

Optimization of quantum algorithms for near-term quantum computers

PROEFSCHRIFT

TER VERKRIJGING VAN
DE GRAAD VAN DOCTOR AAN DE UNIVERSITEIT LEIDEN,
OP GEZAG VAN RECTOR MAGNIFICUS PROF. DR. IR. H. BIJL,
VOLGENS BESLUIT VAN HET COLLEGE VOOR PROMOTIES
TE VERDEDIGEN OP WOENSDAG 2 NOVEMBER 2022
KLOKKE 15.00 UUR

DOOR

Xavier Bonet-Monroig

GEBOREN TE XÀTIVA, VALÈNCIA (SPANJE)
IN 1991

Promotor: Prof. dr. C. W. J. Beenakker
Co-promotor: Dr. T. E. O'Brien
Promotiecommissie: Prof. dr. A. Montanaro (University of Bristol)
Prof. dr. L. Visscher (Vrije Universiteit Amsterdam)
Prof. dr. J. Aarts
Prof. dr. A. Achúcarro
Dr. J. Tura i Brugués

Casimir PhD series, Delft-Leiden 2022-31

ISBN 978-90-8593-541-4

An electronic version of this thesis can be found
at <https://openaccess.leidenuniv.nl>

Cover: Two roses that arise from the same root are drawn using quantum gates. Inside them there are mathematical descriptions of quantum circuits. One of the roses has been carefully cared and has grown strong and with bright colors. It is a metaphorical representation of a quantum algorithm that has been successful, with errors fully mitigated. The other rose has not been carefully treated, it has not been able to grow and its colors are dull. It is a metaphorical description of a failed quantum algorithm because errors have not been properly mitigated. Illustration by Cristina Cejudo Vera.

To my family: Anna, Xavi and Ada, and to Mireia for always being there.

“There is a flower... I think she has tamed me...”

Antoine de Saint-Exupéry

Contents

1	Introduction	1
1.1	Preface	1
1.2	Basics of quantum algorithms	2
1.2.1	Quantum state tomography	3
1.2.2	Variational quantum algorithms	6
1.2.3	Classical optimization of variational quantum algorithms	8
1.2.4	Applications to quantum chemistry and material science	10
1.3	Noisy intermediate-scale quantum computers	12
1.3.1	Noise in quantum hardware	13
1.3.2	Quantum error correction	14
1.3.3	Quantum error mitigation	15
1.4	Outline of this thesis	16
2	Low-cost error mitigation by symmetry verification	19
2.1	Introduction	19
2.2	Symmetry verification	20
2.2.1	Ancilla and in-line symmetry verification	21
2.2.2	Variational quantum eigensolvers	22
2.2.3	Post-selected symmetry verification and S-QSE	26
2.3	Simulation of symmetry verification on the hydrogen molecule	28
2.4	Inserting and rotating symmetries	30
2.5	Extending the symmetry verification of the hydrogen molecule	34
2.6	Conclusion	37
2.A	Appendix: Error mitigation of QSE with anticommuting operators	39
3	Experimental error mitigation via symmetry verification in a variational quantum eigensolver	43
3.1	Introduction	43
3.2	Variational quantum eigensolvers for the Hydrogen molecule	44
3.3	Experimental error mitigation via symmetry verification	46
3.4	Effect of symmetry verification on positivity constraints	53

3.5	Conclusion	54
3.A	Appendix: Tomographic reconstruction and limitations . .	55
3.B	Appendix: Constraining the positivity of reduced density matrices	57
3.C	Appendix: Theoretical modeling of the experiment	57
3.C.1	Numerical simulations	57
3.C.2	Exchange gate	58
4	Nearly-optimal measurement scheduling for partial to- mography of quantum states	61
4.1	Introduction	61
4.2	Background	64
4.3	Near-optimal measurement schemes for local qubit and fermion operators	65
4.4	Measuring anti-commuting linear combinations of local fermionic operators	69
4.5	Conclusion	69
4.A	Appendix: Schemes for partial state tomography of qubit k -RDMs	71
4.B	Appendix: Upper bounds on the size of commuting cliques of Majorana operators.	73
4.C	Appendix: Details of measurement schemes for fermionic systems	75
4.D	Appendix: Reducing operator estimation over symmetries	78
4.E	Appendix: Parallel iteration over pairings	80
4.F	Appendix: Measurement circuitry for fermionic RDMs . .	81
4.G	Appendix: Proof that the maximum size of an anti-commuting clique of Pauli or Majorana operators is $2N + 1$	83
4.H	Appendix: Proof of Thm. 1	84
5	Performance comparison of optimization methods for variational quantum algorithms	87
5.1	Introduction	87
5.2	Background	88
5.3	Three-stage sampling adaptation	90
5.4	Hyperparameter tuning	93
5.5	The sampling noise floor	95
5.6	Conclusion	96
5.A	Appendix: Details on optimization algorithms	99
5.B	Appendix: Numerical experiments	100
5.C	Optimization algorithms hyper-parameters	101

5.D	Appendix: Unitary Coupled-Cluster ansatz based on coupled-cluster amplitudes	104
5.E	Appendix: Variational Hamiltonian ansatz for the Hubbard model	106
6	Quantum simulation of hydrogen molecule on Quantum Inspire	109
6.1	Introduction	109
6.2	Quantum simulation of the Hydrogen molecule	110
6.2.1	Variational quantum eigensolvers	110
6.2.2	Ground-state energy of the Hydrogen molecule via VQE	111
6.2.3	Implementation in Quantum Inspire	111
6.2.4	Results	113
6.3	Conclusion and outlook	115
7	Calculating energy derivatives for quantum chemistry on a quantum computer	119
7.1	Introduction	119
7.2	Background	120
7.3	The quantum chemical Hamiltonian	121
7.4	Energy derivative estimation using eigenstate truncation approximation	121
7.5	Geometry optimization on a superconducting quantum device	124
7.6	Polarizability estimation	128
7.7	Conclusion	129
7.A	Appendix: Classical computation	131
7.B	Appendix: Experimental methods	132
7.C	Appendix: Simulation methods	133
7.D	Appendix: Numerical optimization and approximate Hessian calculations	133
	Bibliography	137
	Samenvatting	161
	Summary	165
	Resum	167
	Curriculum Vitæ	171

List of publications	173
-----------------------------	------------

1 Introduction

1.1 Preface

Quantum computation is a paradigm of computation and information processing in which the laws of quantum mechanics are used to perform calculations. The basic unit of information storage is called a quantum bit or qubit, and it has properties beyond those of its classical counter-part, the bit. The most notable of which is the fact that information on a qubit can be stored in a superposition of the 0 and 1 states,

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle, \alpha_i \in \mathbb{C} \quad (1.1)$$

$$|\alpha_0|^2 + |\alpha_1|^2 = 1. \quad (1.2)$$

The complexity to describe a (pure) N -qubit quantum state grows exponentially with the system size. This is because a complete description of such quantum state has $2^N - 1$ α_i (complex) coefficients or amplitudes

$$|\psi\rangle = \alpha_0|00\dots 0\rangle + \alpha_1|00\dots 1\rangle + \dots + \alpha_{2^N-1}|11\dots 1\rangle. \quad (1.3)$$

A quantum computer is a collection of qubits, and thus a quantum computation is the use of such device to perform a calculation. As a specialized machine, quantum computers require dedicated algorithms that take advantage of quantum mechanical features to solve problems. In the early 90s, Deutsch and Josza [1, 2] showed that a quantum algorithm implemented on a quantum computer could solve a specific classical problem exponentially faster in time compared to the best-known classical algorithm for the same task. In 1997 a ground-breaking development by Peter Shor demonstrated the first quantum algorithm that had an exponential run-time advantage with respect to its classical counterpart for finding the prime factors of an integer [3]. Since then, the interest on quantum computation increased and mostly focused on developing quantum algorithms that show quantum speed-up in solving classically hard problems [5]. In recent years, research in quantum algorithms has gone beyond the search for speed-ups to a wider range of topics such as understanding what makes quantum computers powerful [4], studying quantum states that cannot be

1 Introduction

accessed with a classical computer [6, 7] or designing fine-controlled quantum sensing protocols [8].

In the past two decades ground-breaking developments in experimental quantum physics has opened the door to study the power of quantum computation. Academic and industry laboratories world-wide have been devoted to manufacture prototypical quantum hardware, using a variety of platforms. In the process, researchers have observed that realistic quantum computers are highly sensitive to noise. The large number of imperfect and unwanted operations of current quantum hardware makes it unrealistic to run the originally proposed quantum algorithms. A solution to these problem is to design quantum error correction schemes such that with bigger devices and better control one can approach a noise-free quantum computer. Unfortunately, with the current state-of-the-art quantum hardware, a useful, fault-tolerant quantum computer will require millions of qubits [9, 10]. Until million-qubit devices are achieved we remain in an intermediate era, where the potential for application remains, but where low-cost error mitigation must replace provable error correction techniques.

This thesis studies various aspects of near-term quantum computing; from noise characterization and error mitigation schemes to information extraction of quantum states, as well as classical optimization methods for circuits. The results of this work cover a series of topics that are necessary to demonstrate beyond-classical computational advantage in noisy quantum hardware.

1.2 Basics of quantum algorithms

A quantum algorithm is a set of instructions given to a quantum computer to perform a task. It involves preparing a quantum state $|\psi\rangle$, evolving the state $|\psi\rangle \rightarrow U|\psi\rangle$ via unitary operations ($UU^\dagger = U^\dagger U = I$), and measure observables. A classical post-processing step is generally used to interpret the extracted information and/or to continue with a new algorithm iteration.

Typically, quantum registers are initialize in their 0 state, such that

$$|\psi_{\text{initial}}\rangle = |0\rangle^{\otimes N}. \quad (1.4)$$

This state is evolved applying a sequence of quantum (unitary) gates that transform the quantum state

$$|\psi_{\text{final}}\rangle = U_k U_{k-1} \dots U_0 |0\rangle^{\otimes N}. \quad (1.5)$$

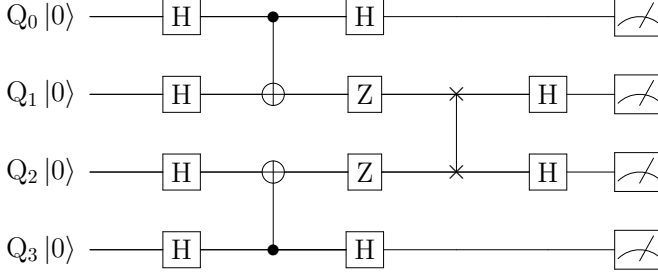


Figure 1.1: An example of a 4-qubit quantum circuit. The initial state $|\psi_0\rangle = |0000\rangle$ is transformed by the use of one- and two-qubit gates to a different state. The boxes at the end of the circuit represent a measurement apparatus, where information about $|\psi_1\rangle$ is extracted.

The information is extracted from the quantum device by measuring measuring one or more qubits. From the 2^N bits of information encoded in the quantum state, only a maximum of N bits of information can be obtained after a measurements operation. This is a huge bottleneck when using a quantum computer.

The system is projected through the Born's rule [11] to the computational state described by these measurements, and the rest of the information is lost. An example of a quantum circuit with single-qubit (Hadamard, Z-gate) and two-qubit (CNOT, SWAP) gates is depicted in Figure 1.1. In this figure the time goes from left to right, sequentially applying unitary operations, the U_k in eq. 1.5, at each time-step. The left side of the figure indicates the initial state $|\psi_0\rangle = |0000\rangle$. After the evolution the quantum state is measured, action depicted by the measurements apparatus at the end of the figure.

1.2.1 Quantum state tomography

In the preface of this thesis we stated that a (pure) N -qubit quantum state is described by $2^N - 1$ complex coefficients. Quantum state tomography is the process of obtaining the full description of a given quantum state $|\psi\rangle$ as described in eq. 1.3. One can describe a pure quantum state by its

1 Introduction

density matrix representation,

$$\rho = |\psi\rangle\langle\psi| = \sum_{i,j=0}^{2^N-1} \beta_{i,j} |i\rangle\langle j|. \quad (1.6)$$

The density matrix formalism of quantum states is strictly more descriptive as one can also represent non-pure (mixed) quantum state,

$$\rho_{\text{mixed}} = \frac{1}{2} [|0\rangle\langle 0| + |1\rangle\langle 1|]. \quad (1.7)$$

One might think that the equal superposition state,

$$|\psi\rangle = \frac{1}{\sqrt{2}} [|0\rangle + |1\rangle], \quad (1.8)$$

also has a probability $\frac{1}{2}$ of returning 0 or 1. By writing its density matrix representation

$$\rho_{\text{superposition}} = |\psi\rangle\langle\psi| = \frac{1}{2} [|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|], \quad (1.9)$$

and comparing eq. 1.7 with respect to eq. 1.9 one can see that the mixed state is not the same as the equal superposition state. The density matrix ρ is a complex $2^N \times 2^N$ matrix in the same Hilbert space of the qubit states [43] with entries $\beta_{i,j} = \alpha_i \cdot \alpha_j^*$. ρ can be expressed as a linear combination of Pauli operators, $\hat{P}_i \in \{I, X, Y, Z\}^{\otimes N}$, as they form a basis of the Hilbert space of the qubits,

$$\rho = \frac{1}{2^N} \sum_{i=0}^{4^N-1} c_i \hat{P}_i. \quad (1.10)$$

In this form, quantum state tomography is reduced to simply estimate the coefficients

$$c_i = \frac{\text{Trace}[\rho \hat{P}_i]}{2^N}. \quad (1.11)$$

On a quantum computer, eq. 1.11 cannot be exactly computed. After every measurement of an N-qubit quantum state ρ only a maximum of N bits of information can be retrieved. This is formalized by the so-called Holevo bound [20], and is one of the greatest bottlenecks when performing quantum computations. One can only obtain an estimate of c_i by using

M copies of ρ and performing a projective measurement on the basis of the operator P_i (see Fig. 1.2). The estimator of c_i

$$\mathbb{E}[c_i] = \langle P_i \rangle, \quad (1.12)$$

with variance

$$\text{Var}[\mathbb{E}[c_i]] = \frac{1 - \langle P_i \rangle^2}{M}. \quad (1.13)$$

The probability of $\mathbb{E}[c_i]$ being $\epsilon > 0$ close to the true value c_i is given by the Chebyshev's inequality

$$\mathbb{P}[|\mathbb{E}[c_i] - c_i| \geq \epsilon] \leq \frac{\text{Var}[c_i]}{M\epsilon^2}. \quad (1.14)$$

This equation provides a bound on the number of copies of ρ needed to estimate a given coefficient to a certain precision

$$M \sim \frac{1}{\epsilon^2}. \quad (1.15)$$

A full characterization of an N-qubit quantum state requires one to estimate each of the 4^N c_i coefficients. With the method described here the time complexity of running quantum state tomography will scale as

$$\text{Time} = \mathcal{O}(\epsilon^2 \cdot 4^N). \quad (1.16)$$

Newer quantum state tomography methods [21] have been able to reduce the time-complexity to

$$\text{Time} = \epsilon \cdot \exp \mathcal{N}. \quad (1.17)$$

thus making quantum state tomography unfeasible even for a small-sized quantum computers.

In practice, we don't need to estimate every expectation value of a quantum state. Given an operator

$$\hat{O} = \sum_i o_i P_i, \quad (1.18)$$

where P_i are Pauli operators, we need only measure those $\langle P_i \rangle$ that have non-zero o_i coefficients. For instance, the electronic structure problem, a central problem in quantum chemistry, only requires one to characterize the state on $\mathcal{O}(N^4)$ operators where N refers to the number of spin-orbitals. This is an exponential reduction in the number of measurements with respect to characterizing the full quantum state [108, 119].

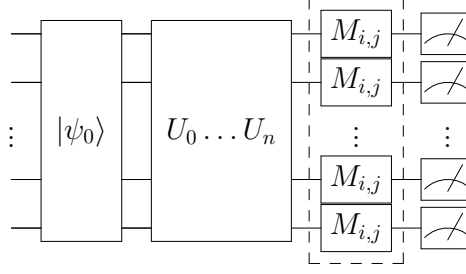


Figure 1.2: Quantum state tomography is done by measuring the state in different basis rotations before extracting the information the qubits, $M_{i,j}$ gate in the figure. The boxed gates in the figure are the tomography rotations that allow to extract the information from the quantum state in different bases. In general, there are 4 possible basis rotations, namely the Pauli matrices I , X , Y and Z . Recover the full quantum state requires to prepare all possible combinations of the Pauli matrices in all possible qubits. However, typically we are only interested in a subset of these $M_{i,j}$ combinations to extract the most important information of the quantum state.

1.2.2 Variational quantum algorithms

Variational quantum algorithms are a family of quantum algorithms in which a set of classical parameters are used to control the quantum state of a device [39–41, 87]. These parameters are included during the evolution step of the algorithm such that some of the quantum gates are parametrized. Therefore, the parametrized quantum state is $|\psi(\vec{\theta})\rangle = U(\vec{\theta})|0\rangle^{\otimes N}$. Figure 1.3 shows an example of a parametrized quantum circuit, where single-qubit rotation gates can be modified based on the classical parameters $\vec{\theta} = \theta_0 \dots \theta_n$.

Variational quantum algorithms require an additional element, namely a real-valued cost function $\mathcal{C}(\vec{\theta})$ to optimize the parameters of the quantum state. \mathcal{C} is typically the expectation value of a quantum observable \hat{O} , a Hermitian operator in the Hilbert space of the qubits. On a quantum computer Hermitian observables can be measured by expressing them as a sum of Pauli operators, e.g. $\hat{P}_i \in \{I, X, Y, Z\}^{\otimes N}$,

$$\hat{O} = \sum_i c_i \hat{P}_i \rightarrow \mathcal{C}(\vec{\theta}) = \langle \hat{O} \rangle = \sum_i c_i \langle \hat{P}_i \rangle, \quad (1.19)$$

with $c_i \in \mathbb{R}$. In Sec. 1.2.1 we described how to characterize a N-qubit quantum state through its Pauli decomposition (eq. 1.10). We can use

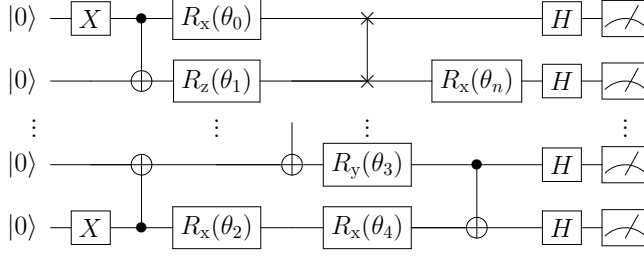


Figure 1.3: A parametrized quantum circuit is formed by combining fixed and variable gates. In this example we show fixed gates Hadamard (H), X, CNOT or SWAP gates and variable rotation gates around specified axes (X, Y or Z). Although here every parametrized gate has its own parameter, it is possible that multiple gates share the same parameter.

(partial) quantum state tomography can be used to estimate the cost function \mathcal{C} of a variational quantum algorithm. However, the number of operators to estimate now need not be exponential in the system size. In fact, most of the problems of interest only need $\text{poly}(N)$ operators. While the Pauli decomposition is a widely use strategy to measure any quantum observable \hat{O} on a quantum computer there exist other methods to estimate \hat{O} such as low-rank factorization [132] or classical shadows of quantum states [65–68].

An evaluation of the cost function involves preparing and measuring the quantum state multiple times to build statistics. Therefore, we only have access to a sampled cost function that depends on the parameters $\vec{\theta}$ and the number of measurements M

$$\bar{\mathcal{C}}(\vec{\theta}, M) = \sum_i c_i [\langle \hat{P}_i \rangle + \epsilon_i(M)]. \quad (1.20)$$

The sampling noise manifests in the form of a random variable ϵ_i drawn from a binomial distribution with variance $\sigma_i^2 \sim \frac{1}{M}$ (see eq. 1.13).

In a variational quantum algorithm one attempts at finding the set of parameters $\vec{\theta}$ of a fixed parametrized quantum circuit that minimize $\mathcal{C}(\vec{\theta})$. Thus, a classical optimization algorithm is used to update $\vec{\theta}$ after every evaluation of \mathcal{C} . The optimization loop continues until some convergence or hard-stop criteria have been met. At the end of the algorithm we are left with a set of parameters $\vec{\theta}_{\text{opt}}$ that minimize $\bar{\mathcal{C}}$. The state generated by the parametrized quantum circuit with $\vec{\theta}_{\text{opt}}$ is an approximation to the

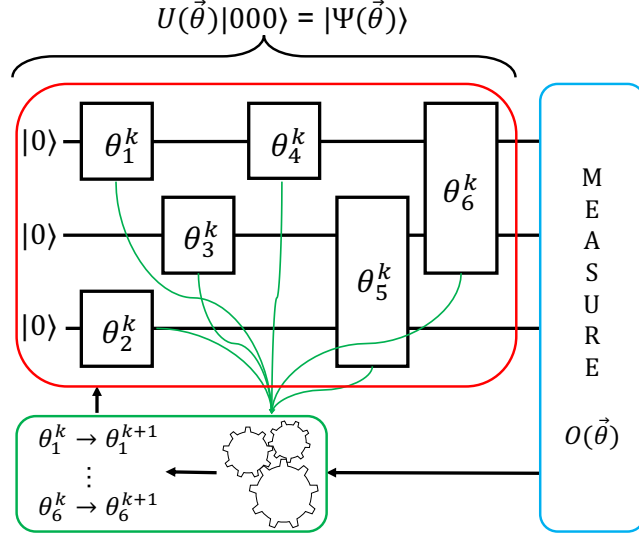


Figure 1.4: A schematic representation of a variational quantum algorithm. First the quantum device is initialized in a known quantum state $|000\rangle$, followed by a sequence of parametrized quantum operations that prepare a final quantum state $|\Psi(\vec{\theta})\rangle$ (red box). The quantum state is then measured to obtain the observable $\hat{O}(\vec{\theta})$ (blue box), which is fed into a classical optimization loop. A new set of angles $\vec{\theta}^{k+1}$ is given by the optimizer (green box). The new parameters are then used in the red box step, and the process is repeated until some convergence criteria is met. Figure taken with permission from ref. [19]

ground-state of the observable used as a cost function \hat{O} . An schematic representation of a variational quantum algorithm is depicted in fig. 1.4.

1.2.3 Classical optimization of variational quantum algorithms

Quantum computers promise to solve problems that are out-of-reach for the largest and fastest existing supercomputers. However such quantum advantage does not imply that quantum computers will replace their classical counter-parts. In fact, classical computers are a crucial part of a quantum computer architecture for its control [30]. Similarly, in variational quantum algorithms a classical optimization method is used to find the optimal values of the parametrized quantum circuit with respect to

the cost function (Sec. 1.2.2). An optimization algorithm is a piece of software that attempts to find the parameters \vec{x} that minimize (or maximize) some cost function $\mathcal{F}(\vec{x})$. To find such optimal parameters, an optimizer suggests candidates $\vec{x}_{1\dots n}$ to be evaluated on the cost function $\mathcal{F}(\vec{x}_{1\dots n})$ sequentially. From those evaluations the optimizer updates a set of rules such that a new candidate with a lower (or higher) $\mathcal{F}(\vec{x}_{n+1})$ is found. The search for candidate parameters continues until some convergence criterion is met (e.g. no improvement in $\mathcal{F}(\vec{x})$, changes in parameters smaller than some threshold, etc.) or some hard stop criterion is reached (e.g. total number of function evaluations, maximum time allowed, etc.). It is clear that a successful implementation of a variational quantum algorithm strongly depends on the ability of the classical computer to optimize the given cost function.

Quantum devices need the assistance of a classical optimization method for their characterization and calibration [26–30]. However, the question of how optimal these methods for quantum-optimal control tasks is uncertain and currently a very active research field. Similarly, the optimization of variational quantum algorithms has largely been with off-the-shelf optimizers. Typically optimizers assume that the cost function is deterministic, but in variational quantum algorithms we only have access to a probabilistic output of it, $\bar{\mathcal{C}}$. For this reason, the results in the literature do not show a clear trends, thus making difficult to assess the optimality of commonly used optimizers for variational quantum algorithms.

For decades, computer scientist have been developing strategies to systematically compare and benchmark optimization algorithms for classical cost functions [22–25]. From the field of classical optimization algorithms we have learnt that there is not a universal optimizer: every problem is better optimized with a particular method and for every instance there are optimal hyper-parameters of the optimizer. This reflects the heuristic nature of the optimization algorithms. Progress in variational quantum algorithms requires us to extend these systematic analysis to the newly accessible cost functions through optimizer benchmark and landscape analysis. Recently, some work has been put into the design of optimizers specifically for variational quantum algorithm specific optimizers [138–141] but many questions remain unknown about their capabilities to optimize larger problems.

1.2.4 Applications to quantum chemistry and material science

Quantum computers have been marked as a tool to study problems in the domain of quantum chemistry and material science. Typically, we are interested in calculating the spectrum and/or some properties (e.g. energy-gradients, polarization, magnetization) of the physical systems described by a Hamiltonian. One of such problems in quantum chemistry is the electronic structure problem, which describes the behaviour of the electrons around the nuclei, under the Born-Oppenheimer approximation which assumes that the nuclei are fixed particles. The Hamiltonian that describes this system is

$$\hat{H} = \sum_{pq} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s, \quad (1.21)$$

where \hat{a} and \hat{a}^\dagger are the fermion creation and annihilation operators acting on the fermionic modes p, q, r, s . The coefficients h_{pq} and h_{pqrs} are the set of one- and two-electron integrals evaluated classically using a fixed basis set functions [31]. The Fermi-Hubbard model is another example of a system of interest for material science [32, 33],

$$H_{\text{Hubbard}} = H_t + H_U = -t \sum_{(i,j),\sigma} \left(a_{i\sigma}^\dagger a_{j\sigma} + a_{j\sigma}^\dagger a_{i\sigma} \right) + U \sum_i n_{i\uparrow} n_{i\downarrow}. \quad (1.22)$$

It describes the interaction between particles on a lattice of $n_x \times n_y$ sites. Particles can tunnel to the nearest-neighbor site with hopping strength t but they are prevented from tunneling to an occupied site of the same spin σ with on-site potential strength U .

In quantum chemistry and material science most of the relevant properties of the systems are described by the eigenvalues and eigenstates of the Hamiltonian (e.g. the ground- and low-lying energy states of the system). Therefore, the goal is to accurately compute these energies. Within the variational paradigm, this can be achieved by preparing approximate ground states via variational circuits (Sec. 1.2.2), and extracting energies or other expectation values using the techniques described in Sec. 1.2.1.

Particles in nature are observed to be of two types: fermions or bosons. Bosonic systems have a symmetric wave-function

$$\Psi_{\text{boson}}(\vec{x}_2, \vec{x}_1) = \Psi_{\text{boson}}(\vec{x}_1, \vec{x}_2), \quad (1.23)$$

and thus the exchange of two bosons leaves the system invariant. The symmetric nature of bosons implies that multiple particles can occupy

the same space. In contrast, fermionic systems have an anti-symmetric wave-function

$$\Psi_{\text{fermion}}(\vec{x}_2, \vec{x}_1) = (-1)\Psi_{\text{fermion}}(\vec{x}_1, \vec{x}_2), \quad (1.24)$$

and an exchange of two fermions changes the sign of the wave-function [12]. Typically, many-body quantum systems are studied through the *creation* (a^\dagger) and *annihilation* (a) or second-quantization formalism. The action of these operators on the vacuum state $|0\rangle$ is as follows:

$$a_i|\vec{0}\rangle = 0 \quad (1.25)$$

$$a_i^\dagger|\vec{0}\rangle = |00100\rangle, \quad (1.26)$$

where a particle at the i -th position has been created. The creation and annihilation operators of a bosonic system obeys the commutation relation

$$[a_i, a_j^\dagger] = a_i a_j^\dagger - a_j^\dagger a_i = \delta_{ij}. \quad (1.27)$$

However a fermionic system follows the anti-commutation relation

$$\{a_i, a_j^\dagger\} = a_i a_j^\dagger + a_j^\dagger a_i = \delta_{ij}. \quad (1.28)$$

The anti-symmetric nature of interacting fermionic systems can pose a problem when studying them via e.g. classical Monte-Carlo methods. The so-called fermionic sign problem appears because one must track all fermionic exchanges to account for the correct sign in the wave-function. This sign problem has been proven NP-hard [13], which is strong evidence that calculating ground states of interacting fermionic systems scales exponentially in the system size. On the contrary a qubit can store the information of a fermionic site in a one-to-one correspondence. Polynomial-time quantum algorithms under fair assumptions to solve systems of interacting fermions have been discovered [14–17, 31, 76, 78, 79, 95, 187]. This translates into an exponential speed-up in computing ground states of interacting fermions using a quantum computers with respect to classical methods.

Variational quantum algorithms are suited to approximate the energy spectrum of such fermionic interacting problems in the near-term [18, 19]. However, a fermionic representation is non-native on a quantum computer; the native operations are in terms of qubit (Pauli) operators. Luckily there exist methods to transform fermionic operators onto a polynomial number of Pauli operators. The so-called fermion-to-qubit mappings [34–38] ensure that the anti-commutation relations of the fermionic operators is respected by the Pauli operators.

In the past decade we have seen a large number of experimental demonstration of variational quantum algorithms for quantum chemistry and material science problems. Variational quantum eigensolvers were originally proposed for the task of approximating the ground-state of the Hydrogen molecule in a photonic quantum device [39, 40]. Many more experiments in a wide variety of platforms have followed [41, 42, 44, 45]. The largest variational quantum algorithm for chemistry was run in 2021 with 12 qubits to compute the Hartree-Fock energy of several molecules [46].

1.3 Noisy intermediate-scale quantum computers

In the roadmap towards error-free quantum computers based on quantum error correction schemes there will exist error prone intermediate-scale hardware sufficiently large to challenge the limits of classical computers. This idea led to the definition of noisy intermediate-scale quantum (NISQ) computation by Preskill [47] to explore the capabilities of quantum hardware pre-quantum error correction. As a newly introduced form of quantum computation, NISQ era needs tailored quantum algorithms that take into account the noise limitations of the hardware. A whole new research field has opened to understand, improve and develop quantum hardware and algorithms for noisy intermediate-scale quantum computers [48].

Quantum computation with NISQ devices requires tailored software and hardware to minimize the noise effects. The first necessary step is to understand as accurately as possible the limitations of the hardware. Only then it is possible to explore algorithms and problems that are suited to them. An alternative solution is to construct hardware which is specifically design for a task or problem, namely building analog quantum computers. In this way, the capabilities of the hardware are maximized (for the specific problem) but their usability is largely constrained.

Regardless of the approach, it is likely that NISQ computing will be best suited for highly specialized problems in the field of quantum mechanics. The simplest approach is to use the quantum hardware to prepare a quantum state $|\psi\rangle$, and extract relevant information using quantum state tomography described in Sec. 1.2.1. Ideally, the quantum state is not accessible with a classical computer, and thus we can obtain information which is otherwise impossible to get.

1.3.1 Noise in quantum hardware

A quantum computation involves storing and transforming quantum information on a quantum system. Unfortunately quantum information is fragile and extremely sensitive to interactions between the quantum system and its surroundings. Some level of noise in any quantum processing unit will be unavoidable because we need to interact with it to perform a computation. Therefore, it is important to study which errors occur during the execution and which physical process might have caused such error. By understanding these effects we can construct noise models that reproduce the errors of a quantum device accurately and reliably, allowing us to benchmark the quality of the quantum hardware.

The simplest noise models that reliably describe errors on a quantum processing unit are the so-called bit- and phase-flip errors. A bit-flip error is the process by which the state of a single qubit $|\psi\rangle$ changes from $|\psi\rangle = |0\rangle \rightarrow |1\rangle$ (or vice-versa) with probability p_{bit} and remains unchanged with probability $1 - p_{\text{bit}}$. Similarly, a phase-flip error describes the change in the relative phase of the quantum state $|\psi\rangle = |0\rangle + |1\rangle \rightarrow |\psi\rangle = |0\rangle - |1\rangle$ of a single qubit with probability p_{phase} . The bit- and phase-flip error models can be combined to a single noise model such that with probability p_{bp} a single qubit undergoes a bit and phase flip error, and with $1 - p_{\text{bp}}$ the qubit remains unchanged. Although these noise models describe the errors at a single-qubit level one can use them to approximate the errors of a multi-qubit device by extending the error model to all qubits of the system with the same probability. It is possible to make the bit-phase flip model more accurate by assigning different probabilities of error to each individual qubit. Despite its simplicity the bit- and phase-flip errors are a good approximation to the errors occurring to a quantum device. These noise models are particularly relevant for quantum error correction [60, 85], where repeated measurement of a system projects coherent sources of error to discrete events.

Another important source of error in quantum hardware is the depolarizing noise channel. It describes the process by which a quantum state becomes a maximally mixed state with probability p_{dep} . This process can be seen as the quantum state losing its “quantumness” because the maximally mixed state is a classical state with equal probability of measuring 0 or 1 in a single qubit in the computational basis 1.7.

In reality, noise in a quantum device is not discrete, and its quantum state is continuously modified in time by the action of the environment. Noise models that account for the time-dependent nature of these interactions are needed to describe the errors that occur during the execution

1 Introduction

of a quantum algorithm in a noisy device more accurately. An example of such time-dependent interaction is the loss of energy of the quantum device to the environment also known as amplitude damping. The rate at which the energy (or amplitude) of the quantum state is dissipated to the surroundings is given by T_1 ; after time t , the probability that a system prepared in the excited state has decayed to the ground state is $1 - e^{-\frac{t}{T_1}}$. Another source of error in a quantum processing unit is the loss of information without energy dissipation. This is the phase-flip channel, and it is characterized by the pure dephasing rate T_ϕ ; the rate at which the phase decays for a given time t . It is common to use the total loss of phase information rate T_2 instead of the pure dephasing rate. T_2 is the rate at which the phase of a quantum state is damped. It defines the time-scale at which the probability of the phase damping happening is given by $p_{\text{phase}}(t) = 1 - e^{-\frac{t}{T_2}}$. It can be shown that the coherence times T_1 and T_2 are related such that [43]

$$T_1 \geq 2T_2. \quad (1.29)$$

The T_1 and T_2 coherence times are relevant figures to characterize a quantum processing unit. First, they set an upper bound on the number of operations that can be performed on the device before a quantum computation breaks down due to errors. Secondly, they are a standard metric to compare across different qubit platforms. Beyond the errors models introduced in this section, many more error sources are known to affect quantum hardware (see ref. [30] for an overview of noise sources in superconducting qubits): leakage to higher-energy states, idle cross-talk, pulse cross-talk, measurement cross-talk.

