variational algorithms”. In: *Physical Review A* 92 (2015), p. 042303. DOI: `10.1103/PhysRevA.92.042303`.

[253] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush. “Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction”. In: *PRX Quantum* 2 (2021), p. 030305. DOI: `10.1103/PRXQuantum.2.030305`.

[254] Joonho Lee and Martin Head-Gordon. “Regularized Orbital-Optimized Second-Order Møller–Plesset Perturbation Theory: A Reliable Fifth-Order-Scaling Electron Correlation Model with Orbital Energy Dependent Regularizers”. In: *Journal of Chemical Theory and Computation* 14 (2018), pp. 5203–5219. DOI: `10.1021/acs.jctc.8b00731`.

[255] Joshua W. Hollett and Pierre-François Loos. “Capturing static and dynamic correlation with ΔNO-MP2 and ΔNO-CCSD”. In: *The Journal of Chemical Physics* 152 (2020). DOI: `10.1063/1.5140669`.

[256] Hugh G. A. Burton. “Hartree–Fock critical nuclear charge in two-electron atoms”. In: *The Journal of Chemical Physics* 154 (2021). DOI: `10.1063/5.0043105`.

[257] Leon Balents. “Spin liquids in frustrated magnets”. In: *Nature* 464 (2010), pp. 199–208. DOI: `10.1038/nature08917`.

[258] Tian-Heng Han, Joel S. Helton, Shaoyan Chu, Daniel G. Nocera, Jose A. Rodriguez-Rivera, Collin Broholm, and Young S. Lee. “Fractionalized excitations in the spin-liquid state of a kagome-lattice antiferromagnet”. In: *Nature* 492 (2012), pp. 406–410. DOI: `10.1038/nature11659`.

[259] Jinsheng Wen, Shun-Li Yu, Shiyan Li, Weiqiang Yu, and Jian-Xin Li. “Experimental identification of quantum spin liquids”. In: *npj Quantum Materials* 4 (2019), p. 12. DOI: `10.1038/s41535-019-0151-6`.

[260] Johnnie Gray. “quimb: A python package for quantum information and many-body calculations”. In: *Journal of Open Source Software* 3 (2018), p. 819. DOI: `10.21105/joss.00819`.

[261] M.R. Hestenes and E. Stiefel. “Methods of conjugate gradients for solving linear systems”. In: *Journal of Research of the National Bureau of Standards* 49 (1952), p. 409. DOI: `10.6028/jres.049.044`.

[262] Skipper Seabold and Josef Perktold. “Statsmodels: Econometric and Statistical Modeling with Python”. In: *9th Python in Science Conference*. 2010, pp. 92–96. DOI: `10.25080/Majora-92bf1922-011`.