Depending on the level of accuracy desired to assess the quality of a quantum program on a specific device we can build highly accurate noise models by running characterization experiments. Such experiments provide researchers with information about coherent over-rotations, thermal fluctuations or leakage to spurious states. All of them can be combined in a device-specific noise model that can reproduce the experimental data to high accuracy. In this thesis we develop one of such errors models in a two-qubit superconducting quantum device, modelling all known sources of noise.

1.3.2 Quantum error correction

Since early in the development of quantum computation it was known that quantum states are fragile, and that a successful quantum algorithm will

require the quantum states to be stable for long periods of time. Quantum error correction was developed to overcome such a challenge. Building upon knowledge from classical error correction where additional bits of information are used to store redundant information used to detect errors, quantum error correction codes use additional qubits with the same goal. The first quantum error correction scheme was proposed independently by Shor [57] and Steane [58]. Subsequently, the stabilizer formalism was developed by Gottesman [59, 60] to investigate properties of quantum error correcting codes.

Outstanding progress has been done in the domain of quantum error correction in the past decade. In particular, several experimental realizations of the surface code [61] have been done with superconducting quantum processors [62, 63] and the largest surface code experiment to date [64]. Quantum error correction is a very active research field, constantly making progress towards efficient implementations of quantum error correcting codes for fault-tolerant universal quantum computers. Unfortunately at the time of this thesis the theoretical predictions on the error rates and/or number of qubits required to implement fault-tolerant quantum algorithms are orders of magnitude away from existing state-of-the-art quantum devices. Therefore, it is expected that fully-fledged quantum error corrected machines will be available in the upcoming decades.

1.3.3 Quantum error mitigation

Quantum error mitigation defines a family of techniques that aim at reducing the error on a noisy quantum computation by detecting errors by adding qubits and/or measurements without the ability to correct them. In order to develop a successful quantum error mitigation method one must consider the capabilities of the hardware such that the cost of running the error mitigation does not worsen the unmitigated quantum computation. Recently much progress has been done in developing and improving the performance of noisy quantum algorithm with low-cost overhead error mitigation strategies. Despite the effort in reducing the computational overhead of quantum error mitigation strategies they all require some steps that will take large polynomial or even exponential computational resources with the system size. Therefore, these strategies are thought to be used in computational tasks involving a moderate number of qubits where the cost will still be within the capabilities of the hardware.

A broad range of quantum error mitigation strategies have been proposed following developments of variational quantum algorithm for near-

1 Introduction

term quantum hardware. Quantum error mitigation protocols can be split into two groups depending on which errors they target. The first class involves reducing errors at the hardware-level, for example:

- *Zero-noise (Richardson) extrapolation* [49, 50]: a zero-noise extrapolated observable is found by artificially increasing the noise level of the gates and calculating the value of the observable. Then, its zeroth order value is computed by fitting a polynomial function to the measured points.
- *Quasi-probabilistic gate decomposition* [50, 51]: gate-level noise is averaged-out by randomly adding noisy gates to the target circuit such that some of the errors are cancelled. The observables are measured from different random circuits, thus finding an average value of all of them.

Alternatively one can remove errors at the problem-level by exploiting previous knowledge of the target system. Two of the main quantum error mitigation of this type are

- *Symmetry verification* [52–54]: the quantum state at the end of a noisy quantum circuit is checked to verify if it respects the symmetries of the target problem. The state is then projected to the correct sub-space that respects these symmetries.
- *Virtual state distillation* [55, 56]: M copies of the noisy quantum state are used to approximate the pure state by distillation of the leading eigenvalues of the noisy copies. As M increases the quantum state approaches the pure state exponentially fast.

In the upcoming years it is expected that the quantity and quality of the qubits on quantum hardware will be significantly larger. At that point we hope to achieve quantum advantage in specific problems by carefully designing algorithms that reduce the noise on these devices together with powerful quantum error mitigation protocols.

1.4 Outline of this thesis

In the rest of the introduction we provide an overview of this thesis where a brief description of every chapter is presented.

Chapter 2

In this chapter of this thesis we cover the topic of quantum error mitigation to improve the results of faulty quantum computations. We developed a quantum error mitigation strategy that uses the inherent symmetries of a target problem as a way to detect errors during a variational quantum algorithm run. To perform the symmetry-verification step we design shallow quantum circuits that signal if an error occurred by measuring one or more qubits. However such a process incurs on a cost in terms of additional gates and qubits, and might lead to less error mitigation power. To overcome the additional cost we develop a cost-free approach in which symmetry-verification can be done as a post-processing step, and only requires a polynomial number of additional observables.

Chapter 3

This chapter covers the work done in collaboration with an experimental laboratory to demonstrate the power of symmetry-verification on a two superconducting qubit experiment. First we perform an in-depth study of the noise sources affecting the experiment and provide a theoretical prediction of the experiment. Then we study the power of symmetry-verification for the experiment, with a theoretical prediction of the experiment after symmetry-verification. Finally, we compare our predictions with the experimental data finding an outstanding agreement between them.

Chapter 4

The next part of the thesis continues in the domain of near-term quantum algorithms. We focus on how to extract the necessary relevant information of a quantum state without reaching exponentially large times for the task. First we find analytical lower-bounds on the number of independent measurement circuits required to extract elements of k -reduced density matrices in both spin- $\frac{1}{2}$ and fermions. Then, we find a binary partition scheme strategy to design the measurement circuits that is asymptotically optimal for spin- $\frac{1}{2}$ up to logarithmic factors, and matches the best-ever strategy. A similar method is then used to devise measurement scheduling of fermionic 1-, and 2-RDMs, finding an exact solution for the former and an asymptotically optimal solutions for the latter (up to a constant factor).

Chapter 5

In this chapter we focus our attention on the third part of any computation using noisy intermediate-scale quantum hardware, the optimization of variational quantum algorithms. The goal of this chapter is to assess the limitation of existing classical optimization algorithms for the task of optimization a variational quantum algorithm under noise. First, we devise a set of numerical experiments that allow us to compare between several off-the-shelf optimizers across multiple target problems, showing that only two of these methods are able to find accurate results under sampling noise conditions. These two methods are then compared with their best hyper-parameters, finding a comparable performance among them. Our final result is the formal definition of the sampling noise floor: any value within a region defined by the sampling level can be measured as the best-ever result. The sampling noise floor is an artifact that can lead to erroneous parameters during an optimization, greatly affecting the final result of the computation.

Chapter 6

In this chapter we describe a user case test on the Dutch quantum computing facilities Quantum Inspire. The goal of this research is to assess the limitations of the current infrastructure to perform NISQ computations. Our results and conclusion are currently being used to develop the new generation of Quantum Inspire.

Chapter 7

In the last chapter of this thesis we develop quantum algorithms to calculate energy derivatives for quantum chemistry. The theory of this work covers a large amount of resource estimates for several quantum algorithms in both noisy intermediate-scale quantum and fault-tolerance regime. A small test example of how to calculate gradients using Newton's method is shown using simulated and experimental data with an outstanding agreement.

2 Low-cost error mitigation by symmetry verification

2.1 Introduction

Noisy, intermediate scale quantum (NISQ) devices have begun to appear in laboratories around the world. These devices have performance rates around or just below the quantum error correction threshold [69–73], but are lacking the number of qubits required for full fault-tolerant quantum computing. This raises the open question of whether the upcoming generation of quantum computers will provide a quantum advantage over classical computers, and in which fields this might be achieved [47, 74, 75]. In particular, for the area of digital quantum simulation, it has been suggested that variational quantum eigensolvers [39] may be sufficiently low-cost to be performed on ~ 50 qubits [76–79]. Around this point, solving the many-body problem exactly becomes too challenging for classical computers, and a slight quantum edge might be available above current approximations.

In lieu of full error correction techniques, much attention is being turned to error mitigation techniques, which, although non-scalable, promise modest improvements at low cost. Previous work has focused on active error minimization, whereby data is obtained at artificially increased error rates and then extrapolated to zero [49, 50, 80, 81], and on probabilistic error cancellation, where an ensemble of noisy circuits is applied such that they average to the target error-free circuit [50, 51]. More specific techniques have been developed for quantum simulation, and in particular for variational quantum eigensolvers. A technique developed for exploring the low-energy excited subspace of a quantum system, the quantum subspace expansion, has been shown to have error mitigation as a side-effect [82, 83].

In this chapter we investigate error mitigation via verification of symmetries found in quantum circuits, in particular those in physical systems. This is a low-cost version of the stabilizer parity checks ubiquitous in quantum error correction [84, 85]. We develop multiple protocols to perform

symmetry verification, both repeatedly throughout a quantum circuit and as a single post-processing step. The latter can be related to a variant of the quantum subspace expansion [82]. We study the sensitivity of symmetry verification to different noise channels, and demonstrate how it can be optimized by adding new symmetries and rotating existing symmetries to be more sensitive to local noise.

2.2 Symmetry verification

Our study is motivated by the presence of symmetries in quantum mechanical systems. In such systems, one has a Hamiltonian \hat{H} , and is usually interested in studying the properties of ground or low-lying eigenstates of the system. A (unitary) symmetry of a system is a unitary operator \hat{S} that commutes with the Hamiltonian - $[\hat{H}, \hat{S}] = 0$. When this is true, \hat{H} may be block diagonalized within the eigenspaces of \hat{S} . Then, if one were to study eigenstates of \hat{H} on a quantum computer, one may perform such a study entirely within a single target eigenspace \mathcal{S} of \hat{S} . In real-world quantum computers, noise may shift the state of the computer outside of the target eigenspace \mathcal{S} . By verifying during or at the end of a calculation that the system remains in \mathcal{S} , and throwing away results where this is not the case, it is thus possible to make our quantum computation less sensitive to these types of noise.

Verification of a symmetry is performed by measurement and post-selection which is typically performed in the computational basis (the eigenstates $|0\rangle$ and $|1\rangle$ of a single qubit). The Pauli operators \mathbb{P}^N may be rotated into this basis relatively easily (see Sec. 2.2.1), and as such are a good class from which to draw symmetry operators. If $\hat{S} \notin \mathbb{P}^N$, but the target eigenspace \mathcal{S} lies within the eigenspace of a Pauli operator \hat{P} , then measuring \hat{P} presents a low-cost alternative to measuring \hat{S} , though this may provide less error mitigation in the case where the eigenspace of \hat{P} is strictly larger than \mathcal{S} . In general, symmetry verification will work with any construction of a projector valued measurement $\{\hat{M}_i\}$ where one projector \hat{M}_S projects onto the target eigenspace \mathcal{S} . We note that phase estimation [86] provides a generic construction for such a measurement, although this is a rather high cost circuit (in particular requiring the ability to apply the symmetry \hat{U} on the quantum computer). This requirement for measurement implies that symmetry verification cannot be extended to antiunitary symmetries (nor to symmetries that anticommute with the Hamiltonian), as these do not lead to eigenspaces that can be projected into.

The projector valued measurement $\{\hat{M}_i\}$ is the more general object for symmetry verification than the symmetry \hat{S} . In an arbitrary quantum circuit at an arbitrary time, if we know by any means that the state $|\psi\rangle$ in the absence of error satisfies $\hat{M}_s|\psi\rangle = |\psi\rangle$, measuring $\{\hat{M}_i\}$ on the noisy state ρ and post-selecting will project to the state

$$\rho_s = \frac{\hat{M}_s \rho \hat{M}_s}{\text{Trace}[\hat{M}_s \rho]}. \quad (2.1)$$

Then, we have

$$\text{Trace}[\rho_s |\psi\rangle\langle\psi|] = \frac{\text{Trace}[\rho |\psi\rangle\langle\psi|]}{\text{Trace}[\hat{M}_s \rho]} \geq \text{Trace}[\rho |\psi\rangle\langle\psi|], \quad (2.2)$$

and our new state ρ_s has strictly greater overlap with the target $|\psi\rangle$ than the pre-selection ρ (unless $\hat{M}_s \rho \hat{M}_s = \rho$, in which case $\rho_s = \rho$). Such a procedure can be immediately extended to multiple operators $\hat{S}_1, \hat{S}_2, \dots$, as long as $[\hat{S}_i, \hat{S}_j] = 0$. (If this is not the case, sequential symmetry verification projects between different eigenspaces, which is inefficient and greatly increases the number of experiments that must be thrown away.) Symmetry verification may also be repeated at multiple points during a quantum circuit, by inserting measurement of \hat{S} in between gates, as long as we expect the state $\langle\psi(t)|$ to be an eigenstate of \hat{S} at time t during the circuit. We call such protocols ‘bulk’ symmetry verification, as opposed to ‘final’ symmetry verification at the end of the an experiment.

2.2.1 Ancilla and in-line symmetry verification

The simplest form of the symmetry verification involves the use of an ancilla qubit to measure the Pauli symmetry \hat{S} . Let us write $\hat{S} \in \mathbb{P}^N$ in terms of its tensor factors; $\hat{S} = \otimes_i \hat{S}_i$, and let N_S be the number of non-trivial $\hat{S}_i = \{X, Y, Z\}$. To each such \hat{S}_i , we can associate a corresponding rotation $\hat{R}_i = \{\exp(i\frac{\pi}{2}Y), \exp(-i\frac{\pi}{2}X), \mathbf{1}\}$ (such that $\hat{R}_i|\hat{S}_i = 1\rangle = |0\rangle$). The verification circuit is then shown in Fig. 2.1(a). For each non-trivial \hat{S}_i , the corresponding qubit is rotated by \hat{R}_i , then performs a controlled-NOT gate on the ancilla qubit, and finally is rotated by \hat{R}_i^{-1} . This requires that the ancilla qubit be coupled to each qubit in the system register that it measures, which is in general not possible in a quantum circuit. As a low-cost alternative (Fig. 2.1(b)), the ancilla qubit may be shuffled along the system register via SWAP gates as it performs the controlled phase gate. In either case, as the ancilla qubit must interact with each register qubit individually, the circuit depth must be $O(N_S)$.

2 Low-cost error mitigation by symmetry verification

It is possible to forego the ancilla qubit in symmetry verification, by instead encrypting the symmetry \hat{S} onto the computational degree of freedom of a qubit within the system itself, which is then read out. In Fig. 2.2(a) we give an example circuit for this in-line symmetry verification, with circuit depth only $O(\log(N_S))$. This logarithmic depth requires qubits to be coupled as a binary tree, which is not possible in systems which allow only local couplings. In general, for such a d -dimensional local coupling, the depth of the circuit must be at least $O(N_S^{1/d})$, being the minimum depth of a light-cone encompassing N_S qubits. In Fig. 2.2(b) we give such a circuit for a system with linear connectivity. Even when all-to-all coupling is available, the $O(\log(N_S))$ -depth circuit (Fig. 2.2(a)) may not be preferable, as the duty cycle for each qubit (i.e. the period of time between the first and last gate each qubit is involved in) is length $O(\log(N_S))$. By contrast, the duty cycle of an individual qubit during the circuit in Fig. 2.1(b) is $O(1)$. A short duty cycle implies that qubits can be used to perform other operations while the symmetry verification is ongoing, reducing the time cost when this circuit is performed as a small block of a larger computation.

2.2.2 Variational quantum eigensolvers

As an example target algorithm for symmetry verification, we consider ground state preparation for a Hamiltonian \hat{H} via a variational quantum eigensolver [39, 87].

Variational quantum eigensolvers (VQE) are natural candidates for final symmetry verification, and common classes of VQEs are also natural candidates for bulk symmetry verification. In particular, for fermionic systems (such as the electronic structure problem), global fermion parity is conserved, making it a prime target for symmetry verification. (At low energy, for non-superconducting systems, the particle number is often conserved as well, but this is not a Pauli operator, and is much more difficult to measure.) Using the Jordan-Wigner transformation on an N -fermion Hamiltonian, this symmetry takes the form $Z^{\otimes N}$. Most VQEs consist of creating an approximate starting state (such as the Hartree-Fock state) that respects this symmetry, and then performing multiple local rotations that continue to respect this symmetry. This is true of both the unitary coupled cluster (UCC) ansatz [39], and the quantum approximate optimization algorithm (QAOA) [88]. In the former, the ansatz is taken as

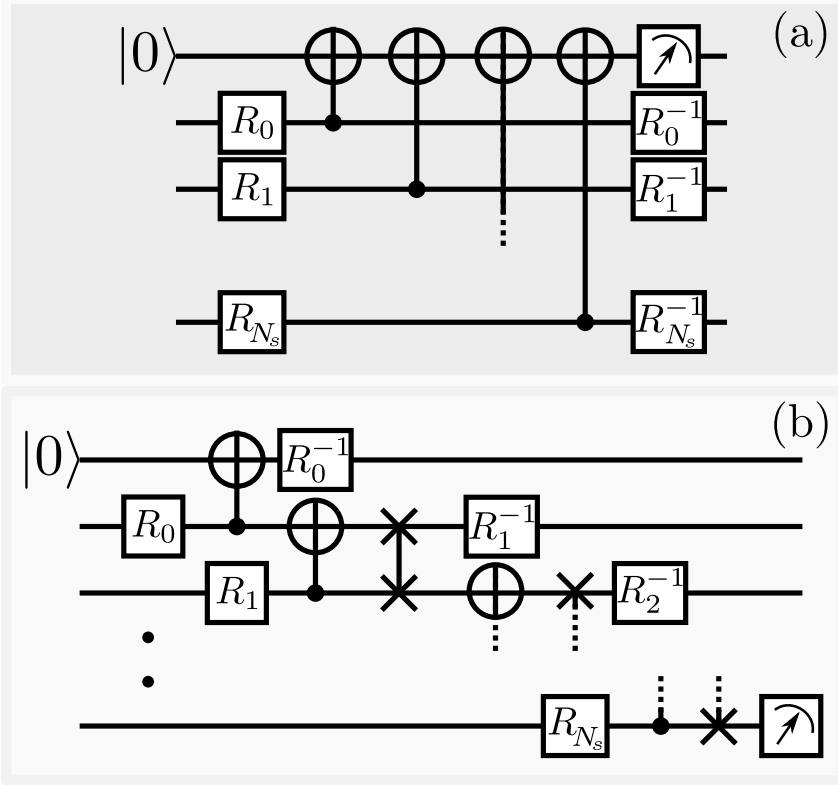


Figure 2.1: Quantum circuit for ancilla symmetry verification of a symmetry \hat{S} . (a) A simple circuit entangling all qubits with a single ancilla qubit. The rotations \hat{R}_i depend on the tensor components \hat{S}_i on each qubit i (relationship given in text). (b) A circuit making an identical measurement to that in (a), but with only local CNOT and SWAP two-qubit gates. A SWAP between qubit 0 and the ancilla is not required because the Bell state prepared after the first CNOT is symmetric between the two qubits (this is not the case for the remaining qubits).

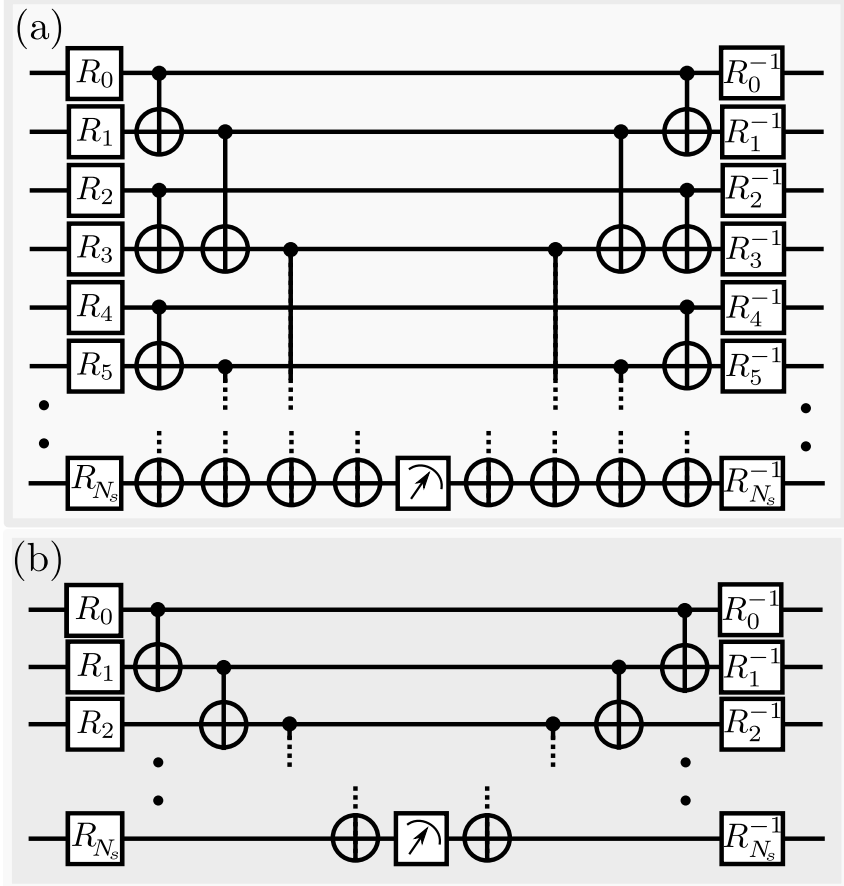


Figure 2.2: Quantum circuits for in-line symmetry verification. (a) The optimal verification circuit has $O(\log(N_s))$ depth, but requires long-range connectivity between qubits, which is not available on many architectures. (b) In the presence of linear connectivity, an $O(N_s)$ depth verification circuit is optimal.

the expansion of the cluster operator $e^{\hat{T}-\hat{T}^\dagger}$

$$\hat{T} = \sum_n \hat{T}^{(n)}, \quad (2.3)$$

$$\hat{T}^{(n)} = \sum_{i_1, \dots, i_n; j_1, \dots, j_n} \theta_{j_1, \dots, j_n}^{i_1, \dots, i_n} \left(\prod_{m=1}^n \hat{c}_{i_m}^\dagger \right) \left(\prod_{m=1}^n \hat{c}_{j_m} \right) \quad (2.4)$$

where the θ parameters are taken as the free parameters to be optimized, and the sum is a sum over empty molecular orbitals to the left of the semi-colon, and filled molecular orbitals to the right. This exponentiation is typically performed by the Trotter-Suzuki expansion, leaving a series of unitaries

$$\prod_{i,j} e^{\theta_j^i (\hat{c}_i^\dagger \hat{c}_j - \hat{c}_j^\dagger \hat{c}_i)} \prod_{i,j;k,l} e^{\theta_{k,l}^{i,j} (\hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_k \hat{c}_l - \hat{c}_l^\dagger \hat{c}_k^\dagger \hat{c}_j \hat{c}_i)} \dots \quad (2.5)$$

each of which respects fermion parity. QAOA for the electronic structure problem consists of performing steps of time evolution alternating between the Hartree-Fock Hamiltonian and the electronic-structure Hamiltonian, both of which respect fermion parity. Thus, for both ansatz, bulk symmetry verification could be performed between individual steps of the time evolution.

Although symmetry verification promises a final state with greater overlap with the ground state, it does not promise a necessarily lower energy. Let us write the (un-normalized) symmetry-accepted state ρ_s , and the symmetry-rejected state ρ_r . If our measurement was perfect, we would have

$$\rho_s = \hat{M}_s \rho \hat{M}_s, \quad \rho_r = (\mathbb{I} - \hat{M}_s) \rho (\mathbb{I} - \hat{M}_s). \quad (2.6)$$

Then, $\text{Trace}[\hat{H}\rho] = \text{Trace}[\hat{H}\rho_r] + \text{Trace}[\hat{H}\rho_s]$. Now, suppose the rejected state ρ_r has lower energy than the accepted state ρ_s ;

$$\frac{\text{Trace}[\hat{H}\rho_r]}{\text{Trace}[\rho_r]} < \frac{\text{Trace}[\hat{H}\rho_s]}{\text{Trace}[\rho_s]}. \quad (2.7)$$

We can calculate

$$\begin{aligned} \text{Trace}[\hat{H}\rho] &= \text{Trace}[\rho_r] \frac{\text{Trace}[\hat{H}\rho_r]}{\text{Trace}[\rho_r]} + \text{Trace}[\rho_s] \frac{\text{Trace}[\hat{H}\rho_s]}{\text{Trace}[\rho_s]} \\ &< (\text{Trace}[\rho_r] + \text{Trace}[\rho_s]) \frac{\text{Trace}[\hat{H}\rho_s]}{\text{Trace}[\rho_s]} = \frac{\text{Trace}[\hat{H}\rho_s]}{\text{Trace}[\rho_s]}, \end{aligned}$$

and our symmetry-verified state would be higher in energy than the initial state as well. As the energy of ρ_r lies strictly above the ground state, failure of symmetry verification must imply ρ_s has sufficiently large overlap

with high-energy states. As such, we would suggest that such a failure implies the energy of ρ itself is not to be trusted.

2.2.3 Post-selected symmetry verification and S-QSE

Conveniently, when a quantum computation requires calculating the expectation values of a set of Pauli operators, symmetry verification may be performed via post-processing of the expectation values themselves (with possibly some additional measurements), rather than requiring additional quantum circuitry. Suppose we want to calculate the expectation value of $\hat{P} \in \mathbb{P}^N$ on our state ρ following projection onto the $\hat{S} = s (= \pm 1)$ subspace of our symmetry $\hat{S} \in \mathbb{P}^N$. The projector onto this subspace may be written $\hat{M}_s = \frac{1}{2}(1 + s\hat{S})$. Then, the expectation value of \hat{P} on the state ρ_s targeted by the symmetry verification can be expanded using Eq. 2.1

$$\begin{aligned} \text{Trace}[\hat{P}\rho_s] &= \text{Trace}\left[\hat{P} \frac{\hat{M}_s \rho \hat{M}_s}{\text{Trace}[\hat{M}_s \rho]}\right] \\ &= \frac{\text{Trace}[\hat{P}\rho] + s\text{Trace}[\hat{P}\hat{S}\rho]}{1 + s\text{Trace}[\hat{S}\rho]}, \end{aligned} \quad (2.8)$$

where we have used the cyclic property of the trace and the fact that $[\hat{P}, \hat{M}_s] = 0$ to write $\text{Trace}[\hat{P}\hat{M}_s \rho \hat{M}_s] = \text{Trace}[\hat{P}\hat{M}_s \rho]$, and expanded our definition of \hat{M}_s . The expectation values $\text{Trace}[\hat{S}\rho]$, $\text{Trace}[\hat{P}\rho]$, and $\text{Trace}[\hat{P}\hat{S}\rho]$ may be then calculated using the unverified state ρ , and substituted into Eq. 2.8 to obtain the verified result. By avoiding additional quantum circuitry, we expect this method to outperform both ancilla and in-line symmetry verification. However, we note that post-selection cannot be used for bulk symmetry verification (as we cannot measure these expectation values during the circuit). Furthermore, it cannot be used in algorithms where the output is not an expectation value $\text{Trace}[\hat{P}\rho]$.

Post-selected symmetry verification can be observed to be identical to a form of the quantum subspace expansion (QSE) [82]. Originally designed to investigate the low-energy excited states around the ground space found by a variational quantum eigensolver, QSE works by taking a set of excitation operators $\{\hat{E}_i\}$, which can be applied to the approximated ground state $|\psi(\vec{\theta})\rangle$ to obtain a set of states $|\phi_i\rangle = \hat{E}_i|\psi(\vec{\theta})\rangle$. The spectrum of the Hamiltonian within the manifold spanned by these states can be calculated as the solution to the generalized eigenvalue problem

$$\hat{H}_{\text{QSE}}|\xi\rangle = \lambda \hat{B}_{\text{QSE}}|\xi\rangle. \quad (2.9)$$

Here, \hat{H}_{QSE} is the Hamiltonian matrix projected into the spanned manifold

$$[\hat{H}_{\text{QSE}}]_{i,j} = \text{Trace}[\hat{H}|\phi_i\rangle\langle\phi_j|], \quad (2.10)$$

and \hat{B}_{QSE} is the overlap matrix,

$$[\hat{B}_{\text{QSE}}]_{i,j} = \text{Trace}[|\phi_i\rangle\langle\phi_j|], \quad (2.11)$$

to account for the fact that $|\phi_i\rangle$ and $|\phi_j\rangle$ are in general not orthogonal. In the presence of noise, although the state $|\phi_i\rangle$ is not well defined (as our noisy state ρ is not a pure state), the operators $|\phi_i\rangle\langle\phi_j| = \hat{E}_i\rho\hat{E}_j^\dagger$ remain well-defined, and the expectation values in Eqs. 2.10 and 2.11 are still able to be measured in an experiment.

The set $\{\hat{E}_i\}$ is usually taken to be the set of low-order polynomials in qubit or fermion operators [82, 83]. However, if the set $\{\mathbb{I}, \hat{S}\}$ is chosen as excitation operators, the solution to the generalized eigenvalue problem is the same as that obtained by post-selection. To show this, we expand

$$\begin{aligned} \text{Trace}[\hat{H}\rho_s] &= \sum_i h_i \text{Trace}[\hat{P}_i\rho_s] \\ &= \sum_i \frac{\text{Trace}[h_i\hat{P}_i\rho] + s\text{Trace}[h_i\hat{P}_i\hat{S}\rho]}{1 + s\text{Trace}[\hat{S}\rho]} \\ &= \frac{\text{Trace}[\hat{H}\rho] + s\text{Trace}[\hat{H}\hat{S}\rho]}{1 + s\text{Trace}[\hat{S}\rho]}. \end{aligned} \quad (2.12)$$

Next, we calculate the QSE matrices (using the commutation of \hat{H} and \hat{S})

$$\hat{H}_{\text{QSE}} = \begin{bmatrix} \text{Trace}[\hat{H}\rho] & \text{Trace}[\hat{H}\hat{S}\rho] \\ \text{Trace}[\hat{H}\hat{S}\rho] & \text{Trace}[\hat{H}\rho] \end{bmatrix}, \quad (2.13)$$

$$\hat{B}_{\text{QSE}} = \begin{bmatrix} 1 & \text{Trace}[\hat{S}\rho] \\ \text{Trace}[\hat{S}\rho] & 1 \end{bmatrix}. \quad (2.14)$$

Assuming that $\text{Trace}[\hat{S}\rho] \neq 1$, \hat{B}_{QSE} is invertible, the problem reduces to finding the (regular) eigenvalues of

$$\hat{B}_{\text{QSE}}^{-1}\hat{H}_{\text{QSE}} = \frac{1}{1 - \text{Trace}[\hat{S}\rho]^2} \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix}, \quad (2.15)$$

where

$$\alpha = \text{Trace}[\hat{H}\rho] - \text{Trace}[\hat{S}\rho]\text{Trace}[\hat{H}\hat{S}\rho], \quad (2.16)$$

$$\beta = \text{Trace}[\hat{H}\hat{S}\rho] - \text{Trace}[\hat{H}\rho]\text{Trace}[\hat{S}\rho]. \quad (2.17)$$

The eigenvalues of this matrix take the form

$$\lambda = \frac{1}{1 - \text{Trace}[\hat{S}\rho]^2}(\alpha \pm \beta) \quad (2.18)$$

$$= \frac{\text{Trace}[\hat{H}\rho] \pm \text{Trace}[\hat{H}\hat{S}\rho]}{1 \pm \text{Trace}[\hat{S}\rho]}, \quad (2.19)$$

which can be seen to be equal to those found in Eq. 2.12. We call this version of the quantum subspace expansion symmetry-QSE, or S-QSE for short.

This result is not surprising; it was suggested in [82] to account for symmetries during QSE by projecting \hat{H}_{QSE} and \hat{B}_{QSE} into the symmetry subspace, which achieves the same result as in the above. However, this demonstrates that one may account for symmetries via a version of QSE without calculating the full linear response. Moreover, this implies that S-QSE corrects for both coherent and incoherent errors that project out of the $\hat{S} = s$ subspace. By contrast, QSE with an operator that anticommutes with the Hamiltonian can only correct coherent errors (see appendix). S-QSE may be immediately combined with other forms of QSE, for example linear response QSE, by including both sets of operators as excitations.

2.3 Simulation of symmetry verification on the hydrogen molecule

To first investigate symmetry verification in a simple setting, we use a VQE to find the ground-state energy of H_2 on two qubits. This follows previous experimental demonstrations [39, 41, 42, 83]. We take the STO-3G basis for H_2 , which has four spin-orbitals, and convert this into a qubit Hamiltonian via the Bravyi-Kitaev transformation. The four spin-orbitals require four qubits to represent them on, but in this representation the Hamiltonian is diagonal on two of the qubits, which may be removed. The remaining two-qubit Hamiltonian takes the form

$$\hat{H} = h_0\mathbb{I}\mathbb{I} + h_1\mathbb{I}Z + h_2Z\mathbb{I} + h_3XX + h_4YY + h_5ZZ, \quad (2.20)$$

where h_i are sums of integrated two and four-body terms from the original electronic structure problem. The calculation of these terms, and the Bravyi-Kitaev transformation itself, were performed using the psi4 [89] and OpenFermion [90] packages. The Hamiltonian can be seen to commute with the symmetry $\hat{S} = ZZ$. Our ground state wavefunction has

2.3 Simulation of symmetry verification on the hydrogen molecule

non-trivial overlap with the Hartree-Fock wavefunction, which is in the $ZZ = -1$ subspace; this is then our target subspace. We follow the unitary coupled cluster ansatz of [41], which consists of exciting our system to the $|01\rangle$ state, and performing the unitary rotation

$$\hat{U}(\theta) = e^{-i\theta X_0 Y_1}. \quad (2.21)$$

This unitary rotation may be decomposed using standard methods [31]. As described previously, the VQE procedure consists of fixing θ , repeatedly preparing $|\psi(\theta)\rangle$ and measuring collections of terms in the Pauli decomposition of \hat{H} until a good estimate of the energy $E(\theta)$ is found. This is then repeated at varying θ as demanded by a classical optimizer until a minimum $E(\theta)$ is found [39].

We compare the performance of the three symmetry verification protocols described previously as a final symmetry verification step. The ancilla symmetry verification is performed in the same manner as Fig. 2.1(a). The in-line symmetry verification is performed in a manner similar to Fig. 2.2(a), but as this is final symmetry verification, we have no need to undo the symmetry measurement. Instead, to measure the expectation value of a Pauli operator $\text{Trace}[\rho \hat{P}]$, we can propagate \hat{P} through the symmetry verification circuit [91] and measure the corresponding Pauli term. It is then sufficient to rotate the control qubit to recover the expectation values $\langle \mathbb{I}Z \rangle$ and $\langle XX \rangle$. From this we may calculate all other expectation values in Eq. 2.20 using the fact that $ZZ = -1$. For this problem, S-QSE not only requires no additional circuitry, but also no additional measurements (all required terms are in the Pauli decomposition of the Hamiltonian).

To test symmetry verification in the presence of realistic noise, we simulate our chosen experiment using the quantumsim density matrix simulator [92]. We take gate error models and parameters similar to previous simulation work based on experimental data of state-of-the-art superconducting transmon qubits [92]. Errors in transmon qubits are dominated by decoherence times, which we take at a base level to be $T_1 = T_2 = 20 \mu\text{s}$. This should be compared to single and two-qubit gate times of 20 ns (giving a total circuit length without symmetry verification of 220 ns). Single and two-qubit gates suffer from additional dephasing noise of 0.01 and 1% respectively. We assume that single-shot measurement (for verification purposes) has a read-out error of 1%, and that error in tomographic measurements and pre-rotations (used to calculate the expectation values themselves) can be cancelled by linear inversion tomography [93, 94].

Using the above error model, we observe (Fig. 2.3) that the un-mitigated VQE (blue points) achieves an error in the energy of approximately 0.01 –

0.04 hartree across the bond dissociation curve. This error is improved upon by all symmetry verification techniques. S-QSE (red diamonds) provides the largest improvement of all symmetry verification protocols, as no additional errors are introduced. The S-QSE circuit is observed to give approximately a five-fold improvement over the unmitigated circuit, while ancilla (orange crosses) and in-line (green squares) symmetry verification show an approximately two-fold and three-fold improvement respectively. The differences between S-QSE and other forms of symmetry verification emphasize the importance of minimizing the verification cost in bulk symmetry verification (where S-QSE is no longer available).

We now investigate the effect of different noise channels on the performance of symmetry verification. Any noise channel that commutes with the symmetry operators evolves the system state within the target subspace, which symmetry verification explicitly does not mitigate. The analysis of which channels have this property can be reduced to an analysis over \mathbb{P}^N , as if we mitigate Pauli errors $\hat{P}_i \in \mathbb{P}^N$, we also mitigate any linear combination of them [84]. In the above circuit, the ZZ symmetry commutes with any single-qubit Z errors, making the protocol prone to the T_ϕ (pure dephasing) channel, but it anticommutes with single-qubit X -errors, making the protocol resilient against the T_1 (amplitude decay) channel. To investigate this, in Fig. 2.4 we calculate the error in determining the ground state energy near the minima of the bond dissociation curve (0.75Å bond distance) using S-QSE, as we vary T_1 and T_ϕ . We turn all other error sources off, and vary T_1 (T_ϕ) with $T_\phi = 20 \mu\text{s}$ ($T_1 = 20 \mu\text{s}$) fixed. In the absence of error mitigation, the two decoherence sources have almost identical effect (deviation approximately 10^{-2} hartree). However, in the presence of error mitigation, the susceptibility of the VQE to T_1 noise is noticeably smaller than to T_ϕ noise - up to a factor of two over the range of decoherence times plotted. We note that S-QSE does not make our circuit second-order sensitive to T_1 noise. This can be understood as X -errors at some points during our VQE circuit are rotated to Z -errors by later gates in the circuit, preventing their mitigation.

2.4 Inserting and rotating symmetries

As observed in the previous section, verifying single symmetries has a marked effect on the performance of a quantum circuit, but will not catch and remove all sources of noise. In this section we suggest how one may improve upon this by adding additional symmetries to the quantum algorithm, and by rotating existing symmetries to make them more sensitive

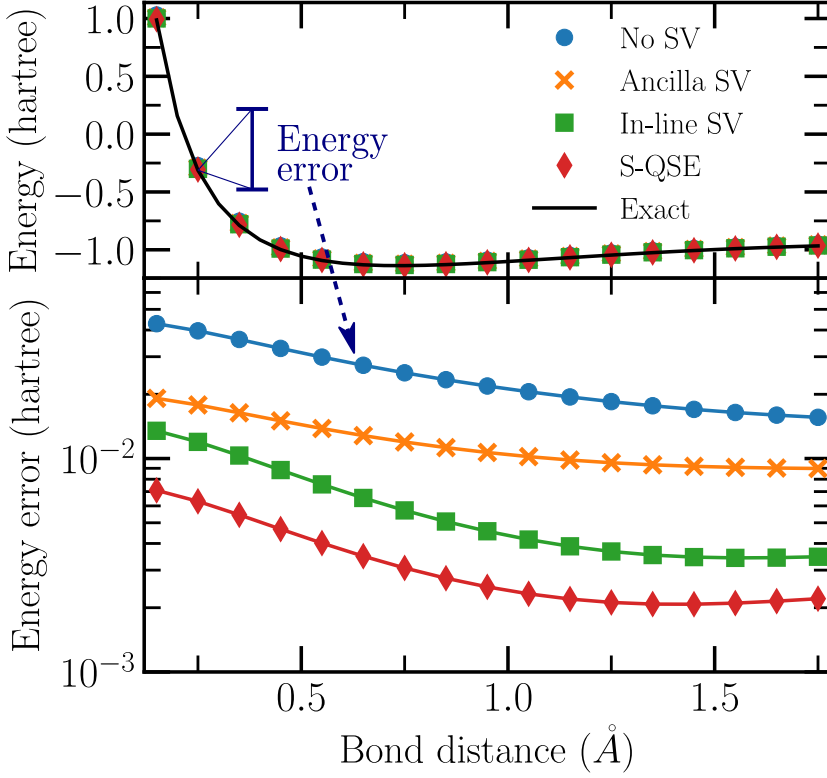


Figure 2.3: (Color online) Accuracy of the VQE over the entire bond dissociation curve using the different symmetry verification methods mentioned in the text (labelled in legend). (top) The target curve of H_2 , compared to the exact result (black line). (bottom) Log plot of the difference between the black lines and points in the above plot.

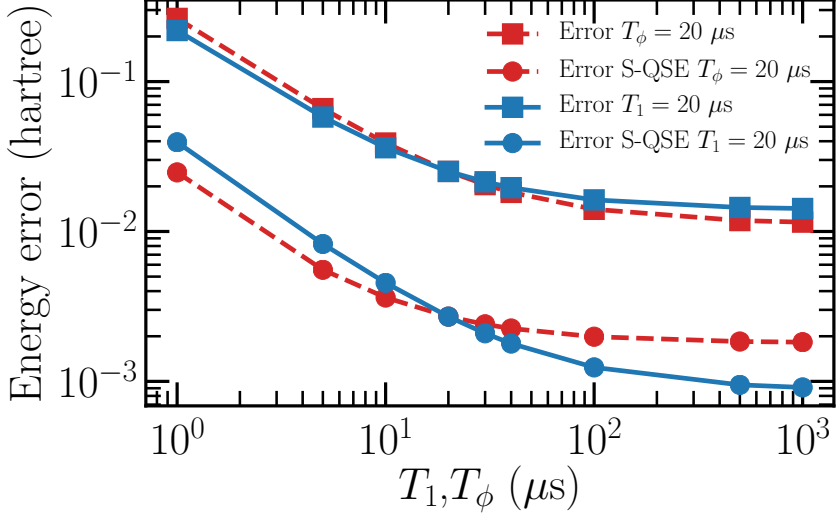


Figure 2.4: (Color online) Effect of varying decoherence times on the VQE accuracy. With all other error sources turned off, T_1 is varied with $T_\phi = 20 \mu\text{s}$ fixed (red-dashed curves), and T_ϕ is varied with $T_1 = 20 \mu\text{s}$ fixed (blue-solid curves). We plot the error in estimating the ground-state energy for the unmitigated experiment (squares), and the circuit mitigated with S-QSE (circles). Data points for the blue and red curves are identical at $T_1 = T_\phi = 20 \mu\text{s}$, as can be seen from the complete overlap.

to errors on the underlying quantum hardware. In the language of quantum error correction, this is a low-cost attempt to increase the distance of the detection code.

We first suggest a method to extend an N -qubit Hamiltonian \hat{H} , given a Pauli operator $\hat{P} \in \mathbb{P}^N$, to an $N + 1$ -qubit Hamiltonian \hat{H}_{ext}

$$\hat{H}_{\text{ext}} = \begin{bmatrix} \hat{H} & 0 \\ 0 & \hat{P}\hat{H}\hat{P} \end{bmatrix}. \quad (2.22)$$

Both blocks of \hat{H}_{ext} can be seen to have the same eigenspectrum (as this is unaffected by the unitary rotation of \hat{P}), and \hat{H}_{ext} commutes with the operator

$$\begin{bmatrix} 0 & \hat{P} \\ \hat{P} & 0 \end{bmatrix} = X\hat{P}, \quad (2.23)$$

which is then the new symmetry operator. This mapping corresponds to mapping Pauli operators $\hat{Q} \in \mathbb{P}^N$ in the original problem to

$$\hat{Q}_{\text{ext}} = \begin{cases} \mathbb{I}\hat{Q} & \text{if } [\hat{Q}, \hat{P}] = 0 \\ X\hat{Q} & \text{if } \{\hat{Q}, \hat{P}\} = 0 \end{cases}. \quad (2.24)$$

To implement this in the algorithm itself, we note that every circuit can be decomposed into a product of unitary rotations

$$\prod_j e^{i\theta_j \hat{Q}_j}, \quad \hat{Q}_j \in \mathbb{P}^N, \quad (2.25)$$

where a single $\hat{Q} \in \mathbb{P}^N$ may be repeated in the product. Adding the symmetry then consists of replacing these rotations by rotations around the transformed operator \hat{Q}_{ext} (as per Eq. 2.24), and re-decomposing the operations into a circuit (using e.g. the methods of [31, 95]). If \hat{H} had a previous set of symmetries \hat{S}_i , these are transformed to a new set $\hat{S}_{i,\text{ext}}$ (following Eq. 2.24), that commute with both \hat{H}_{ext} and the additional symmetry $X\hat{P}$. This extension method is particularly suitable for digital quantum simulation, as circuits are often generated in the form of Eq. 2.25. This is the case for traditional Hamiltonian simulation [96], quantum phase estimation [31], and the UCC QSE discussed previously, all of which require exponentiating an operator via the Suzuki-Trotter expansion [97].

Beyond choosing the number of symmetries in a problem, one may wish to choose how these symmetries appear in the problem. In particular, sets of symmetries may be found that anticommute with all local

operators, which should increase the mitigation power of the verification protocol against local sources of noise. (For example, the N -qubit operators $X^{\otimes N}$ and $Z^{\otimes N}$ with even N .) Any two groups of M Pauli operators are unitarily equivalent as long as they satisfy the same commutation and multiplication rules (e.g. $\mathbb{I}Z, Z\mathbb{I}$, and ZZ are equivalent to XX, YY and ZZ , but not to $\mathbb{I}X, \mathbb{I}Y$ and $\mathbb{I}Z$). To find such unitary transformations, we suggest decomposing them into rotations of the form $\hat{R} = e^{i\frac{\pi}{2}\hat{Q}}$ for $\hat{Q} \in \mathbb{P}$, which transforms

$$\hat{P} \in \mathbb{P} \rightarrow \hat{R}^\dagger \hat{P} \hat{R} = \begin{cases} \hat{P} & \text{if } [\hat{P}, \hat{Q}] = 0 \\ i\hat{P}\hat{Q} & \text{if } \{\hat{P}, \hat{Q}\} = 0 \end{cases}. \quad (2.26)$$

Rotations of this form have a few desirable properties. Their effect is easy to calculate classically, and they transform Pauli operators to Pauli operators. Furthermore, each \hat{R} leaves half of the Pauli group unchanged. This allows for some choice of rotations to leave desired symmetries (or other operators) already present in the problem invariant, while other terms are rotated.

2.5 Extending the symmetry verification of the hydrogen molecule

We now demonstrate the verification of multiple symmetries by extending the previous VQE simulation of H_2 . We transform the electronic structure Hamiltonian onto a qubit representation this time via the Jordan-Wigner transformation. This gives the four-qubit Hamiltonian

$$\begin{aligned} \hat{H} = & h_I \mathbb{I} + \sum_i h_i Z_i + \frac{1}{2} \sum_{i \neq j} h_{i,j} Z_i Z_j \\ & + h_s (X_0 Y_1 Y_2 X_3 + Y_0 X_1 X_2 Y_3 \\ & - X_0 X_1 Y_2 Y_3 - Y_0 Y_1 X_2 X_3), \end{aligned} \quad (2.27)$$

which has symmetries $\hat{S}_0 = Z_0 Z_1$, $\hat{S}_1 = Z_0 Z_2$, and $\hat{S}_2 = Z_0 Z_1 Z_2 Z_3$. In the Bravyi-Kitaev transformation these symmetries were the additional qubits that were thrown away. We choose again the unitary coupled cluster ansatz for the VQE, which can be reduced to the operator¹

$$\hat{U}(\theta) = e^{i\theta Y_0 X_1 X_2 X_3}. \quad (2.28)$$

¹The cluster operator for this system is a sum of 8 four-qubit terms, however the action of each term on the Hartree-Fock starting state is identical, so only one is needed.

2.5 Extending the symmetry verification of the hydrogen molecule

As in the two-qubit case, the VQE circuit consists of preparing the system in the Hartree-Fock state $|1100\rangle$, applying $U(\theta)$ and measuring the variational energy, for a total circuit time of 400 ns.

The above set of symmetries still commute with all single-qubit Z errors, so we rotate our problem to increase the mitigation power of symmetry verification. We choose the rotation

$$\hat{R} = e^{i\frac{\pi}{2}Y_0X_2}e^{i\frac{\pi}{2}Y_1X_3}. \quad (2.29)$$

This transforms the symmetry $\hat{S}_0 \rightarrow X_0X_1X_2X_3$, whilst leaving \hat{S}_1 and \hat{S}_2 unchanged. The resulting set of symmetries do not commute with any single-qubit X or Z operator, as required. To create the transformed circuit, we need to transform both our starting state $|1100\rangle \rightarrow \hat{R}|1100\rangle$, and the UCC unitary ansatz

$$\hat{U}(\theta) \rightarrow \hat{R}\hat{U}\hat{R}^\dagger = e^{i\theta Y_0Z_1X_2}. \quad (2.30)$$

The transformed circuit incurs an additional cost from this initial application of \hat{R} , but this is balanced by the reduced weight of the transformed cluster operator, resulting in a total circuit time of 440 ns.

In Fig. 2.5, we compare the performance of the two different circuits above to the two-qubit circuit of Fig. 2.3, with and without the addition of S-QSE. At small bond distance ($\lesssim 0.75\text{\AA}$), the target ground state (in the absence of rotation by \hat{R}) is roughly a computational basis state, which is immune to dephasing errors. At this point, all three verification protocols perform roughly similarly, despite the unmitigated four-qubit simulations performing significantly worse than the unmitigated two-qubit simulation. At large bond distance ($\gtrsim 0.75\text{\AA}$), the ground state is prone to T_2 noise, at which point we see the rotated 4-qubit S-QSE simulation significantly outperforming its counterparts. At the largest distance studied, this simulation achieves a two-fold reduction in error compared to the two-qubit S-QSE simulation, despite using twice as many qubits and a twice as long circuit. By comparison, unrotated S-QSE on four qubits cannot protect against the T_2 noise accumulated over the simulation, and performs a factor of two worse than the two-qubit S-QSE simulation. This clearly demonstrates the need to optimize symmetry verification protocols to account for errors present in the system as this technique is scaled up to larger computations. Over the entire bond-dissociation curve, the rotated four-qubit S-QSE simulation outperforms its unmitigated counterpart by over an order of magnitude.

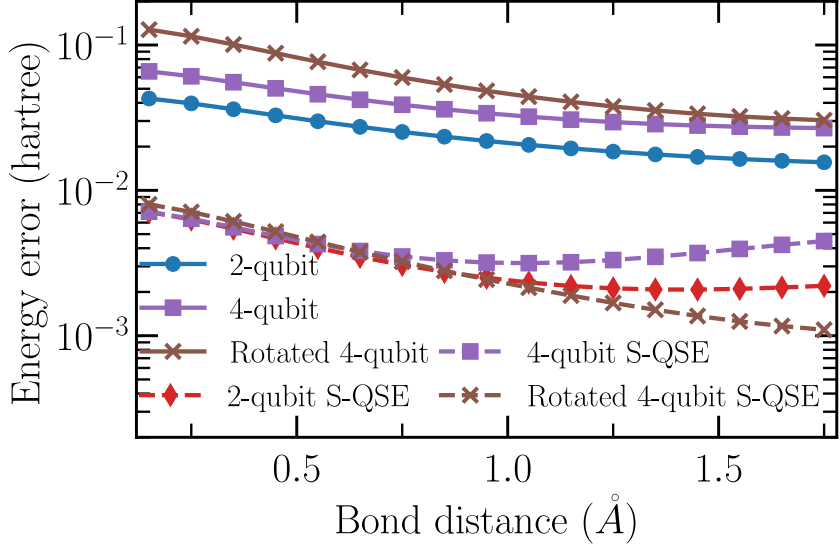


Figure 2.5: (Color online) Adding and adjusting symmetries to optimize symmetry verification. The blue (dots) and red (diamonds) curves correspond to their coloured (shaped) counterparts in Fig. 2.3, whilst the purple (squares) and brown (crosses) curves come from a four-qubit simulation of H_2 using the two protocols described in the text. The dashed lines represent the S-QSE versions of their solid counterparts. Error parameters on all qubits are the same for all simulations (parameters given in the text).

2.6 Conclusion

In this paper we have presented a new low-cost strategy for error mitigation, which we call symmetry verification. We have discussed various ways in which it can be applied to different algorithms, and various methods to optimize the mitigation power against common sources of error. We have demonstrated these protocols on a simulated VQE experiment of H_2 , and observed that they outperform the unmitigated result over the entire bond-dissociation curve by around an order of magnitude.

Although the above techniques are very promising for small experiments, much work needs to be done optimizing symmetry verification for mid-range experiments in the NISQ era. The addition and choice of symmetries needs to be investigated further to minimize the resulting circuit depth. Further study is also needed on the optimal number of symmetry verifications to be added to a circuit, both to maximise mitigation and minimize run-time (which increases exponentially in the number of verifications made). Finally, given the obvious connection between symmetry verification and the stabilizer formalism of quantum error correction, it is natural to ask whether one can mix the two to transform slowly between mid-size NISQ circuits and large-scale fault-tolerant ones.

Appendix

2.A Appendix: Error mitigation of QSE with anticommuting operators

In this appendix we repeat the analysis of QSE from the text, but with an operator \hat{A} that anti-commutes with the Hamiltonian \hat{H} . Let us assume to begin that \hat{A} is unitary. Such an operator cannot be simultaneously diagonalized with \hat{H} , and so we have no result from symmetry verification to compare with. Given an eigenstate $\hat{H}|\psi\rangle = E|\psi\rangle$, we have that $\hat{H}\hat{A}|\psi\rangle = -\hat{A}\hat{H}|\psi\rangle = -E\hat{A}|\psi\rangle$, and so the presence of an anticommuting operator splits the eigenstates of \hat{H} into pairs of equal magnitude but opposite sign energies (known as eigenstates of different chirality). If $\hat{A} = \hat{A}^\dagger$, the eigenstates of \hat{A} itself are the equal superpositions

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|\psi\rangle \pm \hat{A}|\psi\rangle). \quad (2.31)$$

For QSE, we must calculate the operators \hat{H}_{QSE} and \hat{B}_{QSE} .

$$\hat{B}_{\text{QSE}} = \begin{bmatrix} 1 & \text{Trace}[\hat{A}\rho] \\ \text{Trace}[\hat{A}^\dagger\rho] & 1 \end{bmatrix}. \quad (2.32)$$

$$\hat{H}_{\text{QSE}} = \begin{bmatrix} \text{Trace}[\hat{H}\rho] & \text{Trace}[\hat{H}\hat{A}\rho] \\ \text{Trace}[-\hat{H}\hat{A}^\dagger\rho] & -\text{Trace}[\hat{H}\rho] \end{bmatrix}. \quad (2.33)$$

Again assuming $|\text{Trace}[\hat{A}\rho]|^2 \neq 1$, we can invert \hat{B}_{QSE} and calculate

$$\hat{B}_{\text{QSE}}^{-1}\hat{H}_{\text{QSE}} = \frac{1}{1 - |\text{Trace}[\hat{A}\rho]|^2} \begin{bmatrix} \alpha & \beta \\ -\beta^* & -\alpha^* \end{bmatrix}, \quad (2.34)$$

where

$$\alpha = \text{Trace}[\hat{H}\rho] + \text{Trace}[\hat{A}\rho]\text{Trace}[\hat{H}\hat{A}\rho] \quad (2.35)$$

$$\beta = \text{Trace}[\hat{H}\hat{A}\rho] + \text{Trace}[\hat{H}\rho]\text{Trace}[\hat{A}\rho]. \quad (2.36)$$

2 Low-cost error mitigation by symmetry verification

The solution to the equation is

$$E_{\text{QSE}}^2 = \frac{|\alpha|^2 + |\beta|^2}{(1 - |\text{Trace}[\hat{A}\rho]|^2)^2} \quad (2.37)$$

$$= \frac{\text{Trace}[\hat{H}\rho]^2 + |\text{Trace}[\hat{H}\hat{A}\rho]|^2}{1 - |\text{Trace}[\hat{A}\rho]|^2}. \quad (2.38)$$

To understand the gain in energy, $|\text{Trace}[\hat{H}\hat{A}\rho]|^2$, let us first consider a single set of opposite chirality states $|\psi\rangle$ and $\hat{A}|\psi\rangle$ (with energy $\pm E$). We first note that if ρ is an incoherent superposition of the eigenstates,

$$\rho = |a|^2 |\psi\rangle\langle\psi| + |b|^2 \hat{A}|\psi\rangle\langle\psi|\hat{A}, \quad (2.39)$$

$\text{Trace}[\hat{H}\hat{A}\rho] = \text{Trace}[\hat{A}\rho] = 0$ (as $\langle\psi|\hat{A}|\psi\rangle = 0$), and QSE strictly does not improve on the estimate of the ground state energy. We next consider the opposite situation, where ρ is a coherent superposition of eigenstates:

$$\begin{aligned} \rho = & (\cos(\theta)|\psi\rangle + \sin(\theta)e^{i\phi}\hat{A}|\psi\rangle) \\ & \times (\cos(\theta)\langle\psi| + \sin(\theta)e^{-i\phi}\langle\psi|\hat{A}^\dagger). \end{aligned} \quad (2.40)$$

We can calculate

$$\text{Trace}[\hat{H}\rho] = E \cos(2\theta), \quad (2.41)$$

$$\text{Trace}[\hat{A}\rho] = \sin(2\theta)(1 + Ae^{i\phi}), \quad (2.42)$$

$$\text{Trace}[\hat{H}\hat{A}\rho] = E \sin(2\theta)(Ae^{i\phi} - 1), \quad (2.43)$$

where $A = \langle\psi|\hat{A}^2|\psi\rangle$ (so $|A| \leq 1$, and for $\hat{A} \in \mathbb{P}^N$, $A = 1$). This gives

$$E_{\text{QSE}}^2 = E^2 \frac{\cos^2(2\theta) + \sin^2(2\theta)\chi_+}{1 - \sin^2(2\theta)\chi_-}, \quad (2.44)$$

$$\chi_{\pm} = (1 \pm Ae^{i\phi})(1 \pm Ae^{-i\phi}). \quad (2.45)$$

We see that if $A = 1, \phi = \frac{\pi}{2}$, QSE corrects the coherent error entirely, whilst if $A = 1, \phi = 0$ it has no effect. This implies that QSE cannot correct coherent rotations of ρ from $|\psi\rangle$ towards an eigenstate of \hat{A} . This is in keeping with the general observations in [82] for the performance of QSE as an error mitigation strategy.

If \hat{A} is not unitary, then $\hat{A}^\dagger \hat{A}$ is a Hermitian operator that commutes with \hat{H} . Importantly, if $\{\hat{A}, \hat{H}\} = 0$, $\{\hat{A}\hat{H}, \hat{H}\} = 0$ as well, giving a second anticommuting operator that is in general non-unitary. This could be

2.A Appendix: Error mitigation of QSE with anticommuting operators

used directly in QSE, although the analysis of Sec. 2.2.3 no longer holds unless $\hat{A}^\dagger \hat{A} \in \mathbb{P}^2$. For symmetry verification, we require the form of the projector \hat{M}_a onto the correct $\hat{A}^\dagger \hat{A}|\psi\rangle = a|\psi\rangle$ subspace. This is a difficult task in general to construct (for $\hat{A}\hat{H}$, it is equivalent to diagonalizing the Hamiltonian). We have been unable to construct any further bounds on the performance of QSE as an error mitigation strategy for a general Hermitian operator, nor for an operator which neither commutes nor anti-commutes with \hat{H} . This is, however, an interesting direction for future research.

3 Experimental error mitigation via symmetry verification in a variational quantum eigensolver

3.1 Introduction

oisy intermediate-scale quantum (NISQ) devices [47], despite lacking layers of quantum error correction (QEC), may already be able to demonstrate quantum advantage over classical computers for select problems [18, 19]. In particular, the hybrid quantum-classical variational quantum eigensolver (VQE) [39, 87] may have sufficiently low experimental requirements to allow estimation of ground-state energies of quantum systems that are difficult to simulate purely classically [76–79]. To date, VQEs have been used to study small examples of the electronic structure problem, such as H_2 [41, 42, 45, 80, 83, 98], HeH^+ [39, 99], LiH [42, 80, 98], and BeH_2 [42], as well as exciton systems [100], strongly correlated magnetic models [80], and the Schwinger model [101]. Although these experimental efforts have achieved impressive coherent control of up to 20 qubits, the error in the resulting estimations has remained relatively high due to performance limitations in the NISQ hardware. Consequently, much focus has recently been placed on developing error mitigation techniques that offer order-of-magnitude accuracy improvement without the costly overhead of full QEC. This may be achieved by using known properties of the target state, e.g., by checking known symmetries in a manner inspired by QEC stabilizer measurements [52, 53], or by expanding around the experimentally-obtained state via a linear (or higher-order) response framework [82]. The former, termed symmetry verification (SV), is of particular interest because it is comparatively low-cost in terms of required hardware and additional measurements. Other mitigation techniques require understanding the underlying error models of the quantum device, allowing for an extrapolation of the calculation to the zero-error

limit [49, 50, 81], or the summing of multiple calculations to probabilistically cancel errors [50, 51, 102].

We experimentally demonstrate the use of SV to reduce the error of a VQE estimating the ground-state energy and the ground state of the H_2 molecule by one order of magnitude on average across the bond-dissociation curve. Using two qubits in a circuit QED processor, we prepare a variational ansatz state via an exchange gate that finely controls the transfer of population within the single-excitation subspace while respecting the underlying symmetry of the problem (odd two-qubit parity). We show that SV improves the energy and state estimates by mitigating the effect of processes changing total excitation number, specifically qubit relaxation and residual qubit excitation. We do this through a full density-matrix simulation that matches the experimental energy and state errors with and without SV, and then using this simulation to dissect the contribution of each error source. Finally, we explore the limitations of SV arising from statistical measurement noise, and find that enforcing the positivity of the fermionic 2-reduced density matrix ties the improvement in energy estimation from SV to the improvement in ground-state fidelity (which was previously not the case).

3.2 Variational quantum eigensolvers for the Hydrogen molecule

A VQE algorithm [39, 87] approximates the ground state $\rho^{(0)}$ of a Hamiltonian \hat{H} by a variational state $\rho^{(\text{raw})}(\vec{\theta})$, with $\vec{\theta}$ a set of parameters that control the operation of a quantum device. These parameters are tuned by a classical optimization routine to minimize the variational energy $E(\vec{\theta}) = \text{Tr}[\rho^{(\text{raw})}(\vec{\theta})\hat{H}]$. In practice, this is calculated by expanding $\rho^{(\text{raw})}(\vec{\theta})$ and \hat{H} over the N -qubit Pauli basis $\mathbb{P}^N := \{I, X, Y, Z\}^{\otimes N}$,

$$\rho^{(\text{raw})}(\vec{\theta}) = \frac{1}{2^N} \sum_{\hat{P} \in \mathbb{P}^N} \rho_{\hat{P}}^{(\text{raw})}(\vec{\theta}) \hat{P}, \quad \hat{H} = \sum_{\hat{P} \in \mathbb{P}^N} h_{\hat{P}} \hat{P}, \quad (3.1)$$

where the Pauli coefficients are given by $\rho_{\hat{P}}^{(\text{raw})}(\vec{\theta}) = \text{Tr}[\hat{P} \rho^{(\text{raw})}]$. The variational energy may then be calculated as

$$E^{(\text{raw})}(\vec{\theta}) = \sum_{\hat{P} \in \mathbb{P}^N} \rho_{\hat{P}}^{(\text{raw})}(\vec{\theta}) h_{\hat{P}}. \quad (3.2)$$

For example, consider the H_2 molecule studied in this chapter. Mapping the Hamiltonian of this system (in the STO-3G basis) onto four qubits

3.2 Variational quantum eigensolvers for the Hydrogen molecule

via the Bravyi-Kitaev transformation [35] and then further reducing dimensions by projecting out two non-interacting qubits [41] gives

$$\begin{aligned}\hat{H}_{\text{H2}} = & h_{II}II + h_{ZI}ZI + h_{IZ}IZ \\ & + h_{XX}XX + h_{YY}YY + h_{ZZ}ZZ,\end{aligned}\quad (3.3)$$

where coefficients $h_{\hat{P}}$ depend on the interatomic distance R . These coefficients may be determined classically using the OpenFermion [90] and psi4 [89] packages. The Pauli coefficients $\rho_{\hat{P}}^{(\text{raw})}$ of the density matrix $\rho^{(\text{raw})}$ are extracted by repeated preparation and (partial) tomographic measurements of the ansatz state. As one only needs those Pauli coefficients $\rho_{\hat{P}}^{(\text{raw})}$ with non-zero corresponding Hamiltonian coefficients $h_{\hat{P}}$, one need not perform full tomography of $\rho^{(\text{raw})}$. However, in a small-scale experiment, full state tomography of $\rho^{(\text{raw})}$ may still be feasible, and may provide useful information for the purposes of benchmarking. In particular, the fidelity of $\rho^{(\text{raw})}$ to $\rho^{(0)}$,

$$F^{(\text{raw})} = \text{Tr}[\rho^{(\text{raw})}\rho^{(0)}], \quad (3.4)$$

is a more rigorous measure of the ability to prepare the ground state than the energy error,

$$\Delta E^{(\text{raw})} = \text{Tr}\left[\left(\rho^{(\text{raw})} - \rho^{(0)}\right)\hat{H}\right]. \quad (3.5)$$

Error mechanisms such as decoherence pull $\rho^{(\text{raw})}$ away from $\rho^{(0)}$, decreasing F and increasing ΔE .

These errors may be mitigated by using internal symmetries $\hat{S} \in \mathbb{P}^{N-1}$ of the target problem, such as parity checks [52, 53]. These checks project $\rho^{(\text{raw})}$ to a symmetry verified matrix $\rho^{(\text{SV})}$ that lies in the $\langle \hat{S} \rangle = s$ subspace of the symmetry. This projection could be performed via direct measurement of \hat{S} on the quantum device, but one may instead extract the relevant terms of the density matrix $\rho^{(\text{SV})}$ in post-processing:

$$\rho_{\hat{P}}^{(\text{SV})} = \frac{\rho_{\hat{P}}^{(\text{raw})} + s\rho_{\hat{S}\hat{P}}^{(\text{raw})}}{1 + s\rho_{\hat{S}}^{(\text{raw})}}, \quad (3.6)$$

The right-hand side may be obtained by partial tomographic measurement of the ansatz state, with at most twice the number of Pauli coefficients

¹As described in Refs. [52, 53], one does not require \hat{S} to be a Pauli operator, however this makes the SV procedure significantly simpler.

3 Experimental error mitigation via symmetry verification

that need to be measured. This upper bound is not always achieved. For example, the \hat{H}_{H_2} Hamiltonian has a $\hat{S} = ZZ$ symmetry, which maps the non-zero Pauli terms in \hat{H}_{H_2} to other non-zero Pauli terms in \hat{H}_{H_2} . Symmetry verification in this problem then does not require any additional measurements to estimate $E^{(\text{SV})}$ beyond those already required to estimate $E^{(\text{raw})}$. Even when it does require additional measurements, SV remains attractive because it does not require additional quantum hardware or knowledge of the underlying error model. One can show that the SV state $\rho^{(\text{SV})}$ may be equivalently obtained via a variant of the quantum subspace expansion (QSE) [82], suggesting an alternative name of S-QSE [52].

One may further minimize the error in a quantum algorithm by tailoring the quantum circuit or the gates within. In a VQE, one wishes to choose a variational ansatz motivated by the problem itself [41, 103] while minimizing the required quantum hardware [42]. To balance these considerations, we suggest constructing an ansatz from an initial gate-set that is relevant to the problem at hand. For example, in the electronic structure problem, the quantum state is generally an eigenstate of the fermion number. When mapped onto qubits, this often corresponds to a conservation of the total qubit excitation number. Gates such as single-qubit Z rotations, two-qubit C-Phase [104], and two-qubit iSWAP [105] gates preserve this number, making these gates a good universal gate set (within the target subspace [106]) for quantum simulation of electronic structure. In the example of H_2 , the total two-qubit parity (ZZ) is indeed conserved and the ground state at any R may be generated by applying to $|01\rangle$ or $|10\rangle$ an exchange gate

$$U_\theta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & i \sin \theta & 0 \\ 0 & i \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.7)$$

with R -dependent optimal exchange angle θ and a follow-up phase correction on one qubit.

3.3 Experimental error mitigation via symmetry verification

We now experimentally investigate the benefits of SV in the VQE of H_2 using two of three transmon qubits in a circuit QED quantum processor

3.3 Experimental error mitigation via symmetry verification

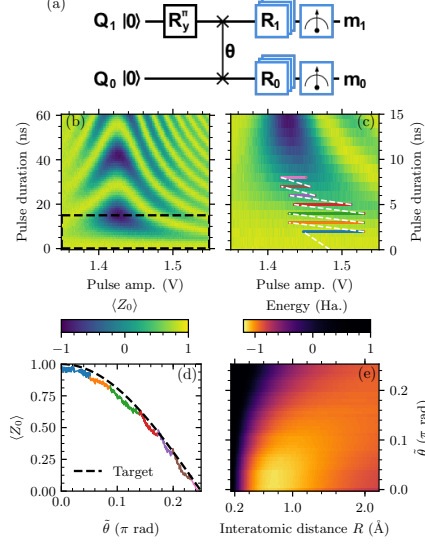


Figure 3.1: Quantum circuit and energy landscape of the variational eigensolver. (a) Quantum circuit for generating and measuring the variational ansatz state. (b) Coherent excitation exchange, produced as Q_0 is fluxed into resonance with Q_1 by a square flux pulse. Pulse amplitude (x axis) parametrizes the frequency to which Q_0 is flux pulsed (~ 1.428 V bringing it on resonance with Q_1). (c) Zoom-in of (b) into the region used in the experiment to control the exchange of population between Q_0 and Q_1 . Colored lines illustrate the hybrid path in pulse duration and amplitude that maps out a finely-adjustable $\tilde{\theta}$ range. (d) Excitation of Q_0 along the hybrid path, showing the matching of the experimentally-defined $\tilde{\theta}$ to the target θ defined in Eq. (3.7) (black dashed curve). Colors [matching (c)] illustrate different pulse durations used in each segment. (e) Landscape of energies $E^{(\text{raw})}(\tilde{\theta}, R)$ as function of the experimentally-defined $\tilde{\theta}$ angle and the interatomic distance R .

3 Experimental error mitigation via symmetry verification

(see details in [44]). The two qubits (Q_0 and Q_1) are coupled by a common bus resonator, and have dedicated microwave lines for single-qubit gating, flux bias lines for local and ns-scale control of their frequency, and dedicated readout resonators coupling to a common feedline for independent readout by frequency multiplexing. We prepare the ansatz state with an efficient circuit [Fig. 3.1(a)] that first excites Q_1 with a π pulse to produce the state $|10\rangle$, and then flux pulses Q_0 into resonance with Q_1 to coherently exchange the excitation population. A sweep of flux-pulse amplitude and duration [Fig. 3.1(b)] reveals the expected chevron pattern that is the hallmark of coherent population exchange between the two qubits, albeit with some asymmetry arising from imperfect compensation of linear distortion in the flux-bias line. To finely control population exchange without being limited by the 1 ns resolution in pulse duration, we stitch together a hybrid path in pulse duration and amplitude. This results in a fine experimental knob $\tilde{\theta}$ (1500 possible settings) that controls population exchange like θ in Eq. (3.7) [Fig. 3.1(c)], although with additional single-qubit phases. The circuit concludes with simultaneous pre-rotation gates on both qubits followed by simultaneous measurement of both qubits, in order to perform tomography of the prepared ansatz state. To fully reconstruct the state, we use an overcomplete set of 36 pre-rotation pairs and extract estimates of the average measurement for each qubit as well as their shot-to-shot correlation using N_{meas} measurements per pre-rotation. Note that single-qubit phase corrections are not required immediately following the exchange gate, as phase rotations can be performed virtually from the fully-reconstructed state.

We now optimize the VQE to approximate the ground-state energy and ground state of H_2 . At each chosen R , we employ the covariance matrix adaptation evolution strategy (CMA-ES) optimization algorithm [107], using $E^{(\text{raw})}$ as cost function and $\tilde{\theta}$ as single variational parameter. The evolutionary strategy optimizes $\tilde{\theta}$ over repeated generations of $N_{\text{pop}} = 10$ samples of $E^{(\text{raw})}(\tilde{\theta})$, each calculated from a raw density matrix $\rho^{(\text{raw})}$ using linear inversion of $N_{\text{meas}} = 10^3$. A typical optimization [Fig. 3.2(a) inset] converges after ~ 20 generations (~ 2 hours). The converged state is finally reconstructed with greater precision, using $N_{\text{meas}} = 10^5$. Figure 2 shows the resulting energy estimate for twelve values of R and the reconstructed optimized state at three such distances. These tomographs show that the optimal solutions are concentrated in the single-excitation subspace of the two qubits, with two-qubit entanglement increasing as a function of R .

Performing the described symmetry verification procedure on the con-

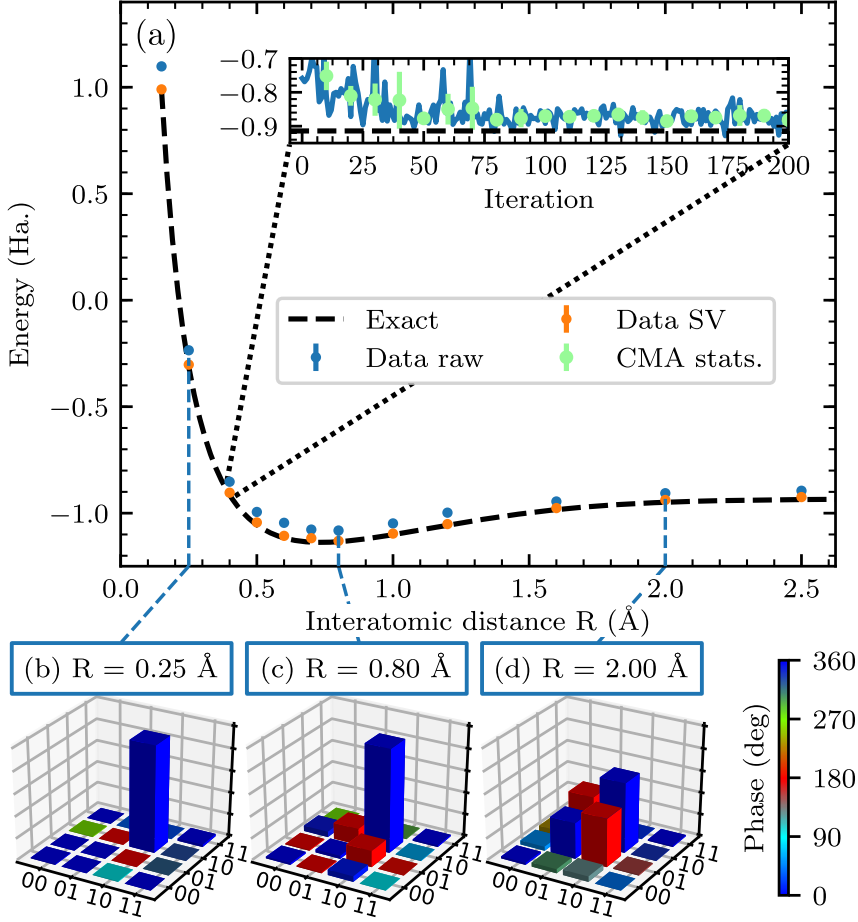


Figure 3.2: Convergence of the VQE algorithm. (a) Experimental VQE estimate of H_2 ground-state energy as a function of interatomic distance R . At each chosen R , we minimize the raw energy $E^{(\text{raw})}$ (blue data points) over the variational parameter $\tilde{\theta}$ using the CMA-ES evolutionary algorithm [107]. Applying SV to the converged solution (orange data points) lowers the energy estimate towards the exact solution (dashed curve). Inset: A typical optimization trace for the convergence of the energy estimate. (b-d) The reconstructed density matrices of the converged states at (b) $R = 0.25$ Å, (c) $R = 0.80$ Å, and (d) $R = 2.00$ Å, showing that the converged states lie mostly in the single-excitation subspace, and that entanglement increases with the interatomic distance R .

3 Experimental error mitigation via symmetry verification

verged states shows improvement across the entire bond-dissociation curve. To quantify the improvement, we focus on the energy error ΔE and the infidelity $1 - F$ to the true ground state, with and without SV (Fig. 3.3). SV reduces the energy error by an average factor ~ 10 and reduces the infidelity by an average factor ~ 9 . In order to quantitatively understand the limits of the VQE optimization, and to clearly pinpoint the origin of the SV improvement, we simulate the experiment via the density-matrix simulator *quantumsim* [92], using an error model built from independently measured experimental parameters [44]. We build the error model incrementally, progressively adding: optimization inaccuracy (the difference between the state ideally produced by the converged θ and the true ground state); dephasing on both qubits (quantified by the measured Ramsey dephasing times T_2^*); relaxation on both qubits (quantified by the measured relaxation times T_1); residual qubit excitations (measured from single-shot histograms with each qubit prepared in $|0\rangle$); and increased dephasing of Q_0 during the exchange gate (quantified by its reduced T_2^* when tuned into the exchange interaction zone). By plotting the errors from each increment of the model, we are able to dissect the observed experimental error into its separate components without [Fig. 3.3(c)] and with [Fig. 3.3(b)] SV. Measured temporal fluctuations of dephasing, relaxation and residual excitation are used to obtain simulation error bars.

The simulation using the full error model shows fairly good matching with experiment for both the ground-state energy error [Figs. 3.3(a,b)] and the state infidelity [Fig. 3.3(c)], without and with SV. The error model dissection shows that the energy error when not using SV is dominated by residual qubit excitations. This is remarkable as the calibrated residual excitations are only 0.25% for Q_0 and 1.34% for Q_1 [44]. The improvement from SV results from the mitigation of errors arising from these residual excitations and from qubit relaxation. This is precisely as expected: these error mechanisms change total qubit excitation number and violate the underlying ZZ symmetry. Using SV changes the dominant error mechanism to optimization inaccuracy. This error could be reduced experimentally by increasing N_{meas} during the optimization, at the cost of increased convergence time. The improvement in state infidelity by SV can be explained along similar lines. We observe some increased deviations between the observed and simulated state infidelity at large R . We attribute these to limitations in our modeling of error during the exchange gate (whose duration increases with R).

3.3 Experimental error mitigation via symmetry verification

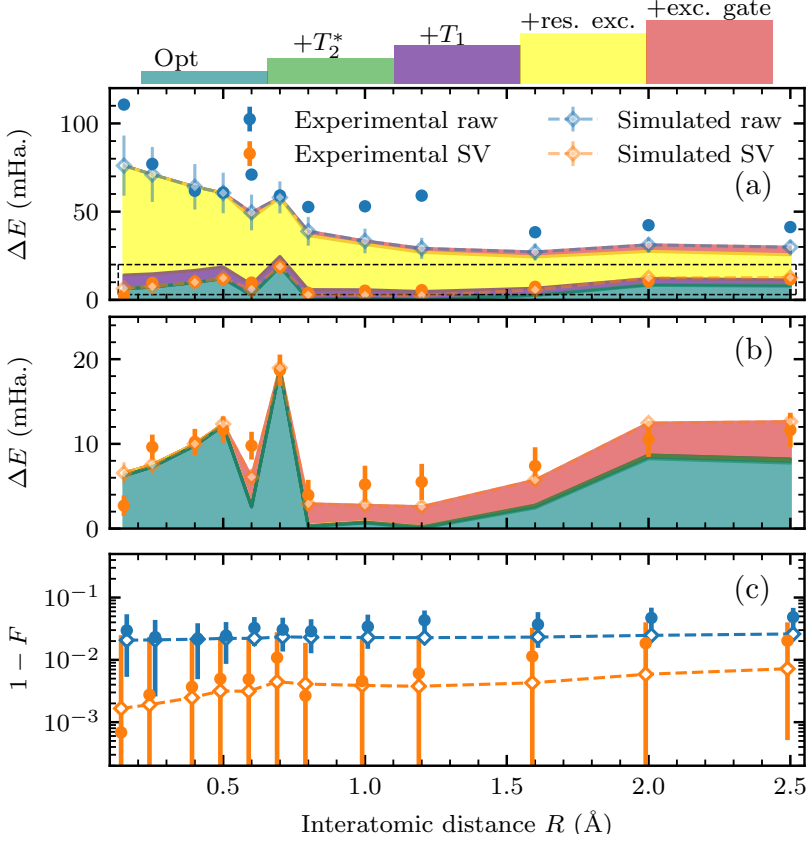


Figure 3.3: Impact of SV in ground-state energy and state fidelity, and dissected error budget. (a) Experimental (solid circles) energy error ΔE without and with SV compared to the result (empty circles) of a full density-matrix simulation using the full error model. The contributions from optimizer inaccuracy, qubit dephasing, qubit relaxation, residual qubit excitations and increased Q_0 dephasing during the exchange gate are shown as shaded regions for the case of no SV applied. Without SV, ΔE is clearly dominated by residual qubit excitation. (b) Zoom-in on experimental and simulated ΔE with SV and corresponding error budget. With SV, the effects of residual excitation and qubit relaxation are successfully mitigated, as predicted in Ref. 52. The remaining energy error is dominated by optimizer inaccuracy. Simulation error bars are obtained by modelling measured fluctuations of T_1 , T_2^* , and residual excitation. (c) Experimental (solid circles) infidelity to the true ground state without and with SV compared to simulation using the full error model (empty circles).

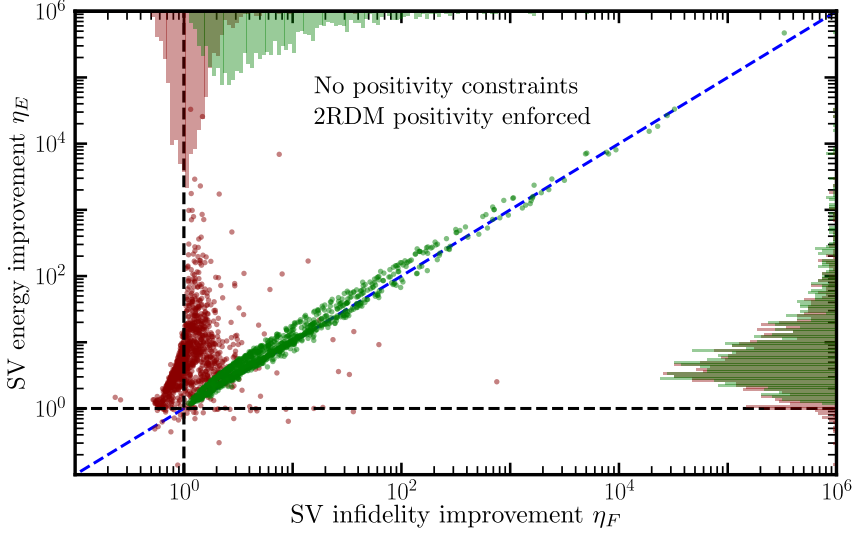


Figure 3.4: Constraining positivity with symmetry verification to mitigate the effect of sampling noise. The experimental data from Fig. 3.3 is split into 100 sample simulations for each R , increasing the sampling noise by a factor of 10 and making it comparable to other sources of experimental error. For each sample, we plot (red) the relative energy error and infidelity [Eq. (3.9)]. Values below 1 (dashed lines) indicate that SV has not provided an improvement, as may be the case when the density matrix has negative eigenvalues. We restore the improvement from SV by constraining the positivity of the 2-reduced density matrix [108] (green). Histograms on the top and right axes show the marginal distribution of the two scatter plots. When the density matrices are constrained to be positive, we observe the points fall along the line $y = x$ (blue dashed line), indicating that SV improves both metrics by the same amount.

3.4 Effect of symmetry verification on positivity constraints

VQEs rely on variational bounding to ensure that the obtained approximation to the ground-state energy is accurate, but this is only guaranteed when the experimental results correspond to a physical state. Our method for calculating the ground-state energy [Eq. (3.1)] independently estimates each Pauli coefficient of the density matrix with error $\propto N_{\text{meas}}^{-1/2}$. Such estimation cannot guarantee a set of Pauli coefficients that could have come from a positive density matrix. This in turn breaks the variational lower bound on the energy estimate, and increases the error in estimates of other properties of the true ground state [108, 109]. As experimental error is reduced, $\rho^{(\text{raw})}$ tends towards a rank-1 density matrix, increasing its chance of being unphysical [109]. Moreover, $\rho^{(\text{SV})}$ is a lower-rank density matrix than $\rho^{(\text{raw})}$ (being projected onto a subspace of the Hilbert space), which implies that unphysicality may be enhanced by SV. The variance in a given term $\rho_{\hat{P}}$ post-SV can be calculated as

$$\text{Var}[\rho_{\hat{P}}^{(\text{SV})}] \approx \frac{3N_{\text{meas}}}{N_{\text{meas}}(1 + \text{Tr}[\rho^{(\text{raw})}\hat{S}])}. \quad (3.8)$$

SV has maximal impact on the quantum state precisely when this denominator is small, so this represents a natural bound for the power of SV as an error mitigation strategy.

The effect of sampling noise may be mitigated somewhat by restricting the fermionic 2-reduced density matrix to be positive (which may be completed in polynomial time) [108]. To investigate the effect of such mitigation, we bin the data used for final tomography of converged states to construct 100 density matrices with $N_{\text{meas}} = 10^3$ at each R , thus increasing the sampling noise by a factor of 10. We wish to study the relative improvement of SV in the two figures of merit, which we quantify as

$$\eta_{\text{E}} = \frac{|\Delta E^{(\text{raw})}|}{|\Delta E^{(\text{SV})}|} \text{ and } \eta_{\text{F}} = \frac{|1 - F^{(\text{raw})}|}{|1 - F^{(\text{SV})}|}, \quad (3.9)$$

when physicality of the raw density matrices is enforced and not. To enforce physicality, we employ a convex optimization routine to find the closest positive semidefinite matrix to the experimentally measured $\rho^{(\text{raw})}$ (closest in the L^2 norm sense on the space induced by the the Pauli basis). We then apply symmetry verification to the post-processed density matrix. Figure 3.4 shows a scatter plot of η_{E} and η_{F} , and relative histograms

of each. Without enforcing physicality, SV makes no significant improvement to the state fidelity, although it almost always improves the energy error. However, when positivity is enforced, SV greatly improves the overlap with the true ground state. We also find that the improvement in the energy from SV is equal to the improvement in fidelity when the starting state is physical, but is relatively uncorrelated when the starting state is not. This makes sense, as the energy gain from SV given a physical matrix comes directly from substituting higher energy states with density on the ground state. It is unclear whether such a strong trend will continue in larger systems without requiring too stringent a positivity constraint. As this is a four-orbital two-electron system, enforcing the positivity of the 2-reduced density matrix enforces positivity on the entire density matrix (which is exponentially difficult in the system size [110]). Testing this scalability is a clear direction for future research ².

3.5 Conclusion

In summary, we have experimentally demonstrated the use of SV to mitigate errors in the VQE of H_2 with two transmon qubits. We implemented an efficient variational ansatz based on an exchange gate producing finely adjustable population transfer in the single-excitation subspace, respecting the ZZ symmetry of the H_2 Hamiltonian. Verification of this symmetry reduced the error of the estimated ground-state energy and the ground state by one order of magnitude on average over the full dissociation curve. A full density-matrix simulation of our system allowed us to budget the contributions from known experimental error mechanisms. We observe that SV mitigates the effect of processes that affect total qubit excitation number, specifically qubit relaxation and residual excitation. Finally, we have investigated the effect of reconstructing density matrices via linear tomographic inversion in the presence of sampling, which voids the guarantee of positivity and in turn the guarantee that SV improves estimation of the ground state. Intriguingly, we observe that when physicality is enforced, the reduction in energy error from SV is directly linked to the increase in fidelity to the ground state. If this observation extends to larger systems, a user can be confident that symmetry-verified Pauli coefficients are accurate for calculations beyond the ground-state energy.

²Note that, for this system, enforcing positivity of the 1-reduced density matrix corresponds to ensuring that all expectation values are bounded between -1 and 1 , and so this does not provide any additional data.

Appendix

3.A Appendix: Tomographic reconstruction and limitations

Tomographic reconstruction was performed with the same technique described in [111]. We provide a brief description here for completeness. For each measurement channel (measurement of Q_1 , measurement of Q_0 , and their correlation), the average measurement outcome is given by $\langle m_i \rangle = \text{Tr}(\hat{M}_i \rho)$, with operator

$$\hat{M}_i = \beta_{II}^i \hat{I}\hat{I} + \beta_{IZ}^i \hat{I}\hat{Z} + \beta_{ZI}^i \hat{Z}\hat{I} + \beta_{ZZ}^i \hat{Z}\hat{Z}, \quad (3.10)$$

and real-valued coefficients β_j^i . Single-shot measurements of Q_0 and Q_1 are 1-bit digitized before correlation and before averaging each of the three channels.

The simultaneously applied measurement pre-rotations R_0 and R_1 consist of the 36 pairs created by drawing each rotation separately from the set $\{I, X_\pi, X_{\pi/2}, Y_{\pi/2}, X_{-\pi/2}, Y_{-\pi/2}\}$. These measurement pre-rotations effectively change the measurement operator to

$$M_i^{k,l} = \text{Tr}\left(R^{k,l,\dagger} \hat{M}_i R^{k,l}\right).$$

There are thus 108 linear equations (36 per channel) linking the averaged measurement to the 15 nontrivial 2-qubit Pauli coefficients (we force $\langle \hat{I}\hat{I} \rangle = 1$). We then extract the Pauli coefficients by performing least-squares linear inversion. Prior to the linear inversion, the measurements are scaled to approximately match the noise in the three channels.

The coefficients β_j^i are obtained from standard calibration measurements. The two qubits are nominally prepared in the four computational states and measured. In total, we perform $7 \times N_{\text{meas}}$ measurements per computational state. The matrix relating the four measurement averages of a channel to the coefficients has elements of the form $\langle \hat{I}\hat{I} \rangle$, $\pm \langle \hat{I}\hat{Z} \rangle$, $\pm \langle \hat{Z}\hat{I} \rangle$ and $\pm \langle \hat{Z}\hat{Z} \rangle$. By taking into account the calibrated residual qubit excitations, which reduce the magnitude of $\langle \hat{I}\hat{Z} \rangle$, $\langle \hat{Z}\hat{I} \rangle$, and $\langle \hat{Z}\hat{Z} \rangle$ from

3 Experimental error mitigation via symmetry verification

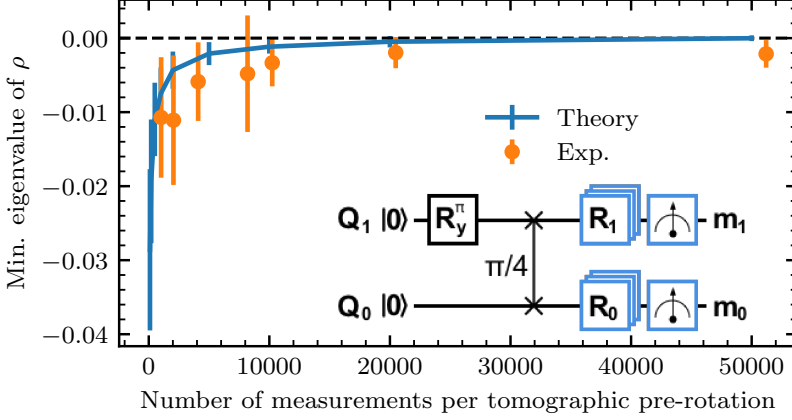


Figure 3.5: Minimal eigenvalue of density matrices obtained from linear tomographic reconstruction with different N_{meas} . Here, the state preparation targets a Bell state using our variational ansatz (inset) with $\theta = \pi/4$ (producing a $\sqrt{\text{iSWAP}}$ gate). Physicality constraints on density matrices restrict their eigenvalues to be non-negative. We observe negative minimum eigenvalues over the entire range of N_{meas} . A *quantumsim* simulation produces a similar trend, asymptotically approaching a physical state by $N_{\text{meas}} \sim 50,000$.

unity, we ensure that the coefficients β_j^i and thus also the operator M_i are not corrupted by residual excitation [112].

Tomography by linear inversion does not ensure physicality of the reconstructed density matrix. We investigate this effect by performing tomography with variable N_{meas} on the state produced by our ansatz with $\theta = \pi/4$ and extracting the minimum eigenvalue of the reconstructed density matrix (Fig. 3.5). A negative minimum eigenvalue manifests unphysicality over the N_{meas} range covered. Our *quantumsim* simulation produces a similar trend, asymptotically approaching a physical state by $N_{\text{meas}} = 5 \times 10^4$. These observations led us to choose $N_{\text{meas}} = 10^5$ for the final state tomography post VQE convergence in Fig. 3, and to further investigate (in Fig. 4) how unphysicality can violate the variational principle, producing reductions in energy from imprecise state reconstruction rather than algorithmic precision.

3.B Appendix: Constraining the positivity of reduced density matrices

Testing whether a N -qubit density matrix ρ is positive is in general QMA-hard [110]. However, if we trace out all but a polynomial number of degrees of freedom of ρ , testing positivity of the resulting reduced density matrix $\rho^{(\text{red})}$ is tractable on classical hardware, and obtaining the closest nearby positive matrix is similarly so. This gives a set of necessary but insufficient physicality conditions for ρ , but enforcing k -local constraints (on a density matrix from a VQE) tends to be sufficient to variationally bound the resulting energies [108]. Following the reduction, we write $\rho^{(\text{red})}$ as a vector over the Pauli basis,

$$\rho^{(\text{red})} = \sum_{\hat{P}} \rho_{\hat{P}}^{(\text{red})} \hat{P}. \quad (3.11)$$

Then, we attempt to find the density matrix $\tilde{\rho}^{(\text{red})}$ closest to $\rho^{(\text{red})}$ in the L^2 -norm

$$\sum_{\hat{P}} \left(\rho_{\hat{P}}^{(\text{red})} - \tilde{\rho}_{\hat{P}}^{(\text{red})} \right)^2, \quad (3.12)$$

subject to the conditions $\tilde{\rho}^{(\text{red})} \succ 0$, and $\tilde{\rho}_I^{(\text{red})} = 1$. This gives a quadratic minimization problem with cone inequality and linear equality constraints, which we solve using interior point methods.

3.C Appendix: Theoretical modeling of the experiment

We use our full-density-matrix simulator *quantumsim* to model the experiment. The error model takes as input parameters the measured values of T_1 , T_2^* and residual excitation for both qubits, and $T_2^{*,\text{red}}$ for Q_0 . We also include the effect of fluctuations on the device parameters by Monte Carlo sampling.

3.C.1 Numerical simulations

The simulations are performed by extracting the full-density-matrix $\rho^{(\text{sim})}$ at the end of the circuit. We use the converged value of $\tilde{\theta}$ at each R to generate the quantum state and extract the Pauli coefficients $\rho_{\hat{P}}^{(\text{sim})}(\tilde{\theta}) =$

3 Experimental error mitigation via symmetry verification

$\text{Tr}[\hat{P}\rho^{(\text{sim})}]$. We add sampling noise to each coefficient, drawn from a zero-mean Gaussian distribution with variance $(1 + \rho_{\hat{P}}^{(\text{sim})})(1 - \rho_{\hat{P}}^{(\text{sim})})/N_{\text{meas}}$, where $N_{\text{meas}} = 4 \times 10^5$. Note that this is greater than the number of measurements per tomographic prerotation in the experiment, as data from multiple tomographic prerotations is used to estimate each Pauli coefficient. To account for fluctuations on the device parameters T_1 , T_2^* , and residual excitations, we average over 10^4 simulations for every R . For each simulation, we draw parameters from independent normal distributions using values in Table I of the supplementary material of Ref. [44]. As the dephasing noise T_2^* depends on T_1 ,

$$\frac{1}{T_2^*} = \frac{1}{T_\phi} + \frac{1}{2T_1}, \quad (3.13)$$

it is more appropriate to sample the pure dephasing rate T_ϕ independently. We calculate the T_ϕ mean and variance (\bar{T}_ϕ , $\text{Var}[\bar{T}_\phi]$) from T_2^* ,

$$\begin{aligned} \bar{T}_\phi &= \frac{1}{\frac{1}{T_2^*} - \frac{1}{2T_1}}, \\ \text{Var}[\bar{T}_\phi] &= \bar{T}_\phi^2 (T_2^*)^{-2} \left[\text{Var}[T_2^*] - \frac{\text{Var}[T_1]}{2T_1^2 (T_2^*)^{-2}} \right]. \end{aligned} \quad (3.14)$$

From the 10^4 simulations we obtain 95%-confidence error bars for ΔE and F as twice the population standard deviation.

3.C.2 Exchange gate

Due to quasi-static flux noise, the angle of the unitary exchange gate (Eq. 7) differs between subsequent applications. Assuming that these fluctuations are fast on the scale of the 2 hour optimization, this may be simulated by integrating over the range of applied gates, resulting in an incoherent noise model. As the gate is not repeatedly applied during a single-shot experiment, this incoherent approximation does not lead to an error in the final result. To perform the integration, we convert our unitary U_θ into a Pauli Transfer Matrix representation (PTM) [113, 114]:

$$[R_\theta]_{i,j} = \frac{1}{2} \text{Tr}[\hat{P}_i U_\theta \hat{P}_j U_\theta^\dagger], \quad \hat{P}_i, \hat{P}_j \in \mathbb{P}^N, \quad (3.15)$$

which may then be integrated over a probability distribution in the deviation δ from the target angle θ :

$$[\tilde{R}]_{i,j} = \int d\delta \, p(\delta) [R_{\theta+\delta}]_{i,j}. \quad (3.16)$$

3.C Appendix: Theoretical modeling of the experiment

We choose for $p(\delta)$ a Gaussian distribution: $p(\delta) = e^{-\frac{\delta^2}{2\sigma^2}}$. In order to obtain the distribution width σ^2 , we note that the same effect causes single-qubit dephasing of Q_0 when fluxed to the exchange point when Q_1 is fluxed away. We may thus estimate σ as

$$\sigma^2 = 1 - e^{-\frac{t_{\text{int}}}{T_2^{*,\text{red}}}}, \quad (3.17)$$

where t_{int} is the exchange gate duration and $T_2^{*,\text{red}}$ the dephasing time of Q_0 at the exchange point (with Q_1 fluxed away). The final gate simulation also includes the effect of amplitude damping on both qubits, and the dephasing of Q_1 at the sweet spot as discrete error channels of duration $\frac{t_{\text{int}}}{2}$ on either side of the gate \tilde{R} .

4 Nearly-optimal measurement scheduling for partial tomography of quantum states

4.1 Introduction

The advent of variational methods, most notably the variational quantum eigensolver [39, 87], inspires hope that useful contributions to our understanding of strongly-correlated physical and chemical systems might be achievable in pre-error corrected quantum devices [41]. Following this initial work, much progress has gone into lowering the coherence requirements of variational methods [115], calculating system properties beyond ground state energies [82, 116, 117], and experimental implementation [42, 83, 98, 118]. However, extracting data from an exponentially complex quantum state is a critical bottleneck for such applications. Initial estimates for the number of measurements required to accurately approximate the energy of a variationally generated quantum state were astronomically large, with bounds for quantum chemistry applications as high as 10^{13} for a system of 112 spin-orbitals in minimal basis [119]. Although improving these results is critical for the scalability of variational approaches, until recently, little effort has been devoted to lowering the number of measurements needed.

A common way to estimate the energy of a quantum state during a variational quantum algorithm is to perform partial tomography [87] on a set of observables which comprise a k -body reduced density matrix (k -RDM)¹ [108]. For instance, the fermionic 2-RDM allows one to calculate such properties as energy [108], energy gradients [117, 120], and multipole

¹While k -body qubit RDMs catalogue correlations between k qubits, k -body fermion RDMs catalogue correlations between k fermions, and thus involve $2k$ fermionic modes; e.g., the elements of the fermionic 2-RDM are the expectation values $\langle c_p^\dagger c_q^\dagger c_r c_s \rangle$.

moments [121] of electronic systems in quantum chemistry and condensed matter problems, and further enables techniques for relaxing orbitals to reduce basis error [82, 122]. By contrast, the qubit 2-RDM plays a vital role in spin systems, as it contains static spin correlation functions that can be used to predict phases and phase transitions [123], and separately contains information to characterize the entanglement generated on a quantum device [124]. Reduced density matrices thus offer a useful and tractable description of an otherwise complex quantum state.

Partial tomography to estimate a reduced density matrix may be performed by separating the observables to be tomographed into sets of mutually-commuting operators. By virtue of their commutation, a unique measurement scheme may be found to measure all operators in a single set simultaneously. Subsequent measurement of non-commuting operators requires re-preparation of the quantum state, so the time required to estimate a target RDM is proportional to the number of unique measurement circuits. Minimizing this number is crucial for the scalability of variational algorithms, as a naive approach requires $\mathcal{O}(N^4)$ unique measurement circuits, which is impractical. Recent work has focused on mapping this problem to that of clique finding or colouring of a graph [125], and applying approximate algorithms to these known NP-hard problems [126]. This achieves constant or empirically determined linear scaling improvements over an approach that measures each term individually [125, 127–130]. However, the commutation relations between local qubit or local fermionic operators has significant regularity not utilised in naive graph-theoretic algorithms. Leveraging this regularity is critical to optimizing and proving bounds on the difficulty of tomography of quantum states.

ref.	partitioning method	circuits based on	partitions	gate count	depth	classical cost	connect.	RDM	sym.
[87]	comm. Pauli heuristic	-	$\mathcal{O}(N^4)$	-	-	$\mathcal{O}(1)$	-	-	-
[42]	compatible Pauli heuristic	single rotations	$\mathcal{O}(N^4)$	N	1	$\mathcal{O}(1)$	linear	yes	no
[108]	n -representability constraints	single rotations	$\mathcal{O}(N^4)$	N	1	$\mathcal{O}(1)$	linear	no	no
[131]	mean-field partitioning	fast feed-forward	$\mathcal{O}(N^4)$	$\mathcal{O}(N)$	$\mathcal{O}(N)$	$\mathcal{O}(N^3)$	full	no	no
[125]	compatible Pauli clique cover	single rotations	$\mathcal{O}(N^4)$	N	1	$\mathcal{O}(N^8 - N^{12})$	linear	yes	no
[127]	comm. Pauli graph coloring	stabilizer formalism	$\mathcal{O}(N^3)$	-	-	-	full	no	no
[130]	a-comm. Pauli clique cover	Pauli evolutions	$\mathcal{O}(N^3)$	$\mathcal{O}(N^2 \log N)$	-	$\mathcal{O}(N^8 - N^{12})$	full	no	no
[128]	comm. Pauli clique cover	symplectic subspaces	$\mathcal{O}(N^3)$	$\mathcal{O}(N^2 / \log N)$	-	$\mathcal{O}(N^8 - N^{12})$	full	no	no
[132]	basis rotation grouping	Givens rotations	$\mathcal{O}(N)$	$N^2/4$	$N/2$	$\mathcal{O}(N^4 \log(N))$	linear	no	Num.
[129]	comm. Pauli clique cover	stabilizer formalism	$\mathcal{O}(N^3)$	$\mathcal{O}(N^2)$	-	$\mathcal{O}(N^8 - N^{12})$	full	yes	no
[133]	comm. Pauli clique cover	stabilizer formalism	$\mathcal{O}(N^3)$	$\mathcal{O}(N^2)$	-	$\mathcal{O}(N^8 - N^{12})$	full	yes	no
[134]	a-comm. Pauli clique cover	Pauli evolutions	$\mathcal{O}(N^3)$	$\mathcal{O}(N^{\frac{3}{2}} \log N)$	-	$\mathcal{O}(N^8 - N^{12})$	full	no	no
here	comm. Majorana pairs	Majorana swaps	$\mathcal{O}(N^2)$	$N^2/2$	N	$\mathcal{O}(N^2)$	linear	yes	Par.
here	a-comm. Majoranas	Majorana rotations	$\mathcal{O}(N^4/\omega)$	ω	$\omega/2$	$\mathcal{O}(\frac{N^4}{\omega})$	linear	no	Par.
here	2-RDM partition bound	-	$\Omega(N^2)$	-	-	-	-	-	-
here	a-comm. clique bound	-	$\Omega(N^3)$	-	-	-	-	-	-

Table 4.1: A history of ideas reducing the measurements required for estimating the energy of arbitrary basis chemistry Hamiltonians with the variational quantum eigensolver. Here N represents the number of spin-orbitals in the basis, and ω is defined in the text. We use the shorthand “comm”. and “a-comm.” for commuting and anti-commuting respectively. The “partitions” column counts the number of unique circuits required to generate at least one sample of each term in the Hamiltonian. Gate counts and depths are given in terms of arbitrary 1- or 2-qubit gates restricted to the geometry of 2-qubit gates specified in the connectivity column. The “classical cost” column reports the overhead to determine the partitions for a given Hamiltonian. In the “RDM” column we report whether the technique is able to measure the entire fermionic 2-RDM with the stated scaling, or just a single expectation value (e.g., of the Hamiltonian). In the “sym.” column we report whether any symmetries of the system commute with all measurements made - this allows for simultaneous measurement, enabling strategies for error mitigation by post-selection at zero additional cost.

In this chapter, we provide schemes for the estimation of fermionic and qubit k -RDMs that minimize the number of unique measurement circuits required, significantly decreasing the time required for partial state tomography over prior art. We demonstrate a scheme to estimate qubit k -RDMs in an N -qubit system in time $\mathcal{O}(3^k \log^{k-1} N)^2$, achieving an exponential increase over prior art. We then prove a lower bound of $\Omega(N^k)$ on the number of state preparations required to estimate fermionic k -RDMs (such as those of interest in the electronic structure problem) using Clifford circuits (including the addition of ancilla qubits prepared in the $|0\rangle$ state) and measurement in the computational basis. We describe protocols to achieve this bound for $k \leq 2$. We detail measurement circuits for these protocols with circuit depths of $\mathcal{O}(N)$ and gate counts of $\mathcal{O}(N^2)$ (requiring only linear connectivity), that additionally allow for error mitigation by symmetry verification [52, 53]. Finally, we detail an alternative scheme to measure arbitrary linear combinations of fermionic k -RDM elements, based on finding large sets of anti-commuting operators. This requires $\mathcal{O}(N^4/\omega)$ measurements, but has a measurement circuit gate count of only $\mathcal{O}(\omega)$ on a linear array, for a free parameter $\omega < N$.

In Tab. 4.1, we provide a history of previous art in optimizing measurement schemes for the electronic structure problem, and include the new results found in this work. We further include the lower bounds for the number of partitions required for anti-commuting and commuting clique cover approaches that were presented in this chapter.

4.2 Background

Physical systems are characterized by local observables. However, the notion of locality depends on the exchange statistics of the system in question. In an N -qubit system, data about all k -local operators within a state ρ is given by the (qubit) k -reduced density matrices, or k -RDMs [108]

$${}^k\rho_{i_1, \dots, i_k} = \text{Trace}_{j \neq i_1, \dots, i_k} [\rho]. \quad (4.1)$$

Here, the trace is over all other qubits in the system. To estimate ${}^k\rho$, we need to estimate expectation values of all tensor products of k single-qubit Pauli operators $P_i \in \{X, Y, Z\}$; we call such tensor products ' k -qubit' operators. In an N -fermion system, data about all k -body operators is contained in the (fermionic) k -body reduced density matrices, which are

²Here and throughout this chapter all logarithms are base two.

4.3 Near-optimal measurement schemes for local qubit and fermion operators

obtained from ρ by integrating out all but the first k particles [108]

$${}^kD = \text{Trace}_{k+1,\dots,N}[\rho]. \quad (4.2)$$

Estimating kD requires estimating the expectation values of all products of k fermionic creation operators c_j^\dagger with k fermionic annihilation operators c_j . For instance, the 2-RDM catalogues all 4-index expectation values of the form $\langle c_p^\dagger c_q^\dagger c_r c_s \rangle$. One can equivalently describe fermionic systems in the Majorana basis,

$$\gamma_{2j} = c_j + c_j^\dagger, \quad \gamma_{2j+1} = i(c_j^\dagger - c_j), \quad (4.3)$$

in which case the fermionic k -RDM may be computed from the expectation values of $2k$ Majorana terms γ_j (e.g. the 2-RDM is computed from expectation of Majorana operators of the form $\langle \gamma_i \gamma_j \gamma_k \gamma_l \rangle$). We call such products $2k$ -Majorana operators for short.

The expectation values of the above operators may be estimated with standard error ϵ by $\mathcal{O}(\epsilon^{-\epsilon})$ repeated preparation of ρ and direct measurement of the operator. This estimation may be performed in parallel for any number of k -qubit operators \hat{P}_i or $2k$ -Majorana operators \hat{G}_i , as long as all operators to be measured in parallel commute. This suggests that the speed of a ‘partial state tomography’ protocol that estimates expectation values of all k -qubit or $2k$ -Majorana operators by splitting them into a set of ‘commuting cliques’ (sets where all elements commute) is proportional to the number of cliques required. In this chapter we focus on optimizing partial state tomography schemes by minimizing this number. Necessarily, our approach will be different for qubit systems (where two spatially separated operators always commute) compared to fermionic systems (where this is often not the case).

4.3 Near-optimal measurement schemes for local qubit and fermion operators

Partial state tomography of qubit k -RDMs can be efficiently performed by rotating individual qubits into the X , Y , or Z basis and reading them out. These rotations define a ‘Pauli word’ $W \in \{X, Y, Z\}^N$, where W_i is the choice of basis for the i th qubit. Repeated sampling of W allows for the estimation of expectation values of any Pauli operator P that is a tensor product of some of the W_i — we say these operators are contained within the word. (The set of all such P is the clique corresponding to W with the property that each P is qubit-wise commuting with the rest of

operators in the word W .) To estimate the k -qubit RDM in this manner, we need to construct a set of words that contain all k -local operators. For $k = 2$, it is sufficient to find a set of words $W \in \{A_0, A_1\}^N$ such that each pair of qubits differ in their choice of letter in at least one word. Then, permuting over $A_0 = X, Y, Z$, and separately $A_1 = X, Y, Z$, extends the set to contain all 2-qubit operators. Such a set can be found via a binary partitioning scheme, for a total of $6\lceil \log N \rceil + 3$ cliques (see App. 4.A for details). This scheme may be further extended to arbitrary $k > 2$ with a complexity $O(3^k \log^{k-1}(N))$. The (classical) computational complexity to generate each word is at most $O(\log(N))$, and $O(N)$ to assign each qubit, making the classical computational cost to generate the set of measurements $\mathcal{O}(e^k N \log^k N)$, which is acceptably small for even tens of thousands of qubits. We have added code to generate the full measurement protocol to the Openfermion software package [90].

Fermionic k -RDMs require significantly more measurements to tomograph than their qubit counterparts, as many more operators anti-commute. In a N -fermion system, the total number of $2k$ -Majorana operators is $\binom{2N}{2k}$, while the size of a commuting clique of $2k$ -Majorana operators may be upper-bounded by $\binom{N}{k}$ in the $N \gg k$ limit (see App. 4.B). As fermionic k -RDMs contain expectation values of $2k$ -Majorana operators, the number of cliques required to estimate all elements in the fermionic k -RDM scales as

$$\binom{2N}{2k} / \binom{N}{k} \sim N^k. \quad (4.4)$$

In terms of the resources requirement to estimate a fermionic k -RDM, this directly implies

Theorem 1. *The number of preparations of an arbitrary N -fermion quantum state ρ required to estimate all terms in the fermionic k -RDM to within an error ϵ , via Clifford operations (including addition of ancilla qubits prepared in the $|0\rangle$ state), and measurement in the computational basis, is bounded below in the worst case as $\Omega(\epsilon^{-2} N^k)$.*

Proof details may be found in App. 4.H. In particular, estimating the fermionic 1-RDM requires repeated preparation of ρ and measurement over at least $2N - 1$ unique commuting cliques, and estimating the fermionic 2-RDM requires repeat preparation and measurement over a number of cliques at least

$$\frac{4}{3}N^2 - \frac{8}{3}N + 1. \quad (4.5)$$

4.3 Near-optimal measurement schemes for local qubit and fermion operators

Maximally-sized cliques of commuting $2k$ -Majorana operators may be achieved via a pairing scheme. If we pair the $2N$ individual 1-Majorana operators into N pairs $\{\gamma_i\gamma_j\}$, the corresponding set of operators $i\gamma_i\gamma_j$ forms a commuting set. Any product of k of these pairs will also commute, so the set of all combinations of k pairs is a commuting clique of exactly $\binom{N}{k} 2k$ -Majoranas. We say that the $2k$ -Majorana operators are contained within the pairing. Curiously, each pairing saturates the bounds found in App. 4.B for the number of mutually commuting $2k$ -Majorana operators in a N -fermion system, and thus this scheme is optimal in the number of $2k$ -Majorana operators targeted per measurement circuit. However, as one $2k$ -Majorana operator may be contained in multiple pairings, it remains to find a scheme to contain all $2k$ -Majorana operators in the minimum number of pairings. For the 1-RDM, it is possible to reach the lower bound of $2N - 1$ cliques by a binary partition scheme, which we detail in App. 4.C. In the 2-RDM case, we have been able to achieve $\frac{10}{3}N^2$ cliques (also detailed in App. 4.C) by a divide and conquer approach. It remains an open question whether the factor $5/2$ between our scheme and the lower bound (Eq. 4.5) can be improved, either by better bounding or a different scheme.

Simultaneous estimation of the expectation value of each observable may be achieved by repeatedly preparing and measuring states in the $i\gamma_i\gamma_j$ basis for all paired γ_i, γ_j in the clique. Measuring the system in this basis is non-trivial and depends on the encoding of the fermionic Hamiltonian onto the quantum device. However, for almost all encodings this requires simply permuting the Majorana labels, which may be achieved by a single-particle basis rotation using Clifford gates (see App. 4.F). This implies that the circuit depth should be no worse than $O(N)$, and will not require T-gates in a fault-tolerant setting. Furthermore, in many cases the measurement circuit should be able to be compiled into the state preparation circuit, reducing its cost further.

Symmetry constraints on a system (i.e. unitary or antiunitary operators S that commute with the Hamiltonian H) force certain RDM terms to be 0 for any eigenstates of the system. For example, when a real Hamiltonian is written in terms of Majorana operators (using Eq. 4.3), it must contain an even number of odd-index 1-Majorana operators, and expectation values of terms not satisfying this constraint on eigenstates will be set to 0. More generally, if a symmetry is a Pauli word W_{symmetry} such that $W_{\text{symmetry}}^2 = 1$, then it will divide the set of all Majorana terms into those which commute with W_{symmetry} and those which anti-commute; products of odd numbers of anti-commuting terms will have zero expectation value on eigenstates of the system. Given n such independent symmetries, each

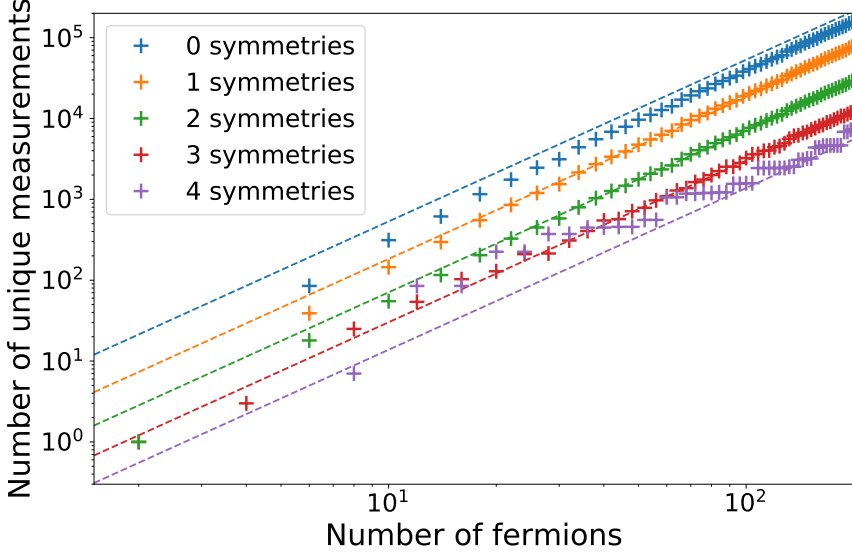


Figure 4.1: Scaling of our Majorana partitioning scheme in the presence of between 0 and 4 symmetry constraints on the system. Dashed lines are from Eq. 4.6

of which commute with half of all 1-Majorana operators (which is typical), we are able to contain all elements of the fermionic 2-RDM in a number of cliques scaling to first order as

$$N^2 \left(\frac{10}{3} 4^{-N_{\text{sym}}} + 2^{1-N_{\text{sym}}} \right). \quad (4.6)$$

(See App. 4.D for details.) In Fig. 4.1, we show the result of an implementation of our scheme for different numbers of symmetries at small N , and see quick convergence to this leading-order approximation for up to 4 symmetries (typical numbers for quantum chemistry problems). Code to generate this measurement scheme has been added to the Openfermion package [90].

4.4 Measuring anti-commuting linear combinations of local fermionic operators

Products of Majorana and Pauli operators have the special property that any two either strictly commute or strictly anti-commute. This raises the question of whether there is any use in finding cliques of mutually anti-commuting Pauli operators. Such cliques may be found in abundance when working with Majoranas — e.g. for fixed $0 \leq j, k, l \leq 2N$, the set $A_{j,k,l} = \{\gamma_i \gamma_j \gamma_k \gamma_l\}$ is a clique of $2N - 3$ mutually anti-commuting operators. Curiously, it turns out that asymptotically larger anti-commuting cliques are not possible - the largest set of mutually anti-commuting Pauli or Majorana operators contains at most $2N + 1$ terms (see App. 4.G for a proof). The number of anti-commuting cliques required to contain all 4-Majorana operators is thus bounded below by $\Omega(N^3)$, matching the numerical observations of [130].

Although sampling each term in an anti-commuting clique A of size L requires $\mathcal{O}(L)$ state preparations, it is possible to measure a (real) linear combination $\mathcal{O} = \sum_{i=1}^L c_i P_i$ of clique elements in a single shot. Since all elements of $A_{j,k,l}$ share three of the same four indices, here we can associated each P_i in the sum over the elements of $A_{j,k,l}$ with the Majorana $P_i = \gamma_i \gamma_j \gamma_k \gamma_l$. Given that $\tilde{O} = (\sum_{i=1}^L c_i^2)^{-1/2} \mathcal{O}$ looks like a Pauli operator ($\tilde{O}^\dagger = \tilde{O}$, $\text{Trace}[\tilde{O}] = 0$), and smells like a Pauli operator ($\tilde{O}^2 = 1$), it can be unitarily transformed to a Pauli operator of our choosing. In App. 4.F, we show that for systems encoded via the Jordan-Wigner transformation, this unitary transformation may be achieved with a circuit depth of only $N - 2 + \mathcal{O}(1)$ 2-qubit gates. It is possible to reduce the depth further by removing Majoranas from the set — if we restrict ourselves to subsets of ω elements of $A_{j,k,l}$, the measurement circuit will have ω gates and be depth ω , but $\mathcal{O}(N^4/\omega)$ such sets will be needed to estimate arbitrary linear combinations of 4-Majorana operators. This makes this scheme very attractive in the near-term, where complicated measurement circuits may be prohibited by low coherence times in NISQ devices.

4.5 Conclusion

Experimental quantum devices are already reaching the stage where the time required for partial state tomography is prohibitive without opti-

mized scheduling of measurements. This makes work developing new and more-optimal schemes for partial tomography of quantum states exceedingly timely. In this chapter, we have shown that a binary partition strategy allows one to sample all k -local qubit operators in a N -qubit system in $\text{poly-log}(N)$ time, reaching an exponential improvement over previous art. By contrast, in fermionic systems we have found a lower bound on the number of unique measurement circuits required to directly sample all k -local operators of $\Omega(N^{\lceil k/2 \rceil})$, an exponential separation. We have developed schemes to achieve this lower bound for $k = 2$ and $k = 4$, allowing estimation of the entire fermionic 2-RDM to constant error in $\mathcal{O}(N^2)$ time. Additionally, we have demonstrated that one can leverage the anti-commuting structure of fermionic systems by constructing such sets of size $1 \leq \omega \leq N$ to measure all 4-Majorana operators in $\mathcal{O}(N^4/\omega)$ time with a gate count and circuit depth of only ω , allowing one to trade off an decrease in coherence time requirements for an increase in the number of measurements required.

Appendix

4.A Appendix: Schemes for partial state tomography of qubit k -RDMs

In this section, we develop methods to minimize the measurement cost for partial state tomography of qubit k -RDMs by minimizing the number of commuting cliques needed to contain all k -qubit operators. To do so, we associate a ‘Pauli word’ $W \in \{X, Y, Z\}^N$ to each clique: by measuring the i th qubit in the W_i basis, we measure every tensor product of the individual Pauli operators W_i . Thus, the clique associated to W contains all k -qubit operators that are tensor products of the W_i — we say these operators are ‘contained’ within the word. We then wish to find the smallest possible set of words such that every k -qubit operator is contained within at least one word.

We construct such a set through a k -ary partitioning scheme, which we first demonstrate for $k = 2$. As motivation, consider that the set of 9 words (with $A, B = X, Y, Z$)

$$W_i^{(A,B)} = \begin{cases} A & \text{if } i < N/2 \\ B & \text{if } i \geq N/2 \end{cases}, \quad (4.7)$$

contains all 2-qubit operators that act on qubits $j < N/2$ and $k \geq N/2$. We may generalize this to obtain all other 2-qubit operators by finding a set of binary partitions $S_{n,0} \cup S_{n,1} = \{1, \dots, N\}$ such that for any pair $0 \leq i \neq j \leq N$ there exists n, a such that $i \in S_{n,a}$, $j \in S_{n,1-a}$. Let us define $L = \lceil \log N \rceil$, and write each qubit index i in a binary representation, $i = [i]_{L-1}[i]_{L-2} \dots [i]_1[i]_0$. Then, for $n = 0, \dots, L-1$ we define

$$i \in S_{n,a} \text{ if } [i]_n = a. \quad (4.8)$$

All $0 \leq i \neq j \leq N$ differ by at least one of their first L binary digits (as shown in Fig. 4.2(a)), so the set of words $W_n^{(A,B)}$, constructed as

$$[W_n^{(A,B)}]_i = \begin{cases} A & \text{if } i \in S_{n,a} \\ B & \text{if } i \in S_{n,1-a}, \end{cases} \quad (4.9)$$

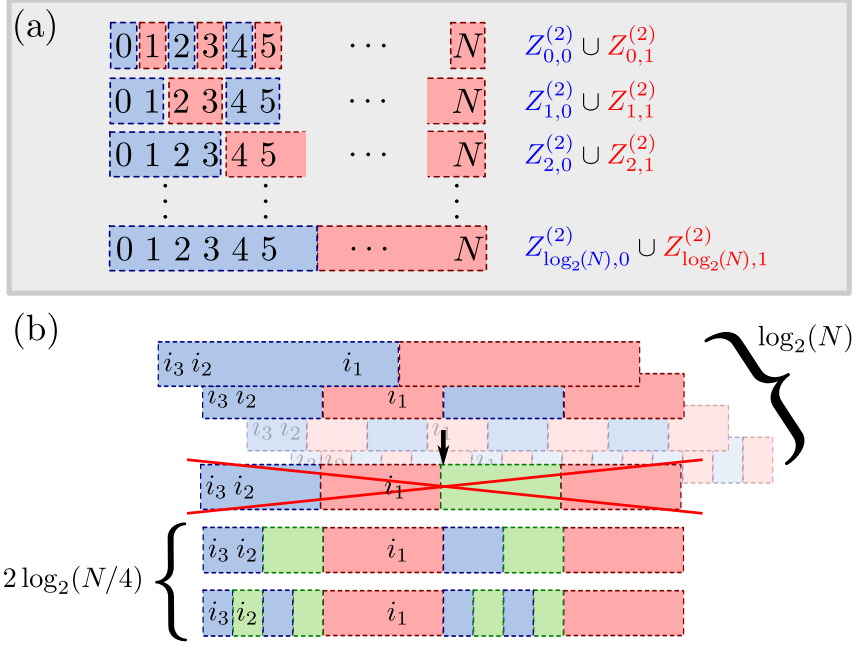


Figure 4.2: Schematics of the binary partition strategy described in text. (a) Scheme to construct $\mathcal{O}(\log N)$ cliques that contain all 2-qubit operators. (b) Extension of the top scheme to a set of $\mathcal{O}(\log^2 N)$ cliques that contain all 3-qubit operators.

defines a set of cliques that contain all 2-qubit operators. As $W_{n,i}^{(A,A)}$ is the same word for every n we need only choose this word once and so the number of cliques may be reduced to $6L + 3$.

To see how the above may be extended to $k > 2$, let us consider $k = 3$. We wish to find 3-ary partitions $\cup_{a=1}^3 S_{n,a} = \{1, \dots, N\}$ that, given any set i_1, i_2, i_3 , we can find some index n for which $i_a \in S_{n,a}$ (allowing for permutation of the i_a). Then, by running over all combinations of X, Y, Z on the three parts of each partition, we will obtain a set of words that contain all 3-qubit operators. We illustrate a scheme that achieves this Fig. 4.2(b). We iterate first over $n = 1, \dots, L$, and find the largest n such that i_1, i_2 and i_3 are split into two subsets by a binary partition. (i.e. where $S_{n,a} \cap \{i_1, i_2, i_3\}$ is non-empty for $a = 0$ and $a = 1$). This implies that two of the indices lie in one part, and one in the other. Without loss of generality, let us assume $i_1 \in S_{n,1}$ and $i_2, i_3 \in S_{n,0}$ (following Fig. 4.2).

4.B Appendix: Upper bounds on the size of commuting cliques of Majorana operators.

It now suffices to find a set of partitions for $S_{n,0}$ so that we guarantee i_2 and i_3 are split in one such partition. We could imagine repeating the binary partition scheme over all $S_{n,0}$; i.e. generating the $\log N$ sets $S_{n,0} \cap S_{n',a}$. However, we can do better than this. As i_1, i_2 and i_3 are not split in any binary partition $S_{n',0}, S_{n',1}$ with $n' > n$, i_2 and i_3 must be in a contiguous block of length $1/2^n$ within $S_{n,0}$. This means that we need only iterate over $n' = 0, \dots, n-1$. We must also iterate over the same number of partitions of $S_{n,1}$, and so the total number of partitions we require is

$$2 \sum_{n=0}^{L-1} n = (L-1)(L-2). \quad (4.10)$$

The above generalizes relatively easily to $k > 2$. Given a set $I = \{i_1, \dots, i_k\}$, we find the binary partition $S_{n,0}, S_{n,1}$ with the largest n that splits I into non-empty sets $I_0 = I \cap S_{n,0}$ and $I_1 = I \cap S_{n,1}$. Then, we iterate over $|I_0|$ -ary partitions of the contiguous blocks of $S_{n,0}$ and the $|I_1|$ -ary blocks of $S_{n,1}$. In total there are $k-1$ possible ways of dividing I (up to permutations of the elements). This implies that at each n we have to iterate over $k-1$ different sub-partitioning possibilities, making the leading-order contribution to the number of cliques

$$(k-1) \sum_{n=0}^{L-1} n^{k-2} \sim L^{(k-1)}, \quad (4.11)$$

and the total number of cliques $\mathcal{O}(3^k \log^{k-1} N)$.

4.B Appendix: Upper bounds on the size of commuting cliques of Majorana operators.

In this appendix, we detail the bounds on the size of commuting cliques of Majorana operators. Let us call the largest number of mutually-commuting k -Majoranas that are a product of l unique terms (i.e. l unique 1-Majoranas) M_l^k . (For an N -fermion system, we will eventually be interested in the case where $l = 2N$.) We wish to bound this number M_l^k by induction. All 1-Majorana operators anti-commute, so $M_l^k = 1$. Then, let us consider the situation where k is even and when k is odd separately. Suppose we have a clique of M_l^k k -Majorana operators with k even. As there are only l unique terms, and these k -Majoranas contain

kM_l^k individual terms each, there must be a clique of $\lceil kM_l^k/l \rceil$ of these operators that share a single term γ_0 . We may write each such operator in the form $\pm\gamma_0\Gamma_i$, where Γ_i . As $[\gamma_0\Gamma_i, \gamma_0\Gamma_j] = 0$ if and only if $[\Gamma_i, \Gamma_j] = 0$, this gives a clique of kM_l^k/l commuting $(k-1)$ -Majorana operators on $l-1$ unique terms, so we must have

$$\left\lceil \frac{kM_l^k}{l} \right\rceil \leq M_{l-1}^{k-1}, \quad k \text{ even.} \quad (4.12)$$

Now, consider the case where k is odd, let us again assume we have a clique of $M^{(k)}$ commuting k -Majoranas. Two products of Majorana operators anticommute unless they share at least one term in common, so let us choose one k -Majorana Γ in our set; each k -Majorana must have at least one of the k terms in Γ , so at least one such term is shared between $\lceil M^{(k)}/k \rceil$ Majoranas in our set. Removing this term gives a clique of $\lceil M^{(k)}/k \rceil$ $(k-1)$ -Majorana operators on $l-1$ unique terms, and so we have

$$\left\lceil \frac{M^k}{k} \right\rceil \leq M_{l-1}^{k-1}, \quad k \text{ odd.} \quad (4.13)$$

These equations may be solved inductively to lowest-order in k to obtain

$$M_l^k \sim l^{\lfloor k/2 \rfloor}. \quad (4.14)$$

This bound can be strengthened in the $l \gg k$ limit, as here the largest commuting cliques of odd- k -Majoranas must share a single term γ_0 . This can be seen as when k is odd, large sets of commuting $(k-1)$ -Majoranas contain many operators that do not share any terms — a set of $k-1$ commuting operators that share a single term can be no larger than approximately $l^{(k-3)/2}$. Formally, let us consider a set C of commuting k -Majoranas, choose $\Gamma \in C$, and write $\Gamma = \gamma_1 \dots \gamma_k$. Then, we may write $C = \cup_i C_i$, where C_i is the subset of operators in C that contain γ_i as a term. If there exists $\Gamma' \in C/C_i$, (i.e. Γ' commutes with all operators in C_i but does not itself contain γ_i), we may divide C_i into k subsets of Majoranas that share the individual terms in Γ' , and so $|C_i| \leq kl^{(k-3)/2}$. If is true for all such C_i , we have then $|C| \leq \sum_i |C_i| \leq k^2 l^{(k-3)/2}$. As this scales suboptimally in the large- l limit³, we must have that C/C_i is empty for some C_i . Then, $C_i = C$, and we can bound

$$M_l^k \leq M_{l-1}^{k-1}, \quad k \text{ odd.} \quad (4.15)$$

³For example, we can achieve better scaling in l via our pairing scheme.

4.C Appendix: Details of measurement schemes for fermionic systems

This leads to the tighter bound (assuming l even)

$$M_l^k \leq \frac{l!!}{(l-k)!! k!!}, \quad (4.16)$$

where the double factorial implies we multiple only the even integers $\leq k$. Then, when $l = 2N$, for even $k = 2n$ we see

$$\begin{aligned} M_{2N}^{2n} &\leq \frac{2N!!}{(2N-2n)!! 2n!!} \\ &= \frac{2^N N!}{2^{N-n} (N-n)! 2^n n!} = \binom{N}{n}. \end{aligned} \quad (4.17)$$

This is precisely the size of the cliques obtained by pairing, proving this scheme optimal in the large- N limit.

In practice, we observe that Eq. 4.15 is true for $k = 3$ whenever $l \geq l_{\text{crit},3} 15$ (i.e. for > 8 -fermion systems). This is because the largest set of commuting 3-Majoranas that do not share a single common element can be found to be (up to relabeling) $\{\gamma_0\gamma_1\gamma_2, \gamma_0\gamma_3\gamma_4, \gamma_0\gamma_5\gamma_6, \gamma_1\gamma_3\gamma_5, \gamma_1\gamma_4\gamma_6, \gamma_2\gamma_3\gamma_6, \gamma_2\gamma_4\gamma_5\}$, which contains 7 terms. The above argument implies that $l_{\text{crit},k}$ scales at worst as k^2 , however the bounds obtained here are rather loose, and we expect it to do far better.

4.C Appendix: Details of measurement schemes for fermionic systems

We now construct asymptotically minimal sets of cliques that contain all 2-Majorana and 4-Majorana operators. 2-Majorana operators that share any term do not commute, so our commuting cliques of 2-Majorana operators must contain only non-overlapping pairs of Majorana terms. Equivalently, we need to find a set of pairings of $\{0, \dots, 2N\}$ such that each pair (i, j) appears in at least one pairing. This may be achieved optimally for N a power of 2 via the partitioning scheme outlined in Fig. 4.3(a). We first split $\{1, \dots, 2N\}$ into a set of $N2^{-n}$ contiguous blocks for $n = 0, \dots, \log(2N)$

$$B_m^n = \{m \times 2^n \leq i < (m+1) \times 2^n\}. \quad (4.18)$$

Then, our cliques may be constructed by pairing the i th element of B_{2m}^n with the $(i+a)$ th element of B_{2m+1}^n (modulo 2^n), as n runs over $0, \dots, \log(N)$

and a runs over $0, \dots, 2^n - 1$. Formally, this gives the set of cliques

$$\begin{aligned} C_{a,n} &:= \{\gamma_\alpha \gamma_\beta, \alpha = (m2^{n+1} + i), \\ &\quad \beta = ((2m+1)2^n + [(i+a) \bmod 2^n]), \\ &\quad m = (0, \dots, N2^{-n} - 1), i = (0, \dots, 2^n - 1)\}, \end{aligned} \quad (4.19)$$

with a total number

$$\sum_{n=0}^{\log N} 2^n = 2N - 1, \quad (4.20)$$

matching exactly the lower bound calculated in the main text. The above technique needs slight modification when N is not a power of 2 to make sure that when $|B_{2m}^n| \neq |B_{2m+1}^n|$, unpaired elements are properly accounted for, but the above optimal scaling may be retained. Code to generate an appropriate set of pairings has been added to the Openfermion package [90].

As all operators in one of the above cliques $C_{a,n}$ commute, their products commute, and the set

$$\{\gamma_i \gamma_j \gamma_k \gamma_l; \gamma_i \gamma_j, \gamma_k \gamma_l \in C_{a,n}\}, \quad (4.21)$$

is clearly a clique of commuting 4-Majorana operators. However, each 2-Majorana operator is guaranteed to be in only one of the cliques $C_{a,n}$, so this will not yet contain all 4-Majorana operators. To fix this, we aim to construct a larger set $\{C_\alpha\}$ of cliques of commuting 2-Majorana operators, such that for every set $\gamma_{i_1}, \gamma_{i_2}, \gamma_{i_3}, \gamma_{i_4}$ there exists one C_α containing both $\gamma_{i_a} \gamma_{i_b}$ and $\gamma_{i_c} \gamma_{i_d}$ (for some permutation of $a, b, c, d = 1, 2, 3, 4$). This may be achieved by the strategy illustrated in Fig. 4.3(b). For each $I = i_1, i_2, i_3, i_4$, choose the smallest n such that $I \subset B_m^n$ for some m . This implies that the $\{B_m^n\}$ split I into two parts - $I_a = I \cap B_{2m+a}^{n-1}$, for $a = 0, 1$, and $|I_0| = 1, 2$ or 3 . Suppose first $|I_0| = 2$, (case 1 in Fig. 4.3(b)). In this case, by iterating over all pairs of elements in B_{2m}^{n-1} and subsequently all pairs of elements in B_{2m+1}^{n-1} , we will at some point simultaneously pair the elements of I_0 and the elements of I_1 , as required. This may be performed in parallel for each m , making the total number of cliques generated at each n $|B_{2m}^{n-1}|^2 = 4^{n-1}$. Now, suppose $|I_0| = 3$ (case 2 in Fig. 4.3) — or $|I_0| = 1$ as the two situations are equivalent. Let $n' < n$ be the smallest number such that $I_0 \subset B_{m'}^{n'-1}$ for some m' , and we may split I_0 into two sets $I_{0,a} = I_0 \cap B_{2m'+a}^{n'-1}$ for $a = 0, 1$. Of the three elements in I_0 , two of them must either lie in $I_{0,0}$ or $I_{0,1}$ - suppose without loss of generality that $|I_{0,0}| = 2$. Then, by iterating over all pairs within $B_{2m'}^{n'-1}$, and all

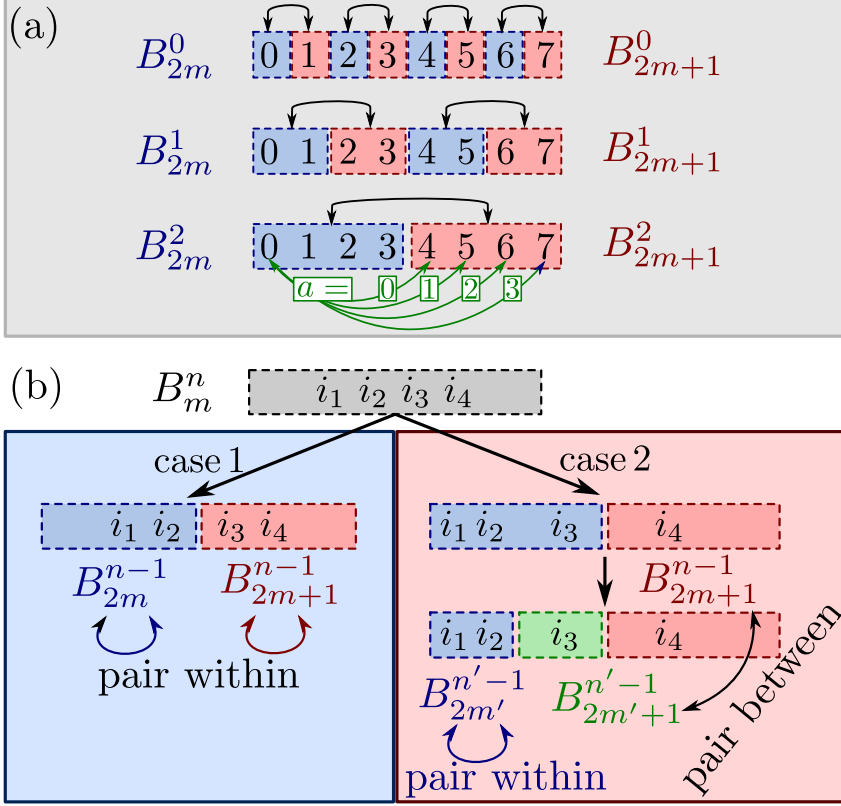


Figure 4.3: Schematic of the fermionic partition strategy for generating cliques that contain all local fermionic operators. (a) a scheme to pair all indices in $\{1, \dots, N\}$ in $\mathcal{O}(N)$ timesteps. (b) The two cases to consider in our strategy to contain all 4-Majorana operators in only $\mathcal{O}(N^2)$ cliques.

pairs between elements of $B_{2m'+1}^{n'-1}$ and B_{2m+1}^{n-1} , we will at some point pair both elements in $I_{0,0}$ and both elements in $I_{0,1} \cup I_{1,1}$.

This pairing needs to occur for all $n > n'$, which implies we need to iterate over all combinations of pairs between elements of $B_{2m'+1}^{n'-1}$ and $\{1, \dots, 2N\}/B_{2m'}^{n'-1}$ (while iterating over pairs within $B_{2m'}^{n'-1}$). This may be performed in parallel for each m' at each n' . First, iterate over all possible pairings of $B_{m_0}^{n'}$ and $B_{m_1}^{n'}$ (which requires $\mathcal{O}(N2^{-n'})$ iterations). Then, iterate over all pairs between $B_{2m_0+a_0}^{n'-1}$ and $B_{2m_1+a_1}^{n'-1}$ for all combinations of $a_0, a_1 = 0, 1$ (requiring $4 \times 2^{n'-1}$ iterations). Simultaneously, iterate over all pairs within $B_{2m_0+1-a_0}^{n'-1}$ and $B_{2m_1+1-a_1}^{n'-1}$ (requiring again $2^{n'-1}$ iterations). This generates $4 \times 4^{n'-1}$ cliques at each n' . The total number of cliques we then require to contain all 4-Majorana operators using this scheme is then

$$\sum_{n'=1}^{\lceil \log N \rceil} N2^{n'} + \sum_{n=1}^{\lceil \log N \rceil + 1} 4^{n-1} \sim \frac{10}{3} N^2. \quad (4.22)$$

4.D Appendix: Reducing operator estimation over symmetries

Given a set $\{S_i\} \subset \mathbb{P}^N$ of N_{sym} mutually-commuting Pauli operators that are symmetries ($[S_i, H] = 0$), we can simultaneously diagonalize both the Hamiltonian and the symmetries, implying that we can find a ground state ρ such that $\text{Trace}[\rho P] = 0$ for each P that does not commute with S_i . In the case of a degenerate ground state eigenspace, not all states will necessarily have this property (as symmetries may be spontaneously broken). However, any such P will not appear in the Pauli decomposition of the Hamiltonian, and so estimation of this RDM term is not necessary to calculate the energy of the state. The commutation of a k -Majorana operator Γ with a Pauli operator symmetry S_i may be seen immediately by counting how many of the k individual terms anti-commute with S_i — if this number is even, then $[\Gamma, S_i] = 0$. This implies that we can separate individual 1-Majorana operators into bins $B_{\vec{s}}$ with $\vec{s} \in \{0, 1\}^{N_{\text{sym}}}$ a commutation label:

$$\gamma_j \in B_{\vec{s}} \rightarrow \begin{cases} \gamma_j S_i = S_i \gamma_j & \text{iff } s_i = 0 \\ \gamma_j S_i = -S_i \gamma_j & \text{iff } s_i = 1. \end{cases} \quad (4.23)$$

4.D Appendix: Reducing operator estimation over symmetries

Let $\vec{s}(\gamma_j)$ denote the label of the bin γ_j may be found in, and we may generalize to all k -Majorana operators $\Gamma = \prod_{l=1}^k \gamma_{j_l}$:

$$s_i \left(\prod_{l=1}^k \gamma_{j_l} \right) = \sum_l s_i(\gamma_{j_l}) \pmod{2}. \quad (4.24)$$

To estimate the symmetry-conserved sector of the 2-RDM, we are then interested in constructing a set of cliques of 4-Majorana operators in $B_{\vec{0}}$. These take the form $\gamma_{j_1} \gamma_{j_2} \gamma_{j_3} \gamma_{j_4}$ where $\vec{s}(\gamma_{j_1}) = \vec{s}$, $\vec{s}(\gamma_{j_2}) = \vec{s} + \vec{\delta}$, $\vec{s}(\gamma_{j_3}) = \vec{s} + \vec{\alpha}$, and $\vec{s}(\gamma_{j_4}) = \vec{s} + \vec{\alpha} + \vec{\delta}$. (Recall here that in binary vector arithmetic, $\vec{a} + \vec{a} \pmod{2} = \vec{0}$.) We construct cliques for the above in two steps. First, we iterate over all quadruples within each bin $B_{\vec{s}}$ (using the methods in App. 4.C). This covers all of the above operators where $\vec{\delta} = \vec{\alpha} = 0$, and may be done simultaneously with cost $\frac{10}{3}B^2$, where B is the size of the largest bin. Then, we iterate between bins $B_{\vec{s}}$ and $B_{\vec{s}+\vec{\beta}}$ for all $\beta \in \{0, 1\}^{N_{\text{sym}}}$ with $\beta_0 = 0$. Such iteration achieves all pairs above — either $\alpha_0 = 0$ (and we pair bins $B_{\vec{s}}$ with $B_{\vec{s}+\vec{\alpha}}$ when we pair $B_{\vec{s}+\vec{\delta}}$ with $B_{\vec{s}+\vec{\delta}+\vec{\alpha}}$), or $\delta_0 = 0$ (and we pair bins $B_{\vec{s}}$ with $B_{\vec{s}+\vec{\delta}}$ when we pair $B_{\vec{s}+\vec{\alpha}}$ with $B_{\vec{s}+\vec{\delta}+\vec{\alpha}}$), or $(\vec{\delta} + \vec{\alpha})_0 = 0$ (and we pair $B_{\vec{s}}$ with $B_{\vec{s}+\vec{\delta}+\vec{\alpha}}$ when we pair $B_{\vec{s}+\vec{\alpha}}$ with $B_{\vec{s}+\vec{\delta}}$). We must perform this pairing in parallel - i.e. construct a set of $2^{N_{\text{sym}}-1}$ -tuples by drawing one element from each $B_{\vec{s}} \times B_{\vec{s}+\vec{\delta}}$ such that every two elements appear in at least one tuple. In App. 4.E we describe how this may be achieved. The total cost of the above is then $2^{N_{\text{sym}}-1}(B^2 + 2B \ln(B) + \ln(B)^2)$. It is common for most symmetries to divide the set of Majoranas in two, in which case $B = 2N \times 2^{-N_{\text{sym}}}$, and our clique cover size is

$$N^2 \left(\frac{10}{3} 4^{-N_{\text{sym}}} + 2^{1-N_{\text{sym}}} \right) + O(N \ln(N)). \quad (4.25)$$

We summarize our method in algorithm 1 (where we use $h(\vec{s})$ as the Hamming weight of a binary vector \vec{s}).

Algorithm 1 Iterate over symmetry-conserved 2-RDM elements. Here, `iterQuad` and `pairBetween` are described in App. 4.C, and `parallelIterate` in Alg. 2

```

Construct bins  $B_{\vec{s}}$ .
quadIter = {}
for  $\vec{s}$  in  $\{0, 1\}^{N_{\text{sym}}}$  do
    quadIter[ $\vec{s}$ ] = iterQuad( $B_{\vec{s}}$ )
end for
while any iterator in quadIter is not stopped do yield (next(iterator)
for iterator in quadIter if iterator is not stopped)
end while
seriesIterate(quadIter)
for  $\vec{\beta}$  in  $\{0, 1\}^{N_{\text{sym}}-1}$ ,  $\beta \neq \vec{0}$  do
    Left-append 0 to beta (i.e.  $\beta = (0, ) + \beta$ )
    quadIter = {}
    for  $\vec{s}$  in  $\{0, 1\}^{N_{\text{sym}}}$ ,  $h(\vec{s} + \vec{\beta}) \geq h(\vec{s})$  do
        quadIter[ $\vec{s}$ ] = pairBetween( $B_{\vec{s}}$ ,  $B_{\vec{s}+\vec{\beta}}$ ).
    end for
    parallelIterate(quadIter)
end for

```

4.E Appendix: Parallel iteration over pairings

If we wish to iterate over all pairs of two lists of L elements each, clearly we must perform at least L^2 total iterations, and the optimal strategy is trivial (two loops). However, if we wish to iterate over all pairs between $K = 3$ or more lists of L elements (i.e. generate a set of K -tuples such that each pair appears as a subset of one tuple), such an optimal strategy is not so obvious. When K is less than the smallest factor of L , a simple algorithm works as described in Algorithm 2. We can see that this algorithm works, for suppose $jk_1 + l = a \pmod L$ and $jk_2 + l = b \pmod L$ for two separate values of j, l - i.e. $j_1k_1 + l_1 = j_2k_1 + l_2 \pmod L$ and $j_1k_2 + l_1 = j_2k_2 + l_2 \pmod L$. Then, we have $j_1(k_1 - k_2) = j_2(k_1 - k_2) \pmod L$, and as k_1, k_2 are smaller than the lowest factor of L , $\gcd(k_1 - k_2, L) = 1$, implying $j_1 = j_2$. This scheme achieves the optimal L^2 total iterations, although the reliance on K being smaller than the lowest factor of L is somewhat unsavoury. We hypothesize that the asymptotic L^2 is indeed achievable for all $K \leq L$, but have not been unsuccessful in our search for a construc-

tion. Instead, for composite L , we suggest padding each list to have length L' , being the first number above L that achieves this requirement. The prime number theorem implies that $L' - L \sim \log(L)$ if $K \leq L$ (as then we require at worst to find the next prime number). This gives the scheme runtime $L^2 + 2L \log(L) + \log(L)^2$, which is a relatively small subleading correction.

Algorithm 2 parallelIterate: Iterate over K lists $dataArray[0], \dots, dataArray[K - 1]$ of L elements, generating all pairs between elements in separate lists. Assumes K less than the smallest factor of L .

```

for  $j = 0$  to  $L - 1$  do
  for  $k = 0$  to  $L - 1$  do
    thisTuple = [dataArray[k][jk + l mod L] for  $k = 0$  to  $K - 1$ ]
    yield thisTuple
  end for
end for

```

4.F Appendix: Measurement circuitry for fermionic RDMs

Direct measurement of products of Majorana operators is a more complicated matter than measurement of Pauli words (which require only single-qubit rotations). However, when the fermionic system is encoded on a quantum device via the Jordan-Wigner transformation [34], a relatively easy measurement scheme exists. Within this encoding, we have

$$i\gamma_{2n}\gamma_{2n+1} = Z_n, \quad (4.26)$$

so if we can permute all Majorana operators such that each pair (γ_i, γ_j) of Majoranas within a given clique is mapped to the form $(\gamma_{2n}, \gamma_{2n+1})$, they may be easily read off. To achieve such a permutation, we note that the Majorana swap gate $U_{i,j} = e^{\frac{\pi}{4}\gamma_i\gamma_j}$ satisfies

$$U_{i,j}^\dagger \gamma_k U_{i,j} = \begin{cases} \gamma_k & \text{if } i, j \neq k \\ \gamma_j & \text{if } k = i \\ -\gamma_i & \text{if } k = j. \end{cases} \quad (4.27)$$

And so repeated iteration of these unitary rotations may be used to 'sort' the Majorana operators into the desired pattern. This may be performed

in an odd-even search format [135] - at each step $t = 1, \dots, N$ we decide for each $n = 1, \dots, N$ whether to swap Majoranas $2n$ and $2n + 1$, and then whether to swap Majoranas $2n$ and $2n - 1$. Within the Jordan-Wigner transformation these gates are local:

$$U_{2n,2n+1} = e^{-i\frac{\pi}{4}Z_n}, \quad U_{2n-1,2n} = e^{-i\frac{\pi}{4}Y_{n-1}Y_n}, \quad (4.28)$$

and so each timestep is depth 3, for a total maximum circuit depth of $3N$ and total maximum gate depth $3N^2$. (To see that only N timesteps are necessary, note that each Majorana can travel up to 2 positions per timestep.) Following the Majorana swap circuit, all pairs of Majoranas that we desire to measure will be rotated to neighbouring positions and may then be locally read out. As each Majorana swap gate commutes with the global parity $\prod_{i=1}^{2N} \gamma_i$, this will be measurable alongside the clique as the total qubit parity $\prod_{i=1}^N Z_i$, allowing for error mitigation by symmetry verification [52, 53]. As the above circuit corresponds just to a basis change, for many VQEs it may be pre-compiled into the preparation itself, negating the additional circuit depth entirely.

As an alternative to the above ideas, it is possible to extend the partitioning scheme for measuring all k -qubit operators to a scheme to sample all fermionic 2-RDM elements via the Bravyi-Kitaev transformation [35, 36]. This transformation maps local fermion operators to $k = \mathcal{O}(\log N)$ qubit operators, and so using our approach the resulting scheme would require $\mathcal{O}(3^k \log^{k-1} N) = (3 \log N)^{\mathcal{O}(\log N)}$ unique measurement. Although this is superpolynomial, it is a slowly growing function for small N and also has the advantage that the measurement circuits themselves are just single qubit rotations. Furthermore, as the set of fermion operators is very sparse in the sense that it has only $\mathcal{O}(N^4)$ terms rather than $N^{\mathcal{O}(\log N)}$ terms, the scheme may be able to be further sparsified.

The measurement scheme to transform a sum of anti-commuting Majorana operators to a single Majorana operator follows a similar scheme to the Majorana swap network, but with the swap gates replaced by partial swap rotations. Let A be a set of anti-commuting Majorana (or Pauli) operators, and then for $P_i, P_j \in A$ the (anti-Hermitian) product $P_i P_j$ commutes with every element in A but P_i and P_j itself. This implies that the unitary rotation $e^{\theta P_i P_j}$ may be used to rotate between P_i and P_j without affecting the rest of A :

$$e^{-\theta P_i P_j} P_k e^{\theta P_i P_j} = \begin{cases} P_k & \text{if } k \neq i, j \\ \cos(\theta) P_i + \sin(\theta) P_j & \text{if } k = i \\ \cos(\theta) P_j - \sin(\theta) P_i & \text{if } k = j \end{cases}. \quad (4.29)$$

4.G Appendix: Proof that the maximum size of an anti-commuting clique of Pauli or Majorana

This rotation may be applied to remove the support of O on individual P_i . For example, if $\theta_1 = \tan^{-1}(\frac{c_1}{c_2})$

$$e^{-\theta_1 P_1 P_2} O e^{\theta_1 P_1 P_2} = \sqrt{c_1^2 + c_2^2} P_2 + \sum_{i=3}^L c_i P_i. \quad (4.30)$$

We extend this to remove support of O on each P_i in turn by choosing $\theta_i = \sqrt{\sum_{j<i} c_j^2} / c_{i+1}$, and then

$$\left(\prod_{i=1}^{L-1} e^{-\theta_i P_i P_{i+1}} \right) O \left(\prod_{i=1}^{L-1} e^{\theta_i P_i P_{i+1}} \right) = \sqrt{\sum_i c_i^2} P_L. \quad (4.31)$$

Following this measurement circuit, O may be measured by reading all qubits in the basis of the final Pauli P_L . Intriguingly, for $P_i, P_{i+1} \in A_{j,k,l}$, we have that $P_i P_{i+1} = \gamma_i \gamma_{i+1}$, which maps to a 2-qubit operator under the Jordan-Wigner transformation (as noted previously). This implies a measurement circuit for these sets may be achieved with only linear gate count and depth, linear connectivity, and no additional ancillas. We can slightly reduce the depth by simultaneously removing the P_i from the “top” and “bottom”; i.e., we remove P_{2N-3} by rotating with P_{2N-4} at the same time as removing P_1 by rotating with P_2 , until after exactly $N - 2$ layers, we have only the term P_N remaining. All generators in this unitary transformation commute with the parity $\prod_{i=1}^{2N} \gamma_i$, implying that it remains invariant under the transformation and may be read out alongside P_N . (This may require an additional $\mathcal{O}(1)$ gates if P_N is not mapped to products of Z_i via the Jordan-Wigner transformation.)

4.G Appendix: Proof that the maximum size of an anti-commuting clique of Pauli or Majorana operators is $2N + 1$

We prove this result in general for the Pauli group \mathbb{P}^N , and note that as the Jordan-Wigner transformation maps Majorana operators to single elements of \mathbb{P}^N , the same is true of this. We first note that elements within an anti-commuting clique $S \subset \mathbb{P}^N$ may not generate each other - let $\prod_{i=1}^n P_i = P_j \in S$, and if n is odd $[P_i, P_j] = 0$ for any P_i in the product, while if n is even $[P_k, P_j] = 0$ for any P_k not in the product. (The one exception to this rule is if one cannot find any such P_k , i.e. when

$P_j = \prod_{i \neq j, P_i \in S} P_i$). Then, note that each element $P \in \mathbb{P}^N$ commutes with precisely half of \mathbb{P}^N , and anticommutes with the other half. This can be seen because a Clifford operation C exists such that $C^\dagger P C = Z_1$, which commutes with all operators of the form $I_1 P'$ and $Z_1 P'$ and anticommutes with all operators of the form $X_1 P'$ and $Y_1 P'$, and these will be mapped to other Pauli operators when the transformation is un-done.

We may extend this result: a set $S = \{P_1, \dots, P_n\}$ of n non-generating anti-commuting elements in \mathbb{P}^N splits \mathbb{P}^N into 2^n subsets $\mathcal{P}_{\vec{b}}$ (with $\vec{b} \in \mathbb{Z}_2^n$), where $Q \in \mathcal{P}_{\vec{b}}$ commutes with P_i if $b_i = 0$ (and anticommutes if $b_i = 1$). To see that all $\mathcal{P}_{\vec{b}}$ must be the same, note that given an operator $Q \in \mathcal{P}_{\vec{b}}$, $P_i P_j Q \in \mathcal{P}_{\vec{b} \oplus \vec{\delta}_i \oplus \vec{\delta}_j}$ (as $P_i P_j$ anti-commutes with P_i and P_j but commutes with all other elements in S), so $|\mathcal{P}_{\vec{b}}|$ and $|\mathcal{P}_{\vec{b} \oplus \vec{\delta}_i \oplus \vec{\delta}_j}|$ are the same size. Similarly, if $Q \in \mathcal{P}_{\vec{b}}$, $P_i Q \in \mathcal{P}_{\vec{b} \oplus \vec{1} \oplus \vec{\delta}_i}$. If n is even, this is sufficient to connect each element in $\mathcal{P}_{\vec{b}}$ to an element in $\mathcal{P}_{\vec{b}'}$, forcing all to be the same size. However, if n is odd the above will not connect $\mathcal{P}_{\vec{b}}$ and $\mathcal{P}_{\vec{b}'}$ unless $|\vec{b}| = |\vec{b}'| \pmod{2}$. We note that $\bigcup_{\vec{b}, |\vec{b}| \pmod{2}=0} \mathcal{P}_{\vec{b}}$ is the set of elements that commute with $\prod_{P_i \in S} P_i$, and thus must be precisely half of \mathbb{P}^N . This proves that the set of operators in \mathbb{P}^N that anticommute with all elements in S is of size $4^N/2^{|S|}$. This must be an integer, so $n \leq 2N$. Then, when $n = 2N$ there is precisely one element that anticommutes with all operators in $S - \prod_{P_i \in S} P_i$, and we may add this to S to get the largest possible set of operators. Such a set is unitarily equivalent to the set of $2N$ Majorana operators γ_i and the global parity $\prod_{i=1}^{2N} \gamma_i$.

4.H Appendix: Proof of Thm. 1

To bound the number of preparations of a state ρ required to estimate a fermionic k -RDM, we first establish a correspondence between the allowed measurement protocols and measurement of a set of commuting Pauli operators on the original state ρ . As $2k$ -Majorana operators are Pauli operators, this implies that an estimate of the expectation value $\langle \Gamma_i \rangle$ of each $2k$ -Majorana operator Γ_i converges with variance

$$\text{Var}(\langle \Gamma_i \rangle) \leq \frac{(1 - \langle \Gamma_i \rangle)(1 + \langle \Gamma_i \rangle)}{4M_i}, \quad (4.32)$$

where M_i is the number of preparations and measurements of ρ in a basis containing Γ_i . We then show the existence of a worst-case state for which this upper bound is tight, which implies that to estimate $\langle \Gamma_i \rangle$ with error ϵ we require $M_i \sim \epsilon^{-2}$ preparations and measurements of ρ in a basis

containing Γ_i . To estimate expectation values of all $\binom{2N}{2k}$ $2k$ -Majorana operators to error ϵ , we need for each operator M_i measurements in a basis containing this operator. As we have established that our measurement scheme only allows such measurements in parallel if the operators commute, the bound derived in App. 4.B directly bounds the number of operators that may be estimated per preparation of ρ to $\binom{N}{k}$, and the result follows by Eq. 4.4.

We now show that our measurement protocol allows only for estimation of commuting Pauli operators. By definition, Clifford operators map Pauli operators to Pauli operators, so any measurement of a state ρ that consists of a Clifford circuit U_{Cl} and subsequent readout in the computational basis is equivalent to a measurement of the commuting Pauli operators $\{U_{\text{Cl}}^\dagger Z_j U_{\text{Cl}}\}$. (The same is true of any tensor products $\{U_{\text{Cl}}^\dagger \otimes_j Z_j U_{\text{Cl}}\} = \{\prod_j U_{\text{Cl}}^\dagger Z_j U_{\text{Cl}}\}$ on ρ — where the \otimes_j is taken over any set of qubits — and the following arguments remain true if Z_j is replaced by $\otimes_j Z_j$). It remains to show that the number of preparations is unaffected by the addition of N_a ancilla qubits in the $|0\rangle$ state. Under such an addition, we may still invert the measurement $U_{\text{Cl}}^\dagger Z_j U_{\text{Cl}} = P_{j,\rho} \otimes P_{j,a}$, where $P_{j,\rho}$ and $P_{j,a}$ are Pauli operators on the system and the ancilla qubits respectively. By construction, the state is separable across the bipartition into system and ancilla qubits, so $\langle P_{j,\rho} \otimes P_{j,a} \rangle = \langle P_{j,\rho} \rangle \langle P_{j,a} \rangle$. Then, as we require our ancilla qubits to be prepared in the $|0\rangle$ state, $\langle P_{j,a} \rangle = 0$ unless $P_{j,a}$ is a tensor product of I and Z , in which case $\langle P_{j,a} \rangle = 1$. If $\langle P_{j,a} \rangle = 0$, a measurement of Z_j does not yield any information about $\langle P_{j,\rho} \rangle$, while if $\langle P_{j,a} \rangle = 1$, a measurement of Z_j yields exactly the same information as a direct measurement of $P_{j,\rho}$. Then, consider two operators $U_{\text{Cl}}^\dagger Z_j U_{\text{Cl}} = P_{j,a} \otimes P_{j,\rho}$ and $U_{\text{Cl}}^\dagger Z_k U_{\text{Cl}} = P_{k,a} \otimes P_{k,\rho}$. We have that $[P_{j,a} \otimes P_{j,\rho}, P_{k,a} \otimes P_{k,\rho}]$ commute, and if $\langle P_{j,a} \rangle = 1$ and $\langle P_{k,a} \rangle = 1$, $P_{j,a}$ and $P_{k,a}$ commute on a term-wise basis (as they are tensor products of I and Z), which implies $[P_{j,\rho}, P_{k,\rho}] = 0$. This shows that the addition of ancilla qubits in the $|0\rangle$ state cannot be used to simultaneously measure non-commuting Pauli operators via Clifford circuits, and our allowed measurements correspond to simultaneous measurement of a set of commuting Pauli operators on ρ , as required.

Finally, we argue for the existence of a state for which Eq. 4.32 is tight. This may not always be the case - by constraining a fermionic k -RDM to the positive cone of N -representable states, Pauli operators with expectation values close to ± 1 (and thus small variance) constrain the expectation values of anti-commuting operators near 0 below this limit. This beneficial covariance is of particular importance when taking

linear combinations of RDM elements e.g. to calculate energies [132], however it requires a state have highly non-regular structure which in general will not be the case (nor known a priori). The simplest example of an unstructured state is the maximally-mixed state on N fermions; by definition all measurements of this state are uncorrelated, and the variance on estimation of all terms is $\text{Var}(\langle \Gamma_i \rangle) = \frac{1}{4M_i}$, which achieves the upper bound in Eq. 4.32. \square

5 Performance comparison of optimization methods for variational quantum algorithms

5.1 Introduction

Recently we have witnessed an explosion of quantum computer prototypes accessible to researchers in academic and industrial laboratories. Existing quantum hardware has already demonstrated the ability to outperform classical computers in specific mathematically contrived tasks [136, 137]. However, it is still unclear whether noisy intermediate-scale quantum (NISQ) [47] hardware can outperform classical computers on practically useful tasks. Here, variational quantum algorithms (VQA) [39, 40, 87] were introduced as a means of preparing classically-hard quantum states by tuning parameters of a quantum circuit to optimize a cost function by utilising a classical optimizer.

The overall performance of VQAs depends on the performance of the classical optimization algorithm. Finding the limitations of these optimization methods for different VQA tasks is critical if they are to impact research and industry. For this, researchers have proposed new classical optimization algorithms that exploit periodic properties of parametrized quantum circuits [138, 139]. Other works have focused on using machine learning techniques to optimize VQAs [140, 141]. These articles benchmark new optimization techniques relative to standard classical optimizers on a wide variety of systems. However, to the best of our knowledge, no extensive comparison of the most common optimization methods for these tasks has been reported yet.

In this chapter we study three aspects that affect the optimization performance in VQAs. We focus on four off-the-shelf optimizers (SLSQP, COBYLA, CMA-ES and SPSA) for the task of finding an approximate ground-state energy of few physical systems. We first look at two dif-

ferent sampling strategies and how they affect the optimization performance with default optimizer hyper-parameters. Then we focus on hyper-parameter tuning of CMA-ES and SPSA; finding a comparable performance between them given optimal hyper-parameters, with the winner depending on the details of the problem. Finally, we investigate the accuracy of the solutions in the presence of stochastic sampling noise. We define a ‘sampling noise floor’: a bound on the accuracy that an optimizer can reach when the optimal parameters are those of the best-ever function evaluation. Additionally, we show that CMA-ES algorithm can outperform this ‘sampling noise floor’ when the optimal parameters are selected from an internal estimate of the optimal candidate. Our main contribution is strong numerical evidence that the optimal parameters of a VQA should not be taken from the best-ever measured function evaluation.

5.2 Background

A variational quantum algorithm attempts to find approximate ground states of an N -qubit quantum system as the output of a circuit $U(\vec{\theta})$ with tunable parameters $\vec{\theta}$. This generates a variational ansatz,

$$|\Psi(\vec{\theta})\rangle = U(\vec{\theta})|\Phi\rangle, \quad (5.1)$$

where the parameters $\vec{\theta} \in [0, 2\pi]^d$ control the rotations of single and two-qubit gates in a quantum circuit implementation of U applied to an initial state $|\Phi\rangle$ (i.e., $U(\vec{\theta}) = U_k(\theta_k)U_{k-1}(\theta_{k-1})\dots U_0(\theta_0)|\Phi\rangle$). During a VQA run, these parameters are tuned to optimize a cost function $\mathcal{C}(\vec{\theta})$, which in our case is the expectation value of a Hermitian observable \hat{O} relative to the state $|\Psi(\vec{\theta})\rangle$,

$$\mathcal{C}(\vec{\theta}) = \langle \hat{O} \rangle = \langle \Psi(\vec{\theta}) | \hat{O} | \Psi(\vec{\theta}) \rangle. \quad (5.2)$$

To measure the expectation value of \hat{O} without additional quantum circuitry, it is typical to write \hat{O} as a linear combination of easy-to-measure operators, i.e., Pauli operators $\hat{P}_i \in \{\mathbb{I}, X, Y, Z\}^{\otimes N}$

$$\hat{O} = \sum_i c_i \hat{P}_i \rightarrow \mathcal{C}(\vec{\theta}) = \langle \hat{O} \rangle = \sum_i c_i \langle \hat{P}_i \rangle. \quad (5.3)$$

A VQA then passes the estimation of the cost function $\mathcal{C}(\vec{\theta})$ to some classical optimization routine to find the values of $\vec{\theta}$ minimizing \mathcal{C} . This optimization loop and the optimizer choices are the focus of this chapter.

To get an estimate of the expectation value $\langle \hat{P}_i \rangle$, one prepares and measures the state multiple times in the \hat{P}_i basis and calculates the mean of the eigenvalues observed. This approximates the cost function \mathcal{C} by an estimator $\bar{\mathcal{C}}$, whose distribution is dependent on the number of repetitions M used to calculate $\langle \hat{P}_i \rangle$,

$$\bar{\mathcal{C}}(\vec{\theta}, M) = \sum_i c_i [\langle \hat{P}_i \rangle + \epsilon_i(M)]. \quad (5.4)$$

Here, ϵ_i is a random variable drawn from a binomial distribution with variance $\sigma_i^2 \sim 1/M$ that is used to simulate the experimental shot or sampling noise. Assuming that Pauli operators are measured independently, the variance of the estimator $\bar{\mathcal{C}}$ may be propagated directly,

$$\text{Var}[\bar{\mathcal{C}}] = \sum_i c_i^2 \sigma_i^2. \quad (5.5)$$

In general, the assumption of independence is violated. One may measure mutually commuting operators in parallel [124, 125, 132–134, 142]. Then the resulting measurement has non-zero covariance [108, 115], which should be accounted for. However, this only introduces a constant factor to the estimation cost, and will not significantly impact the relative optimizer performance. Here, we will use M , defined in Eq. (5.4), as the overall cost for the quantum subroutine which takes $\vec{\theta}$ and M as inputs, and outputs $\bar{\mathcal{C}}(\vec{\theta}, M)$.

To optimize within a VQA, an access to $\mathcal{C}(\vec{\theta})$ is provided to a classical optimizer, which then minimizes the sampled cost function $\bar{\mathcal{C}}(\vec{\theta}, M)$ as a function of the classical parameters $\vec{\theta}$. One can additionally provide estimates of gradients $\nabla_{\theta} \mathcal{C}$ (or higher order derivatives) in order to perform gradient-based (or Newton-like) optimization. To avoid a comparison of the runtime of gradient estimation to that of estimating the raw cost function $\bar{\mathcal{C}}(\vec{\theta}, M)$, we only compare four gradient-free optimization algorithms. Moreover, it has been shown that gradient-based optimization strategies suffer given noisy function evaluations with simple noise structures [143] (e.g., stationary and isotropic noisy covariance) in the sense that 1) the convergence rate to local optima is hampered [144] and 2) such simple noise does not help in escaping from local optima [145]. We select the following optimizers (see Appendices for further details):

1. SLSQP determines a local search direction by solving the second-order local approximation of the cost function that satisfies the constraints,

2. COBYLA uses linear approximations of the target and constrains function to optimize a simplex within a trust region of the parameter space,
3. CMA-ES is a population-based optimization algorithm where the points are drawn from a multivariate Gaussian distribution, whose parameters (covariance matrix and location) are adapted online,
4. SPSA employs a stochastic perturbation vector to compute simultaneously an approximate gradient of the objective.

We compare these algorithms across multiple systems of different sizes and number of parameters considered.

5.3 Three-stage sampling adaptation

Existing state-of-the-art quantum hardware is limited by the stability of the devices, which need to be tuned within time-scales of hours up to a day. This influences a hard limit on the total number of samples we can measure before the devices changes, of the order of $\sim 10^9$. This shot budget becomes the limiting factor in VQAs. One must carefully balance between exploring the parameter space and accurately measuring the cost function. In this spirit, Cade et al. [146] introduced a sampling procedure for SPSA that splits the total shot budget between three stages, resulting in improved performance of VQAs.

Naively, one can fix the total number of shots per Pauli operator and run the optimization until the budget is spent (i.e., if one used 1000 shots per Pauli per function call and allocated a total shot budget of 10^7 per Pauli, this would allow for a total of 10000 evaluations). We refer to this approach as one-stage optimization. Alternatively, one could think of an optimization strategy where the number of shots is increased as the optimization progresses towards better parameters as introduced by Cade et al. We perform a three-stage optimization procedure where the number of samples per Pauli operator increases per phase, reducing the number of total function calls. In our three-stage optimization, we fix the number of shots per Pauli for each stage (i.e., 100-1000-10000 shots for a total budget of 10^7). The number of function evaluations is then computed from a ratio 10:3:1. For every 10 function calls at the first stage, we use 3 function calls in the second stage and 1 function call in the third stage (i.e., 7150-2145-715 evaluations for a total budget of 10^7 shots per Pauli).

We compare the one- and three-stage protocols for the four optimization algorithms previously introduced, aimed at assessing if the three-stage

5.3 Three-stage sampling adaptation

protocol has any evident advantage over standard sampling strategy. For this comparison we use the relative energy error,

$$\Delta_r E = \left| \frac{\mathcal{C}(\vec{\theta}_{\text{opt}}) - E_0}{E_0 - c_0} \right|, \quad (5.6)$$

where $\mathcal{C}(\vec{\theta}_{\text{opt}})$ is the noiseless cost function evaluated at the optimized parameters $\vec{\theta}_{\text{opt}}$ obtained from a noisy optimization. E_0 is the lowest eigenvalue of the problem. c_0 is the coefficient of the identity operator, which is the largest Hamiltonian term. This is the relevant figure of merit to capture the performance of the optimizer as it measures the relative error in estimating the traceless part of the Hamiltonian $H - c_0 I$, requiring the quantum computer. Our numerical experiments are performed under sampling noise with a total shot budget of 10^7 , 10^8 and 10^9 per Pauli operator. In the one-stage method we fix the total number of function evaluations to 10^4 and use 10^3 , 10^4 and 10^5 shots per Pauli operator per function call. In the three-stage procedure the function calls are also fixed at 7150-2145-715 for all budgets, and the shots per Pauli operator at every stage are 10^2 - 10^3 - 10^4 , 10^3 - 10^4 - 10^5 , and 10^4 - 10^5 - 10^6 , respectively. The optimization stops when the shot budget is reached. However, the SLSQP and COBYLA optimization algorithms have a termination criterion that in most cases results in exiting early and not utilizing their full shot budget.

We benchmark the algorithms for three different systems on 8-qubits: H_4 in a chain and square configuration and the 2x2 Hubbard model. The results are shown in figure 5.1. For every problem we run experiments for one-stage (blue, pink and green dots) and three-stage (red, brown, and purple dots) protocols for each optimization. Both SLSQP and COBYLA have a consistent improvement when the three-stage sampling is applied, but are inferior to the performance of CMA-ES and SPSA in all cases. Conversely, for CMA-ES the one-stage sampling has better performance than the three-stage protocol in all three test systems. Finally, SPSA behaves differently with respect to the sampling procedure depending on the system. For the square configuration of H_4 the three-stage sampling has a slightly better energy. In the case of the chain configuration, the one-stage method shows almost a 4-fold improvement over the three-stage method. For the 2x2 Hubbard model the three-stage methods has an order of magnitude improvement in performance. Overall, the best performance across all problems is achieved by SPSA in both one- and three-stage methods (indicated by green and orange ticks on the top of the panel).

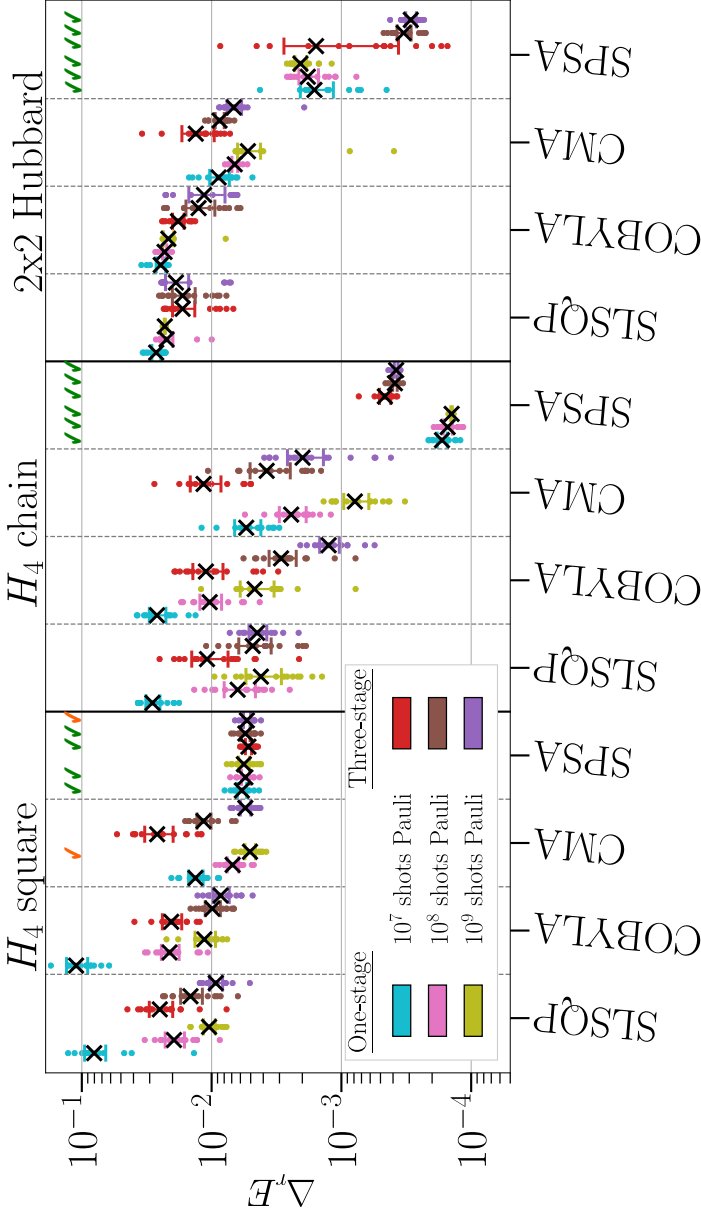


Figure 5.1: Comparison of optimizers with default hyper-parameters and for the one-stage (blue, pink and green) and three-stage (red, brown and purple) method. Black crosses depict the mean value, the error bars are the 95% confidence interval of 15 independent runs. Green ticks mark the overall winner optimization in mean and standard error. Orange ticks mark the overall winner optimization in mean with overlapping error bars to other optimizers.

5.4 Hyperparameter tuning

Most optimization algorithms come with default (hyper-)parameters detailing the optimization. These (hyper-)parameters are either derived under idealized theoretical assumptions, or evaluated from numerical experiments on standard benchmarks. It is common to tune the hyper-parameters of the optimizers when used on a function that has not been previously studied [147, 148]. Similarly, when performing a VQA one should consider hyper-parameter-tuning the optimizer.

Finding optimal hyper-parameters of an optimizer can be costly and generally problem-dependent. A sub-field of classical optimization has been devoted to automatizing such hyper-parameter tuning. Here we use the iterated racing for automatic algorithm configuration [149], IRACE (see Appendix 5.C for a description of the procedure), to tune the SPSA and CMA-ES settings for four molecular systems, H_4 square and chain and H_2O at equilibrium and stretched geometries (corresponding to weakly- and strongly-correlated regimes, respectively). Additionally we perform hyper-parameter tuning of CMA-ES for the Hubbard model on three different configurations; 1x6, 2x2 and 2x3. For SPSA, however, we take the results of ref. [146] where the hyper-parameters were tuned. The hyper-parameters used for the numerical experiments can be found in Tables I and II in Appendix 5.C.

With the tuned hyper-parameters we performed a new set of experiments including new systems: H_2O in its equilibrium and stretched geometries with 10 qubits and Hubbard models on 1x6 and 2x3 lattices requiring 12 qubits. The results of these simulations are shown in figure 5.2. A first observation is that CMA-ES improves on the problems tested without tuning. The overall performance between CMA-ES and SPSA is roughly similar, with SPSA doing better in weakly-correlated problems (H_4 chain, H_2O equilibrium and 2x2 Hubbard model). For strongly-correlated systems, CMA-ES has a slightly better performance with the mean values mostly within error bars (orange ticks in fig. 5.2). Supporting our idea that hyper-parameter tuning is crucial to ensure a good VQA performance is the 5-fold improvement of SPSA on the H_4 chain problem. For the H_4 square only a mild improvement is observed for both SPSA and CMA-ES, both reaching an almost equal relative energy error. A reason of this can be that both optimizers find the optimal parameters accessible by the ansatz, better result requiring a larger circuit.

With regards to the Hubbard model, SPSA performs clearly better on the 2x2 configurations. However CMA-ES is capable of finding much lower

5 Performance comparison of optimization methods

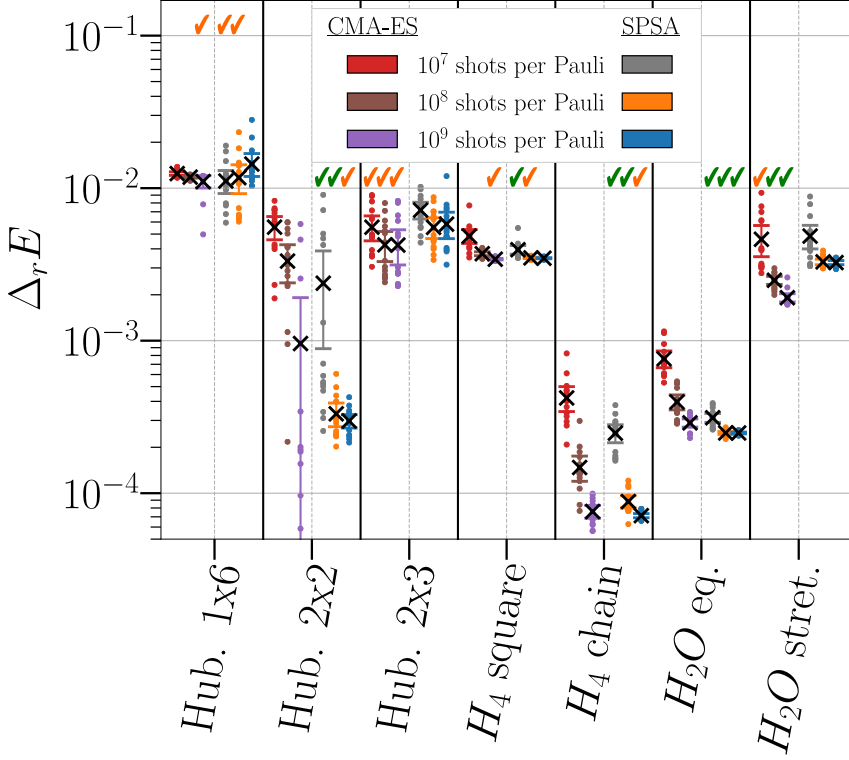


Figure 5.2: Comparison of optimized hyper-parameters of CMA-ES (red, brown and purple dots) and SPSA (grey, orange and blue dots). Black crosses and error bars depict the mean and 95% confidence interval of 15 independent runs. Green and orange represent the same as in fig. 5.1.

points in with 10^8 and 10^9 optimization, suggesting that the optimization landscape is not trivial, and that better results can be found. The 1x6 and 2x3 Hubbard models yield a comparable performance between the optimizers. Again, a possible explanation for this is that they reach the actual optimum.

Finally, we observe that CMA-ES starts to outperform SPSA as the system size and number of parameters increases. This might be an indication that SPSA is not as well-suited for large problems.

5.5 The sampling noise floor

A successful VQA requires the optimization algorithm to return the optimal parameters of $\bar{\mathcal{C}}$. It is common to assign as the optimal candidate the one with the best-ever measured $\bar{\mathcal{C}}$. However, in VQAs, the optimization is performed using a proxy cost function $\bar{\mathcal{C}}$ – a sampled version of the real objective \mathcal{C} . So, it is possible that $\bar{\mathcal{C}}(\vec{\theta})$ returns a value that is lower than its corresponding noiseless evaluation $\mathcal{C}(\vec{\theta})$ due to statistical fluctuations. Any optimizer that assigns the optimal candidate to the point with the best-ever function evaluation of $\bar{\mathcal{C}}$ will, with large probability, return a candidate worse than the global minimum (assuming its existence). The region of points in parameter space that can return the best-ever measured $\bar{\mathcal{C}}$ we refer to as the sampling noise floor, discussed in the following.

Given a cost function $\mathcal{C}(\vec{\theta})$, we attempt to optimize its sampled version $\bar{\mathcal{C}}$ with variance $\text{Var}[\bar{\mathcal{C}}]$. Let us assume $\mathcal{C}(\vec{\theta})$ has a global minimum $\vec{\theta}_g$ with noiseless value \mathcal{C}_g , and that we evaluate $\mathcal{C}(\vec{\theta})$ at multiple $\vec{\theta}$ including one evaluation at $\vec{\theta}_g$. Under sampling noise, the value of the cost function evaluated at $\vec{\theta}_g$ is drawn with some probability $1 - p$ from a confidence interval

$$\Delta_p = [\mathcal{C}_g + m(p)\sqrt{\text{Var}[\bar{\mathcal{C}}]}, -\infty), \quad (5.7)$$

where $m(p) \sim \log(p)$ is the size of the relevant confidence interval for our distribution of $\bar{\mathcal{C}}$. Assuming this distribution is symmetric, whenever $\vec{\omega} \neq \vec{\theta}_g$ satisfies $\mathcal{C}(\vec{\omega}) - m(p)\sqrt{\text{Var}[\bar{\mathcal{C}}]} \notin \Delta_p$, with probability $> (1 - p)$ an evaluation of the noisy cost function will lie above this confidence region, $\bar{\mathcal{C}}(\vec{\omega}) > \Delta_p$. Then, with confidence $> (1 - p)^2$, $\vec{\theta}_g$ will be correctly identified (between these two candidates) as the optimal set of parameters. However, when this is not the case and $\mathcal{C}(\vec{\omega}) - m(p)\sqrt{\text{Var}[\bar{\mathcal{C}}]} \in \Delta_p$, we can no longer be confident that the true minimum $\vec{\theta}_g$ will be identified. This defines a (potentially disconnected) region in parameter space,

$$\Omega(p) = \{\vec{\omega}: \mathcal{C}(\vec{\omega}) < \mathcal{C}(\vec{\theta}_g) + 2m(p)\sqrt{\text{Var}[\bar{\mathcal{C}}]}\}, \quad (5.8)$$

from where alternative candidates can be drawn with probability p . This corresponds to a region of possible cost function values,

$$\mathbb{C}_p = [\mathcal{C}(\vec{\theta}_g), \mathcal{C}(\vec{\theta}_g) + 2m(p)\sqrt{\text{Var}[\bar{\mathcal{C}}]}], \quad (5.9)$$

that an optimizer returning the best-ever measured $\bar{\mathcal{C}}$ may achieve. We define the width of this region, $2m(p)\sqrt{\text{Var}[\bar{\mathcal{C}}]}$, as the sampling noise floor.

This region is not completely defined, as we have not set a value for p . In practice, the value of p depends upon the rate at which the optimizer can converge; an optimizer that converges slowly will encounter more parameter sets $\vec{\omega}$ near the true minimum, increasing the probability that one of these parameter sets might generate a false optimum. We do not have direct access to an estimate for the width of \mathbb{C}_p , but we can still demonstrate the phenomenon numerically.

SPSA and CMA-ES have been designed not to rely on the best-ever function evaluation. In particular, CMA-ES returns two different candidates; the best-ever measured and a so-called favourite. The favourite is computed by the algorithm’s update function at the end of the optimization process and includes all accumulated prior information. As this information includes many more shots than a single function call, in principle it can average out the sampling noise over the optimization landscape, and beat the sampling noise floor. We investigate the effect of sampling noise on these two candidates returned by CMA-ES.

Figure 5.3, shows the results of the sampling noise floor in the optimization performance. For every system we computed the following energy values: $\tilde{\mathcal{C}}(\vec{\theta}_{\text{best}})$, $\mathcal{C}(\vec{\theta}_{\text{best}})$ and $\mathcal{C}(\vec{\theta}_{\text{fav}})$. For each estimated energy we compute the relative energy error $\Delta E = \frac{\mathcal{C}(\vec{\theta}_{\text{opt}}) - E_0}{|E_0 - c_0|}$. Note that the best-ever function evaluation (orange points) is often below the true energy, breaking the variational principle. This is due to the effect of sampling noise, and for a fair comparison we should compare $\mathcal{C}(\vec{\theta}_{\text{best}})$ (red points) and $\mathcal{C}(\vec{\theta}_{\text{fav}})$ (purple points), where the true cost function is evaluated. The mean value of $\mathcal{C}(\vec{\theta}_{\text{best}})$ gives an estimate for the width of the sampling noise floor. We observe that in all cases, the mean of the favourite candidate is below the mean of the best candidate, showing that the optimizer has beaten the sampling noise floor. For the Hubbard model and the H_4 systems, this is not significant (up to a 95% confidence interval), but for the two geometries of the water molecule, the difference is much larger. Choosing the favourite over the best yields up to a 3-fold reduction of error. We believe the difference in performance comes from the different optimization landscapes of the different problems; studying this in detail is a target for future work.

5.6 Conclusion

Variational quantum algorithms have recently prompted significant interest as candidates amenable to near-term hardware. However, the per-

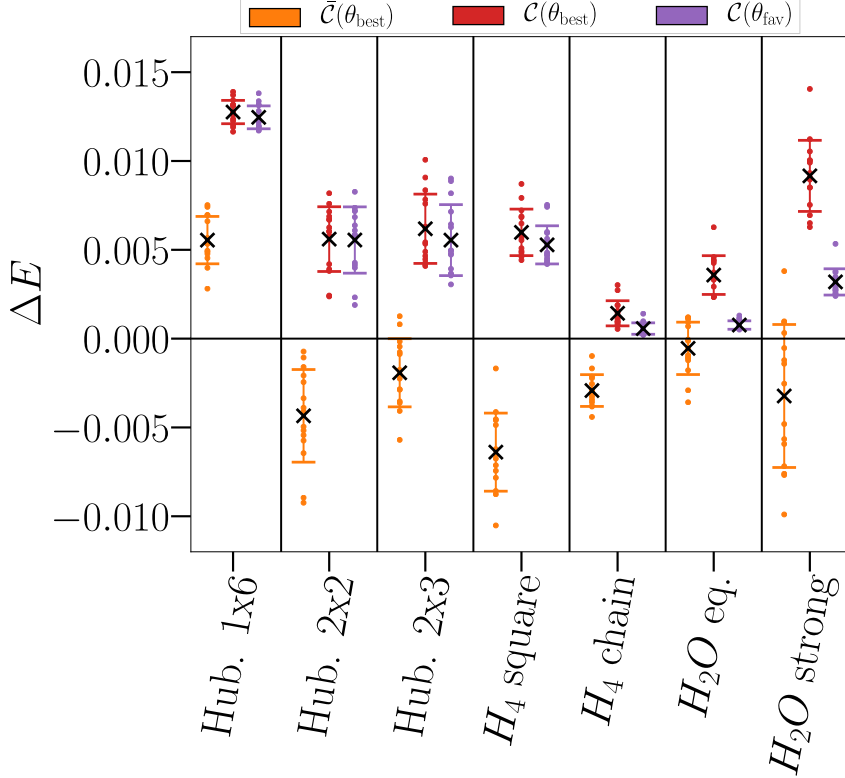


Figure 5.3: Best-ever versus favourite candidate from CMA-ES under noisy optimization. The optimization uses 10^7 shots per Pauli over the course of the experiment with individual estimations of $\bar{C}(\theta)$ are made using only 10^4 shots per Pauli. From left to right: (orange) best-ever measured function evaluation, (red) the best-ever candidate evaluated noiseless, and (purple) favourite candidate evaluated noiseless.

formance of these quantum algorithms relies on a classical optimization of a difficult cost function. This task is in general intractable to solve optimally. It is, therefore, important to benchmark the available optimizers for this purpose. We study the performance of four off-the-shelf optimization algorithms under the effect of sampling noise for the task of finding the ground-state energies. We perform a comparison using default hyper-parameters, and then extend the analysis using a three-stage sampling method from Ref. [146] and by adding hyperparameter tuning. First find that SPSA performs best in both standard and three-stage samplings without tuning. Next, we focus on the performance of SPSA and CMA-ES in the three-stage procedure. We then hyper-parameter-optimize these two methods. With these new parameter settings, SPSA and CMA-ES have a comparable performance on the strongly-correlated systems and a small advantage for SPSA on the weakly-correlated ones. We notice that the advantages of SPSA seem to vanish as problem sizes grow.

Finally, we study the effect of sampling noise on the optimization performance using CMA-ES. We observe that the best-ever function evaluation is may not be a feasible optimal candidate, contrary to the common approach in classical optimization. Specifically, we show that the best-ever result suffers from a sampling noise floor problem that makes any of the parameters within it a potential best-ever result. In contrast, the so-called CMA-ES favourite candidate obtained from its update rule at the end of the optimization shows an overall better mean and standard deviation than its best-ever counter-part, indicating the sampling noise floor can be overcome. We expect that our analytical and numerical results of the sampling noise floor opens a new line of inquiry about optimization methods for VQAs.

Appendix

5.A Appendix: Details on optimization algorithms

In this appendix, we provide a more detailed description of the optimization algorithms used in this chapter.

- Simultaneous perturbation stochastic approximation algorithm (SPSA) [150, 151] is designed for noisy evaluations of a cost function, where a stochastic perturbation vector (for instance, a vector whose components are independently sampled from the Rademacher distribution) is used to simultaneously estimate all partial derivatives at given a point. Compared to the well-known finite difference method to estimate the gradient, which requires $2d$ evaluations of the cost function defined over \mathbb{R}^d , the stochastic approximation always consumes two evaluations, hence saving many function evaluations when the search dimension is high. However, this algorithm does not follow exactly the gradient direction due to the use of stochastic perturbation.
- Constrained Optimization BY Linear Approximations (COBYLA) [152] is designed for constrained derivative-free optimization. It employs linear approximations to the objective and constraint functions via a linear interpolation given $M + 1$ points (or simplex). These approximations are then optimized within a trust region at each step.
- Sequential Least Squares Quadratic Programming (SLSQP) [153, 154] is an implementation¹ of the more general Sequential Quadratic Programming (SQP) approach [155] for solving constrained optimization problems. Loosely speaking, in each iteration, SQP proposes a local search direction by solving a sub-problem defined at the current search point in which the nonlinear cost function is replaced by its local second-order approximation and the constraints

¹We took the implementation from the `scipy` package, which is based on the original software as described in [153].

are approximated by their affine approximation. When there is no constraint, this method degenerates to Newton’s method.

- Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) [156] is the state-of-the-art direct search algorithm for continuous black-box optimization problem, which distinguishes itself from other algorithms in self-adaptation of its internal variables to the energy landscape. Briefly, this algorithm iteratively draws a number of candidate solutions from a multivariate Gaussian distribution, in which the shape of this distribution (e.g., covariance matrix and location) is adapted online based on the evaluated points in its trajectory.

5.B Appendix: Numerical experiments

In this appendix, we describe the numerical experiments used to generate the data for the figures of the chapter. The code and data to reproduce these figures can be found in [157].

To generate the target problems we use the open-source electronic structure package OpenFermion [90]. In addition, we generate the molecular systems with the computational chemistry software Psi4 through the OpenFermion plug-in. The classical numerical simulations are performed using the open-source quantum circuit simulator package Cirq [158]. Regarding the optimization methods we use the Scipy [159] software for COBYLA and SLSQP, PyCMA [160] for CMA-ES and an in-house version of SPSA based on the code in [161].

As described in the main text, we focus on the performance of the optimization methods for VQAs under sampling noise conditions. In order to include the sampling noise in our experiments we compute a noisy expectation value for every Pauli operator in the Hamiltonian with a fixed number of shots, as follows:

1. Prepare the ideal quantum state, measure $\langle P_i \rangle$ and $p = \frac{1 - \langle P_i \rangle}{2}$,
2. sample $\tilde{p} = \mathcal{B}(p, M)$ from a binomial distribution with M shots,
3. compute a noise expectation value $\langle \tilde{P}_i \rangle = 1 - 2\tilde{p}$,
4. calculate the noisy Hamiltonian expectation value as $\langle \tilde{H} \rangle = \sum_i c_i \langle \tilde{P}_i \rangle$.

This is a good approximation to the sampling noise generated by measuring the expectation values of Pauli operators in real hardware, when the number of shots is large enough. Moreover, we avoid the bottleneck of preparing and measuring the same state multiple times.

H_2O	Equilibrium	Stretched
O	(0.0, 0.0, 0.1173)	(0.0, 0.0, 0.0)
H	(0.0, 0.7572, -0.4692)	(0.0, 1.8186, 1.4081)
H	(0.0, -0.7572, -0.4692)	(0.0, -1.8186, 1.4081)

Table 5.1: Table describing the configurations of the atoms for the two water molecule problems used in this chapter.

In the Fermi-Hubbard model experiments (see app. 5.E for further details), we set the parameters of the Hamiltonian to $t = 1.0$ and $U = 2.0$. The ansatz circuit for these problems is constructed using the Variational Hamiltonian Ansatz (VHA) with 5 layer for the 1x6, 2 layer for the 2x2 and 4 layers for the 2x3 Hubbard model. These are the minimum number of layers needed to achieve a ground-state fidelity of 0.99 in ref. [146].

For the H_4 in the chain configuration, the first hydrogen atom is located at 0.0 in all coordinates, then every atom is separated in the x-direction by 1.5\AA . In the square configuration, we fix the hydrogen atoms in 2-dimensions. The positions of the atoms are parametrized by their polar coordinates with $R = 1.5\text{\AA}$ and $\theta = \frac{\pi}{4}$, and we locate them at $(x, y, 0)$, $(x, -y, 0)$, $(-x, y, 0)$, $(-x, -y, 0)$ with $x = R \cos(\theta)$ and $y = R \sin(\theta)$. For the water molecule problems, the (x, y, z) -coordinates of the atoms given in Table 5.1. Additionally, in both of the problems we reduce the active space by freezing the lowest two lowest orbitals, thus reducing the problem from 14 qubits to 10 qubits (or from 7 to 5 spin-orbitals). As a trial state to approximate the ground-state of the molecular systems, we use the so-called Unitary Couple-Cluster ansatz. A detailed description on how we construct the UCC ansatz can be found in a separate appendix 5.D.

Finally, the total number of parameters to be optimized for each target problem can be found in table 5.2.

5.C Optimization algorithms hyper-parameters

Prior to applying the aforementioned optimizers on VQAs, we also optimize the hyper-parameters of those optimizers. Such an extra tuning task aims at bringing up the performance of each optimizer to the maximum, hence facilitating a fair comparison on each problem. To achieve

5 Performance comparison of optimization methods

System	# Parameters
H_4 chain	14
H_4 square	10
H_2O eq.	26
H_2O stret.	26
Hub. 1x6	15
Hub. 2x2	6
Hub. 2x3	16

Table 5.2: Number of parameters of the ansatze for each target problem.

this task efficiently, we utilize the well-known IRACE algorithm for the hyperparameter tuning. Irace has been extensively applied in automated machine learning researches for configuring machine learning models/optimizers [162, 163].

Built upon a so-called iterated racing procedure, this algorithm employs a statistical test (usually the Wilcoxon ranked-sum test) to obtain a robust (with respect to the sampling noise in measured energy values) ranking of hyper-parameter settings, thereby serving as a suitable choosing for our task. The tuning process with IRACE initiates by fixing the subset of optimizer hyper-parameters to be modified, including bounds and potential constraints. Then, a 'race' between randomly sampled values begins. These configurations are evaluated a fixed number of times, and the less favourable configurations are disregarded based on a statistical test. The configurations that survived are then raced again until the budget of evaluations is depleted or the number of configurations is below a threshold. Next, IRACE updates the candidate generation model based on the survival configurations, and generates a set of new configurations to race against the elites. The racing procedure is repeated until the total budget is depleted. The surviving configurations are returned as the optimal configurations of the algorithm. The final hyper-parameters used for the experiments are the average of the survivors are shown in Table 5.3 and 5.4.

In detail, the hyper-parameters we tuned are as follows:

- For SPSA, we use the following ranges for each hyper-parameter: $a \in [0.01, 2]$, $\alpha \in [0, 1]$, $c \in [0.01, 2]$, and $\gamma \in [0, 1/6]$.
- For CMA-ES, we use the following ones: population size $\in [30, 130]$, c mean $\in [0, 1]$, $\mu \in [0, 0.5]$, Damp. factor $\in [0, 1]$, and $\sigma_0 \in [0.25, 1.1]$.

5.C Optimization algorithms hyper-parameters

SPSA	a	α	c	γ
default	0.15	0.602	0.2	0.101
H_4 chain	1.556	0.809	0.106	0.097
H_4 square	0.867	0.593	0.133	0.113
H_2O eq.	0.103	0.878	0.149	0.131
H_2O stret.	0.660	0.743	0.253	0.108
Hubbard	0.15	0.602	0.2	0.101

Table 5.3: List of values for SPSA hyper-parameters used in Fig. 5.2 after tuning using IRACE. We perform hyper-parameter optimization only with H_4 chain and H_2O equilibrium and use the same values for the respective square and stretched configurations. For the Hubbard models we take the default values as ref. [146] suggest their optimality.

CMA-ES²	σ_0	Population	μ	c mean	Damp. Factor
default	0.15	$[4 + 3\log(m)]$	0.5	1.0	1.0
H_4 chain	0.20	149	0.383	0.293	0.665
H_4 square	0.309	99	0.409	0.561	0.852
H_2O eq.	0.344	99	0.460	0.192	0.770
H_2O stret.	0.310	104	0.380	0.802	0.819
Hub. 1x6	0.9131	51	0.3814	0.3614	0.6006
Hub. 2x2	0.8561	113	0.2741	0.6317	0.6771
Hub. 2x3	0.897	128	0.1898	0.988	0.8391

Table 5.4: List of values for CMA-ES hyper-parameters used in Fig. 5.2 after tuning using IRACE. Here, m indicates the number of free parameters of the ansatz in each problem.

For running the irace algorithm, we allocated 500 evaluations of the hyper-parameters as the total budget, as well as a maximum total running time of 7 days, and used two evaluations of each hyper-parameter in the beginning of each race. Also, we used the F-test for eliminating worse configurations in the racing procedure. The finally suggested configurations in Table 5.3 and 5.4 are the best elites from four independent runs of irace.

As for COBYLA and SLSQP, we took their default hyper-parameter settings, i.e., $\rho_{\text{initial}} = 0.1$ and Tolerance= 10^{-8} for COBYLA and $\epsilon = 0.055$ and Tolerance= 10^{-8} .

5.D Appendix: Unitary Coupled-Cluster ansatz based on coupled-cluster amplitudes

Several classes of systems remain challenging to solve, even for the coupled cluster methods considered as the golden standard in quantum chemistry. Those systems are usually plagued by “quasidegeneracy”, meaning that the wavefunction cannot be decomposed into a single leading component. This leads to an important deterioration of methods relying on the single determinant assumption (also said to be mono-reference) [164]. This issue can be partially solved by developing multi-reference coupled cluster approaches (see Refs. [165, 166] for a review). Owing to the recent developments of quantum algorithms in the NISQ-era, there has been a renewed interest in the unitary formulation of coupled cluster (UCC) which is naturally suited for quantum computation and naturally extendable to generate multi-reference wavefunctions [39, 167], while being intractable on classical computers [115]. Several formulations of UCC have been investigated to go beyond the standard UCCSD method where only fermionic excitations from occupied to virtual orbitals (with respect to the reference determinant, usually the Hartree–Fock one) are considered [115, 168–172]. However, the number of operators (and thus the number of parameters) can rapidly become problematic if implemented naively. A powerful approach is provided by the Adaptive Derivative-Assembled Pseudo-Trotter (ADAPT) types of ansatz [172–178], which allows to adaptively increase the number of operators in the ansatz one by one until reaching a given accuracy. In this chapter, we employ a different strategy by taking advantage of the amplitudes extracted from the traditional coupled cluster method performed on a classical computer. In coupled cluster, the exponential ansatz reads as follows

$$|\Psi(\vec{t})\rangle = e^{\hat{T}}|\Phi_0\rangle, \quad (5.10)$$

where $|\Phi_0\rangle$ denotes the reference determinant (like the Hartree–Fock wavefunction) and

$$\hat{T} = \sum_{i=1}^{\eta} \hat{T}_i = \sum_{\mu} t_{\mu} \hat{\tau}_{\mu} \quad (5.11)$$

5.D Appendix: Unitary Coupled-Cluster ansatz based on coupled-cluster amplitudes

(η denotes the total number of electrons) is usually truncated to singles and doubles only:

$$\begin{aligned}\hat{T}_1 &= \sum_{\substack{i \in \text{occ} \\ a \in \text{virt}}} t_a^i \hat{a}_a^\dagger a_i, \\ \hat{T}_2 &= \sum_{\substack{i > j \in \text{occ} \\ a > b \in \text{virt}}} t_{ab}^{ij} \hat{a}_a^\dagger \hat{a}_b^\dagger a_i a_j.\end{aligned}\tag{5.12}$$

One could think of determining the CC amplitudes \mathbf{t} variationally, but this is not convenient in practice because the Baker–Campbell–Hausdorff (BCH) expansion cannot be used (because $\hat{T}^\dagger \neq -\hat{T}$). Tractable implementations rely on a non-variational optimization using the “Linked” formulation:

$$e^{-\hat{T}} \hat{H} e^{\hat{T}} |\Phi_0\rangle = E(\vec{t}) |\Phi_0\rangle.\tag{5.13}$$

The amplitudes are then determined by solving a set of non-linear equations defined by projecting Eq. (5.13) against a set of excited configurations $\{|\mu\rangle\}$ (configurations obtained from the excitation operators in \hat{T}):

$$\langle \mu | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Phi_0 \rangle = 0,\tag{5.14}$$

for which the BCH expansion can be used, as it can be naturally truncated to fourth order.

In this chapter, we computed the coupled cluster amplitudes of all our molecular systems (H_4 chain, H_4 square and H_2O) and defined our UCC ansatz according to these amplitudes. Instead of implementing UCC naively by considering all possible excitations, we only keep the excitation operators for which the corresponding CC amplitude is non-zero. This reduces already the total number of operators (and thus the total number of parameters) significantly. In practice, we use the trotterized-UCC ansatz,

$$|\Psi(\vec{\theta})\rangle = \prod_{\mu} e^{\theta_{\mu}(\hat{\tau}_{\mu} - \hat{\tau}_{\mu}^{\dagger})} |\Phi_0\rangle.\tag{5.15}$$

This trotterized form is an approximation (though it may be mitigated by the classical optimization [179]) which depends on the ordering of the operators. We decided to order the operators with respect to the value of the CC amplitudes in descending order, meaning that the first operator to be applied to the reference state within the UCC ansatz will be the

operator with the highest associated CC amplitude. We figured out that the operators in our ansatz were also the ones picked by the ADAPT-VQE ansatz [172], although the ordering might not be different. However, ADAPT-VQE can add new operators (or select and repeat an already present operator) to reach a higher accuracy. To avoid performing the (somewhat costly) first ADAPT-VQE steps, one could think about using our strategy first and then apply ADAPT-VQE for a few more steps to increase the pool of operators slightly. Note that a stochastic classical UCC can also be employed as a pre-processing step to determine the important excitation operators of the UCC ansatz, as shown in the recent work of Filip et al. [180].

In our numerical experiments the initial state is always the Hartree-Fock state corresponding to the number of electrons in the system. The parameters of the circuit are initialized at 0.0.

5.E Appendix: Variational Hamiltonian ansatz for the Hubbard model

In this section, we provide the details on the variational Hamiltonian ansatz (VHA) used for the Hubbard model problems.

The Fermi-Hubbard Hamiltonian describes the behaviour of fermions on a lattice of $n_x \times n_y$ sites. Fermions can hop to nearest-neighbour sites with some strength t , and observe a repulsion or Coulomb term of strength U to move to the same site with the same spin,

$$H_{\text{Hubbard}} = H_t + H_U = -t \sum_{(i,j),\sigma} \left(a_{i\sigma}^\dagger a_{j\sigma} + a_{j\sigma}^\dagger a_{i\sigma} \right) + U \sum_i n_{i\uparrow} n_{i\downarrow}.$$

One can further split the hopping term with respect to the vertical and horizontal hopping terms $H_t = H_v + H_h$.

The VHA were introduced in ref. [119] as a means of constructing parametrized quantum states motivated by time-evolution by Trotterization for the Hubbard model. However, in our numerical experiments we use the VHA introduced by Cade et al. [146] where the horizontal and vertical terms can be implemented in parallel (see eq.[2] in reference). The parametrized quantum state is constructed as

$$|\Psi(\vec{\theta})\rangle = U(\vec{\theta})|\Phi\rangle = \prod_{l=1}^L e^{i\theta_{v2,l} H_{v2}} e^{i\theta_{h2,l} H_{h2}} \quad (5.16)$$

$$e^{i\theta_{v1,l} H_{v1}} e^{i\theta_{h1,l} H_{h1}} e^{i\theta_{U,l} H_U}. \quad (5.17)$$

5.E Appendix: Variational Hamiltonian ansatz for the Hubbard model

The initial state $|\Phi\rangle$ is the Gaussian state of the non-interacting part of the Hamiltonian, and the parameters of the circuit are set to 0.0.

The Fermi-Hubbard Hamiltonian are generated with the open-source package OpenFermion [90].

6 Quantum simulation of hydrogen molecule on Quantum Inspire

6.1 Introduction

In the last decade the field of quantum computation has experienced an unprecedented expansion due to the development of better and larger quantum computer prototypes. Such developments have brought the attention of governments and companies with the aim of transferring academic knowledge into impactful applications for society. One of the first consequences of such effort has been the appearance of publicly accessible quantum computers via cloud services. This has allowed researchers world-wide to be able to perform their own quantum computations without the need of a specialized laboratory.

An increase in size and quality of quantum hardware is expected in the upcoming years. However, these devices will fall within the noisy intermediate-scale quantum [47] (NISQ) paradigm. It still remains an open question whether NISQ hardware will be able to achieve some level of beyond classical computation. For this reason, it is expected that the main users of cloud-based quantum computing services will be researchers with the goal of exploring how quantum computers can speed-up relevant academic and industrial problems. Therefore, the upcoming generation of cloud-based quantum computers should take into the needs of these researchers.

One of the first public quantum computers *Quantum Inspire* was made available in The Netherlands in a collaboration between TU Delft and TNO [181, 182]. The goal of *Quantum Inspire* is to provide users access to various technologies to perform quantum computations and insights in principles of quantum computing and access to the community [181]. The current version of *Quantum Inspire* allows users to use its simulator, a 5-qubit superconducting quantum processor (Starmon-5) device and a 2-qubit spin qubit processor (Spin-2).

Performing a quantum computation with quantum hardware requires a high level of knowledge and expertise. Fortunately, existing cloud-based quantum computers are built with user-friendly interfaces such that running small experiments is possible with average coding skills. While these simplifications allow anyone to access quantum computers, they typically have a negative impact on the quality and flexibility of the algorithms that can be run on them. In this work we use *Quantum Inspire* to calculate the dissociation curve of the Hydrogen molecule (H_2). Our goal is to explore the limits of *Quantum Inspire* for running NISQ computation, assessing the quality of the results, and the potential time overhead associated with interaction between local computers from users and the cloud-based service.

6.2 Quantum simulation of the Hydrogen molecule

6.2.1 Variational quantum eigensolvers

In the upcoming years it is expected that quantum hardware will be limited to a few hundreds of moderate quality qubits, with stringent limitations in coherence time and gate fidelities. To overcome such difficulties, variational quantum eigensolvers (VQE) were designed to combine quantum and classical resources such that the potential of noisy qubits is maximized. A VQE uses a quantum processor to prepare a parametrized quantum state $|\psi(\vec{\theta})\rangle$ and measure a quantum observable \hat{O} such that

$$\langle O \rangle = \langle \psi(\vec{\theta}) | \hat{O} | \psi(\vec{\theta}) \rangle. \quad (6.1)$$

The expectation value of the observable serves as a cost function to a classical optimization loop that suggests new parameters. This process is repeated until some convergence criteria is met. The quantum state $|\psi(\vec{\theta}_{\text{opt}})\rangle$ is an approximation to the lowest eigenstate of the observable, and respectively its expectation value is an approximation to its lowest eigenvalue. To run a VQE on real quantum hardware, one must prepare the parametrized quantum state through a quantum circuit. Within this quantum circuit, some of the gates have the possibility to be continuously changed through a classical knob $U(\vec{\theta})$ such that

$$|\psi(\vec{\theta})\rangle = U(\vec{\theta})|\psi_0\rangle, \quad (6.2)$$

where $|\psi_0\rangle$ is an initial quantum state (i.e. $|0\rangle^{\otimes N}$, Hartree-Fock, Gaussian state, etc.).

6.2.2 Ground-state energy of the Hydrogen molecule via VQE

The hydrogen molecule (H_2) has become a standard benchmark problem to solve using near-term quantum hardware. Calculating the ground-state energy using a VQE has been done extensively in the past years [39, 41, 42, 44, 45, 83, 98]. The number of qubits required to compute the ground-state energy of this problem in its minimal basis (STO-3G) is 4. However, one can further reduce the problem to 2 qubits by removing qubits from the inherent symmetries of the problem. Following ref. [41] we use the 2-qubit Hamiltonian under the Bravyi-Kitaev transformation at different bond distances R ,

$$H(R) = h_0(R)II + h_1(R)ZI + h_2(R)IZ + h_3(R)ZZ + h_4(R)XX + h_5(R)YY, \quad (6.3)$$

where the coefficients h_i also dependent on R . Their values can be efficiently computed using standard quantum chemistry packages [89, 90].

It is possible to obtain the exact ground-state energy of our target problem with a single parameter quantum circuit. In fig. 6.1 we show the quantum circuit implemented to calculate the ground-state energy of the problem by optimizing the free parameter θ . The parametrized part of the circuit is constructed from the unitary operation

$$U(\theta) = e^{-i\theta X_0 X_1}, \quad (6.4)$$

and decomposed into a circuit by standard methods [31] (see boxed region in fig. 6.1). The circuit is initialized in the Hartree-Fock state $|\psi_0\rangle = |10\rangle$ using a π -rotation around the y-axis. A final phase correction is needed to find the correct solution ($\frac{\pi}{2}$ rotation around the z-axis) which is implemented virtually. The final single-qubit gate rotations are used to measure the different Pauli operators of the Hamiltonian 6.3.

6.2.3 Implementation in Quantum Inspire

We use the infrastructures provided by *Quantum Inspire* [181] to calculate the dissociation curve of the H_2 molecule with the Starmon-5 processor. While *Quantum Inspire* provides a user-friendly web interface we code the quantum circuits with the Software Development Kit (SDK). The circuits are then uploaded to a server and are automatically scheduled and executed. After the completion of the experiment we receive a dataset with a binary shot count for every circuit. The data is downloaded to a local computer, processed to obtain the expectation values of every

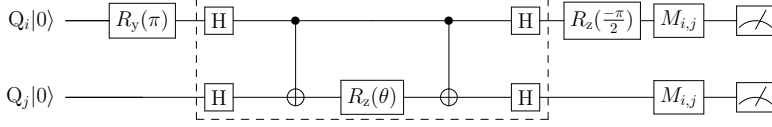


Figure 6.1: Quantum circuit used to calculate the ground-state energy of the H_2 molecule. There is an optimal value of θ such that the circuit reaches the exact ground-state energy of the Hamiltonian in eq. 6.3 for every bond distance R .

operator in eq. 6.3, and added to obtain the final energy. The fact that the data must be locally processed makes it nearly impossible to run an optimization algorithm to optimize the angle. Instead of running the VQE algorithm with the quantum hardware, we obtain the optimal angles from a classical simulation of the perfect algorithm. These optimal angles are implemented in the Starmon-5, using different combination of qubit pairs, as well as simulating them with sampling and measurement error.

Our goal is to run the circuit from fig. 6.1 with the optimal parameters previously obtained. However the actual circuit executed at the hardware level might not be exactly the one we upload to the server because there might be gates that are not native to the hardware. In our experiment we use two qubits of the Starmon-5 device available in *Quantum Inspire*. The Starmon-5 device does not have CNOT and H as native gates, thus they are automatically decomposed onto native gates. Despite the fact that we do not know the exact decomposition, we can make an estimate on the total circuit time following the specifications from the gate time specifications in [181],

$$t_{\text{cir}} = 3 \cdot t_{1q} + 2 \cdot t_H + 2 \cdot t_{\text{CNOT}} + t_{\text{meas}} + t_{\text{init}} \quad (6.5)$$

$$= 3 \cdot 20 + 2 \cdot 40 + 2 \cdot 100 + 2000 + 150 \sim 2500 \text{ ns}. \quad (6.6)$$

Those gates that are not native to Starmon-5 are automatically decomposed into native gates. As we do not know exactly which decomposition takes place we have assumed that $t_H = 40$ ns and $t_{\text{CNOT}} = 100$ ns. These numbers are taken from the fact that a generic single-qubit gate takes 20 ns and a two-qubit gate takes 60 ns, so we assumed two single-qubit gates to decompose a H-gate, and two single-qubit gates and a two-qubit

to decompose a CNOT.

In order to calculate the cost function of the problem, in our case the energy from the Hamiltonian 6.3, we must prepare and measure the state multiple times. Additionally, it is not possible to measure all terms of the Hamiltonian simultaneously, therefore each cost function evaluation requires to prepare N different circuits to estimate all the operators. In our experiments we measure every operator with its own circuit, thus needing a total of six circuits per evaluation. Every circuit is then measured with $M = 2^{12}$ shots. An estimate total time T for a single function evaluation is the

$$T = t_{\text{cir}} \cdot M \cdot N \sim 2.5 \cdot 2^{12} \cdot 6 = 61440 \mu s = 61.44 \text{ ms.} \quad (6.7)$$

However the number we obtained above does not reflect the actual runtime of a single function evaluation. There are time-delays associated with communication to the server, as well as periodical calibration and tune-up steps running in the background from which the user has no control. Therefore the total wall-time of running a single experiment is much longer than expected. In our result we studied the actual wall-clock time that takes to run a single function evaluation of the H_2 molecule.

6.2.4 Results

The results of running the optimal angles on the *Quantum Inspire* infrastructure are shown in fig. 6.2. A first observation is that the simulated results with sampling noise match almost perfectly the exact ground-state energy curve. However when one includes measurement errors in the simulation the energy error increases to 0.6 to 0.7, showing how much impact this has on the accuracy of the algorithm (dashed green curve). By applying measurement corrections the energy error is reduced by almost an order of magnitude in the simulations (dashed red curve).

Turning our attention to the curves computed with the Starmon-5 device we observe a remarkable feature that the same qubit pair with different order does not result in an identical or even similar curve. For example, the pink and light blue curves are obtained using the qubits 1 and 2, but they when qubit $i = 1$ is used the results are up to 6 times worse than when $i = 2$. An exact same trend is observed when the qubits picked are 0 and 2 (brown and green curves). Such an asymmetric behavior of the system with the same qubit pair might be explained by the large differences in coherence times and gate fidelities between qubits. When the qubit with the worst parameters carries the majority of the gates, such

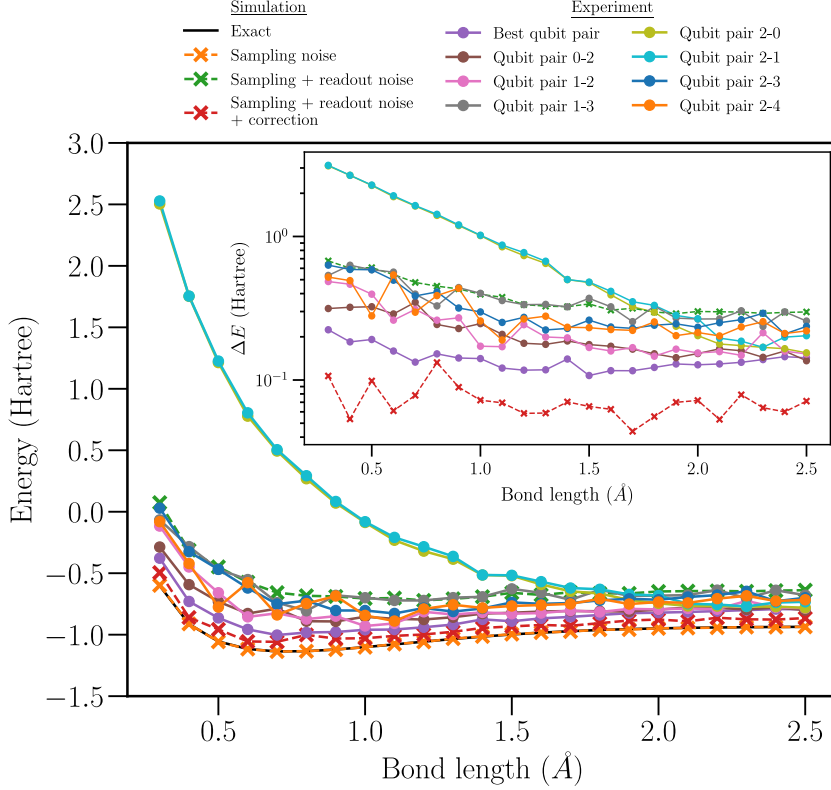


Figure 6.2: Approximate ground-state energy and energy error ΔE (inset) of the H_2 molecule with respect to the bond distance. The solid black line depicts the exact ground-state energy dissociation curve of the problem. Classically simulated energy and energy errors are shown in dashed lines with a cross marker. We simulated three types of errors: sampling noise, sampling noise with measurement errors and sampling noise with measurement and a correction to the measurement error. Energies computed with the quantum device are shown in solid lines with a dot marker. Each curve represents the result of the experiment implemented on a different pair of qubits from the Starmon-5.

a large difference in performance is not surprising. A similar asymmetry is observed when the pair of qubits 2 and 4 is used (purple and orange curves), although it is less pronounced. One possible reason for the less pronounced asymmetry is the fact that these qubits are the best pair in the device as described in the characterization data in ref. [181].

Perhaps the most striking result is given by the qubit pair 1 and 3 (grey curve). We would have expected to see a much worse result in the dissociation curve in this combination because qubits 1 and 3 are not directly coupled, and qubit 2 must be involved to mediate the interaction (see ref. [183] for details of the chip). An additional number of single- and two-qubit gates is required to implement our circuit 6.1 using qubits 1 and 3, and therefore a much worse performance is to be expected. One potential explanation is that *Quantum Inspire* automatically re-assigned the qubit pair to a pair that is directed interacted. Unfortunately, it is not possible for the end-user to know in more detail what happens after the job is submitted to the server.

We are also interested in obtaining the real runtime of performing the experiments on a cloud-based system. For this, we extracted the wall-clock time used to obtain every point in the dissociation curve for every qubit pair used. In figure 6.3 we depict the runtime for every point of the dissociation curve for every qubit pair. For each qubit pair we also calculate the mean and standard error of all the times. In average the total runtime per point is between 85 to 90 seconds. Compared to the estimated runtime from eq. 6.5 of 61.4 milliseconds, it represents more than 4 orders of magnitude in runtime. While our estimation did not include communication delays and other related times, it is surprising that all of them amount to more than a minute long.

We also observe three points that required more than 100 seconds to compute. A reason for them to occur might be because a characterization step took place in between the execution of the computation. Such process occurs periodically and automatically and the end user has no control or knowledge that it has occurred. This is particularly worrisome as the parameters of the device might suddenly change while the computation is not finished, thus making the computational unreliable.

6.3 Conclusion and outlook

Quantum computers promise a computational advantage compared to classical computers for specific problems. In the past decade world-wide efforts from public and private entities have produced a significant num-

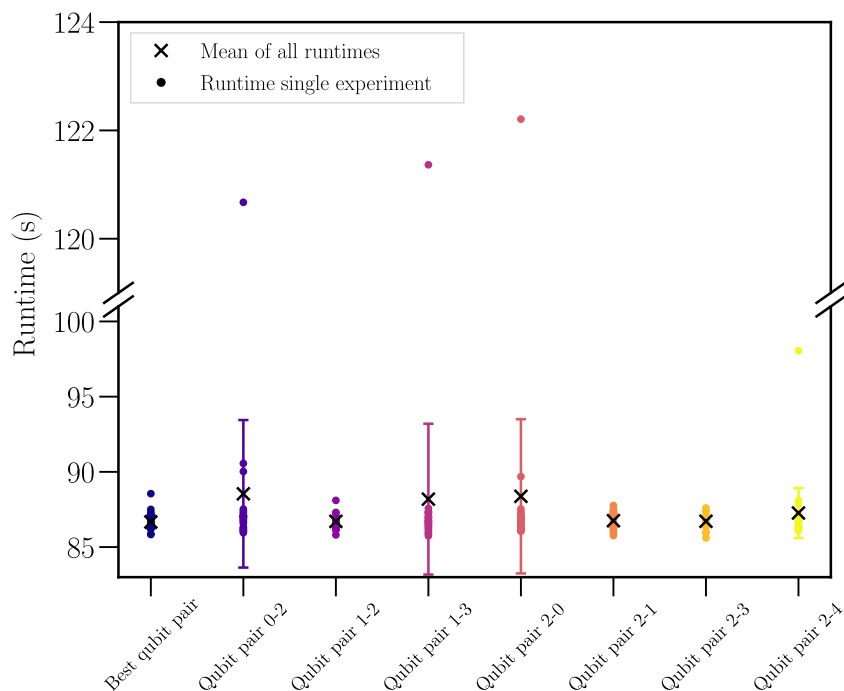


Figure 6.3: Runtime of calculating every point in the dissociation curve (dots) for each qubit pair. The cross for every qubit pair represents the mean time of all points calculated with that qubit pair, the associated error bar is the the 95% confidence interval of the population.

ber of prototypes of quantum computers, yet requiring expensive and specialized equipment that only few laboratories can maintain. Even though the state-of-the-art quantum hardware is not yet at the point where unambiguous quantum advantage can be proved, they serve as a test-bed for researchers to explore their potential. To engage with more researchers some of these laboratories with quantum computer prototypes have opened their resources to the public through cloud-access. In this chapter we explore capabilities of one of such platforms *Quantum Inspire* for research purposes.

Our first result show that even a 2-qubit toy experiment such as the ground-state dissociation curve of the H_2 requires a high level of expertise and knowledge to be performed. First of all, users must be familiar with the programming language of the platform to describe the quantum circuits to be run. Once the quantum program is ready users must upload it to the server, moment at which users have no longer control over the algorithm. In our case we have observed an unexpected good performance from a qubit pair that is not directly coupled. A possible explanation for this result is that the system has automatically implemented the experiment in a different qubit pair. Even though this might be an expected behavior of the system, the providers must carefully specify which actions are taken such that the results are trustworthy.

Secondly we study the runtime of measuring a single point in the H_2 dissociation curve. We show that the time required to obtain this point on the cloud service requires up to 4 orders of magnitude more time than an estimated expected runtime. This is worrisome if one aims at using cloud-based quantum computers for research purposes in the NISQ era. The coherence times of NISQ device is expected to be relatively short, and thus the time required to interact with the hardware must be minimized as much as possible.

7 Calculating energy derivatives for quantum chemistry on a quantum computer

7.1 Introduction

Quantum computers are at the verge of providing solutions for certain classes of problems that are intractable on a classical computer [47]. As this threshold nears, an important next step is to investigate how these new possibilities can be translated into useful algorithms for specific scientific domains. Quantum chemistry has been identified as a key area where quantum computers can stop being science and start doing science [18, 19, 185, 186]. This observation has led to an intense scientific effort towards developing and improving quantum algorithms for simulating time evolution [187, 188] and calculating ground state energies [39, 86, 87, 189] of molecular systems. Small prototypes of these algorithms have been implemented experimentally with much success [39, 41, 42, 83, 100]. However, advances over the last century in classical computational chemistry methods, such as density functional theory (DFT) [190], coupled cluster (CC) theory [191], and quantum Monte-Carlo methods [192], set a high bar for quantum computers to make impact in the field.

The ground and/or excited state energy is only one of the targets for quantum chemistry calculations. For many applications one also needs to be able to calculate the derivatives of the molecular electronic energy with respect to a change in the Hamiltonian [193, 194]. For example, the energy gradient (or first-order derivative) for nuclear displacements is used to search for minima, transition states, and reaction paths [195] that characterize a molecular potential energy surface (PES). They also form the basis for molecular dynamics (MD) simulations to dynamically explore the phase space of the system in its electronic ground state [196] or,

after a photochemical transition, in its electronically excited state [197]. While classical MD usually relies on force-fields which are parameterized on experimental data, there is a growing need to obtain these parameters on the basis of accurate quantum chemical calculations. One can easily foresee a powerful combination of highly accurate forces generated on a quantum computer with machine learning algorithms for the generation of reliable and broadly applicable force-fields [198]. This route might be particularly important in exploring excited state PES and non-adiabatic coupling terms, which are relevant in describing light-induced chemical reactions [199–201]. Apart from these perturbations arising from changing the nuclear positions, it is also of interest to consider the effect that small external electric and/or magnetic fields have on the molecular energy. These determine well-known molecular properties, such as the (hyper)polarizability, magnetizability, A- and g-tensors, nuclear magnetic shieldings, among others.

Although quantum algorithms have been suggested to calculate derivatives of a function represented on a quantum register [184, 202–205], or of derivatives of a variational quantum eigensolver (VQE) for optimization purposes [115, 206], the extraction of molecular properties from quantum simulation has received relatively little focus. To the best of our knowledge only three investigations; in geometry optimization and molecular energy derivatives [207], molecular vibrations [116], and the linear response function [208]; have been performed to date.

In this chapter, we perform a geometry optimization of the H_2 molecule on a superconducting quantum processor, as well as its response to a small electric field (polarizability), and find excellent agreement with the full configuration interaction (FCI) solution.

7.2 Background

Let \hat{H} be a Hamiltonian on a $2^{N_{\text{sys}}}$ -dimensional Hilbert space (e.g. the Fock space of an N_{sys} -spin orbital system), which has eigenstates

$$\hat{H}|\Psi_j\rangle = E_j|\Psi_j\rangle, \quad (7.1)$$

ordered by the energies E_j . In this definition, the Hamiltonian is parametrized by the specific basis set that is used and has additional coefficients $\lambda_1, \lambda_2, \dots$, which reflect fixed external influences on the electronic energy (e.g. change in the structure of the molecule, or an applied magnetic or electric field). An d th-order derivative of the ground state energy with respect to the

parameters λ_i is then defined as:

$$D_{\lambda_1, \lambda_2, \dots}^{d_1, d_2, \dots} = \frac{\partial^d E_0(\lambda_1, \lambda_2, \dots)}{\partial^{d_1} \lambda_1, \partial^{d_2} \lambda_2, \dots}, \quad (7.2)$$

where $d = \sum_i d_i$. As quantum computers promise exponential advantages in calculating the ground state E_0 itself, it is a natural question to ask how to efficiently calculate such derivatives on a quantum computer.

7.3 The quantum chemical Hamiltonian

A major subfield of computational chemistry concerns solving the electronic structure problem. Here, the system takes a second-quantized *ab initio* Hamiltonian, written in a basis of molecular spinors $\{ \phi_p(\mathbf{r}) \}$ as follows:

$$\hat{H} = \sum_{pq} h_{pq} \hat{E}_{pq} + \frac{1}{2} \sum_{pqrs} g_{pqrs} \left(\hat{E}_{pq} \hat{E}_{rs} - \delta_{q,r} \hat{E}_{ps} \right), \quad (7.3)$$

where $\hat{E}_{pq} = \hat{c}_p^\dagger \hat{c}_q$ and \hat{c}_p^\dagger (\hat{c}_p) creates (annihilates) an electron in the molecular spinor ϕ_p . With equation (7.3) relativistic and non-relativistic realizations of the method only differ in the definition of the matrix elements h_{pq} and g_{pqrs} [209]. A common technique is to assume pure spin-orbitals and integrate over the spin variable. As we want to develop a formalism that is also valid for relativistic calculations, we will remain working with spinors in this chapter. Adaptation to a spinfree formalism is straightforward, and will not affect computational scaling and error estimates.

The electronic Hamiltonian defined above depends parametrically on the nuclear positions, both explicitly via the nuclear potential and implicitly via the molecular orbitals that change when the nuclei are displaced.

7.4 Energy derivative estimation using eigenstate truncation approximation

In this section, we present a method for calculating energy derivatives on a quantum computer. For wavefunctions in which all parameters are variationally optimized, the Hellmann–Feynman theorem allows for ready calculation of energy gradients as the expectation value of the perturbing

operator [207, 211]:

$$\frac{\partial E_0}{\partial \lambda} = \langle \Psi_0 | \frac{\partial \hat{H}}{\partial \lambda} | \Psi_0 \rangle. \quad (7.4)$$

This expectation value may be estimated by repeated measurement of a prepared ground state on a quantum computer, and classical calculation of the coefficients of the Hermitian operator $\partial \hat{H} / \partial \lambda$. If state preparation is performed using a VQE, estimates of the expectation values in Eq. 7.4 will often have already been obtained during the variational optimization routine.

The Hellmann–Feynman theorem cannot be so simply extended to higher-order energy derivatives. We may write an energy derivative via perturbation theory as a sum of products of path amplitudes A and energy coefficients f_A . For example, a second order energy derivative may be written as

$$\begin{aligned} \frac{\partial^2 E_0}{\partial \lambda_1 \partial \lambda_2} &= \langle \Psi_0 | \frac{\partial^2 \hat{H}}{\partial \lambda_1 \partial \lambda_2} | \Psi_0 \rangle \\ &+ \sum_{j \neq 0} 2 \operatorname{Re} \left[\langle \Psi_0 | \frac{\partial \hat{H}}{\partial \lambda_1} | \Psi_j \rangle \langle \Psi_j | \frac{\partial \hat{H}}{\partial \lambda_2} | \Psi_0 \rangle \right] \frac{1}{E_0 - E_j}, \end{aligned} \quad (7.5)$$

allowing us to identify two amplitudes

$$\mathcal{A}_1(j) = \langle \Psi_0 | \frac{\partial \hat{H}}{\partial \lambda_1} | \Psi_j \rangle \langle \Psi_j | \frac{\partial \hat{H}}{\partial \lambda_2} | \Psi_0 \rangle, \quad (7.6)$$

$$\mathcal{A}_2 = \langle \Psi_0 | \frac{\partial^2 \hat{H}}{\partial \lambda_1 \partial \lambda_2} | \Psi_0 \rangle, \quad (7.7)$$

and two corresponding energy coefficients

$$f_1(E_0; E_j) = \frac{2}{E_0 - E_j}, \quad f_2 = 1. \quad (7.8)$$

The generic form of a d -th order energy derivative may be written as

$$\begin{aligned} D &= \sum_{\mathcal{A}} \sum_{j_1, \dots, j_{X_{\mathcal{A}}-1}} \operatorname{Re}[\mathcal{A}(j_1, \dots, j_{X_{\mathcal{A}}-1})] \\ &\times f_{\mathcal{A}}(E_0; E_{j_1}, \dots, E_{j_{X_{\mathcal{A}}-1}}), \end{aligned} \quad (7.9)$$

where $X_{\mathcal{A}}$ counts the number of excitations in the path.

One may approximate the sum over (exponentially many) eigenstates $|\Psi_j\rangle$ in Eq. 7.9 by taking a truncated set of (polynomially many) approximate eigenstates $|\tilde{\Psi}_j\rangle$. We call such an approximation the eigenstate

truncation approximation, or ETA for short. However, on a quantum computer, we expect both to better approximate the true ground state $|\Psi_0\rangle$, and to have a wider range of approximate excited states [82, 100, 214–216]. Here, we focus on the quantum subspace expansion (QSE) method of [82]. This method proceeds by generating a set of N_E vectors $|\chi_j\rangle$ connected to the ground state $|\Psi_0\rangle$ by excitation operators \hat{E}_j ,

$$|\chi_j\rangle = \hat{E}_j|\Psi_0\rangle. \quad (7.10)$$

This is similar to truncating the Hilbert space using a linear excitation operator in the (classical) equation of motion coupled cluster (EOMCC) approach [217]. The $|\chi_j\rangle$ states are not guaranteed to be orthonormal; the overlap matrix

$$S_{j,k}^{(\text{QSE})} = \langle\chi_j|\chi_k\rangle, \quad (7.11)$$

is not necessarily the identity. To generate the set $|\tilde{\Psi}_j\rangle$ of orthonormal approximate eigenstates, one can calculate the projected Hamiltonian matrix

$$H_{j,k}^{(\text{QSE})} = \langle\chi_j|\hat{H}|\chi_k\rangle, \quad (7.12)$$

and solve the generalized eigenvalue problem:

$$\hat{H}^{(\text{QSE})}\vec{v}^{(j)} = \tilde{E}_j\hat{S}^{(\text{QSE})}\vec{v}^{(j)} \rightarrow |\tilde{\Psi}_j\rangle = \sum_l \tilde{v}_l^{(j)}|\chi_l\rangle. \quad (7.13)$$

Regardless of the method used to generate the eigenstates $|\tilde{\Psi}_j\rangle$, the dominant computational cost of the ETA is the need to estimate N_E^2 matrix elements. Furthermore, to combine all matrix elements with constant error requires the variance of each estimation to scale as N_E^{-2} (assuming the error in each term is independent). Taking all single-particle excitations sets $N_E \propto N_{\text{sys}}^2$. However, in a lattice model one might consider taking only local excitations, setting $N_E \propto N_{\text{sys}}$. Further reductions to N_E will increase the systematic error from Hilbert space truncation, although this may be circumvented somewhat by extrapolation.

For the sake of completeness, we also consider here the cost of numerically estimating an energy derivative by estimating the energy at multiple points;

$$\frac{\partial^2 E}{\partial \lambda^2} = \frac{1}{\delta \lambda} \left(\frac{\partial E}{\partial \lambda}(\lambda + \delta \lambda/2) - \frac{\partial E}{\partial \lambda}(\lambda - \delta \lambda/2) \right) + O(\delta \lambda^2). \quad (7.14)$$

In this equation, the derivatives can be computed using VQE via the Hellmann–Feynman theorem. One can see that the sampling noise amplified by the division of $\delta \lambda$.

7.5 Geometry optimization on a superconducting quantum device

To demonstrate the use of energy derivatives directly calculated from a quantum computing experiment, we first perform geometry optimization of the diatomic H_2 molecule, using two qubits of a superconducting transmon device. (Details of the experiment are given in Sec. 7.B.) Geometry optimization aims to find the ground state molecular geometry by minimizing the ground state energy $E_0(\mathbf{R})$ as a function of the atomic co-ordinates R_i . In this small system, rotational and translational symmetries reduce this to a minimization as a function of the bond distance $R_{\text{H-H}}$. In Fig. 7.1, we illustrate this process by sketching the path taken by Newton’s minimization algorithm from a very distant initial bond distance ($R_{\text{H-H}} = 1.5\text{\AA}$). At each step of the minimization we show the gradient estimated via the Hellman–Feynman theorem. Newton’s method additionally requires access to the Hessian, which we calculated via the ETA (details given in Sec. 7.B). The optimization routine takes 5 steps to converge to a minimum bond length of 0.749\AA , within 0.014\AA of the target FCI equilibrium bond length (given the chosen STO-3G basis set). To demonstrate the optimization stability, we performed 100 simulations of the geometry optimization experiment on the *quantumsim* density-matrix simulator [92], with realistic sampling noise and coherence time fluctuations (details given in Sec. 7.C). We plot all simulated optimization trajectories on Fig. 7.1, and highlight the median ($R_{\text{H-H}}, E(R_{\text{H-H}})$) of the first 7 steps. Despite the rather dramatic variations between different gradient descent simulations, we observe all converging to within similar error bars, showing that our methods are indeed stable.

To study the advantage in geometry optimization from direct estimation of derivatives on a quantum computer, we compare in Fig. 7.2 our performance with gradient-free (Nelder-Mead) and Hessian-free (conjugate gradient, or CG) optimization routines. We also compare the performance of Newton’s method with an approximate Hessian from Hartree-Fock (HF) theory. All methods converge to near-identical minima, but both Newton methods converge about twice as fast as the raw CG method, which in turn converges about twice as fast as Nelder-Mead. The density-matrix simulations predict that the ETA method Hessians provide less stable convergence than the HF Hessians; we attribute this to the fact that the HF Hessian at a fixed bond distance does not fluctuate between iterations. The density-matrix simulations also predict the CG method performance to be on average much closer to the Newton’s method perfor-

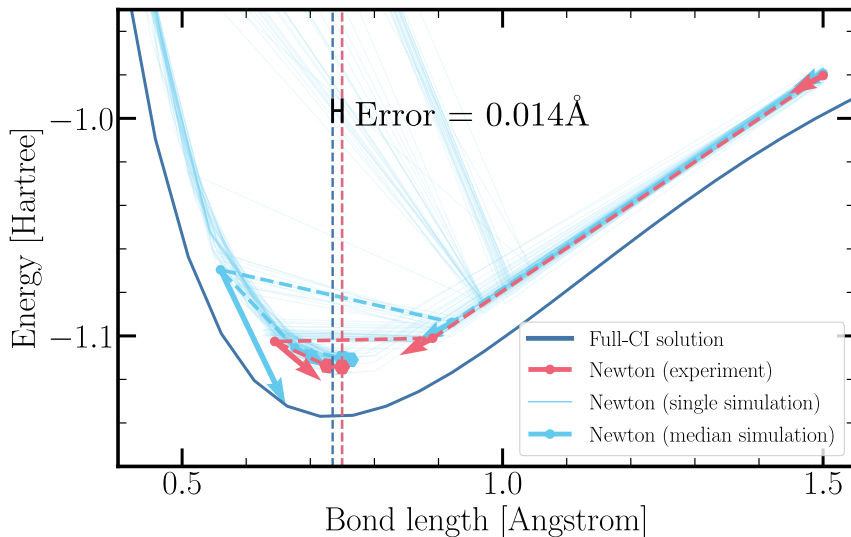


Figure 7.1: Illustration of geometry optimization of the H_2 molecule. A classical optimization algorithm (Newton) minimizes the estimation of the true ground state energy (dark blue curve) on a superconducting transmon quantum computer (red crosses) as a function of the bond distance $R_{\text{H-H}}$. To improve convergence, the quantum computer provides estimates of the FCI gradient (red arrows) and the Hessian calculated with the response method. Dashed vertical lines show the position of the FCI and estimated minima (error 0.014\AA). Light blue dashed lines show the median value of 100 density matrix simulations (Sec. 7.C) of this optimization, with the shaded region the corresponding interquartile range.

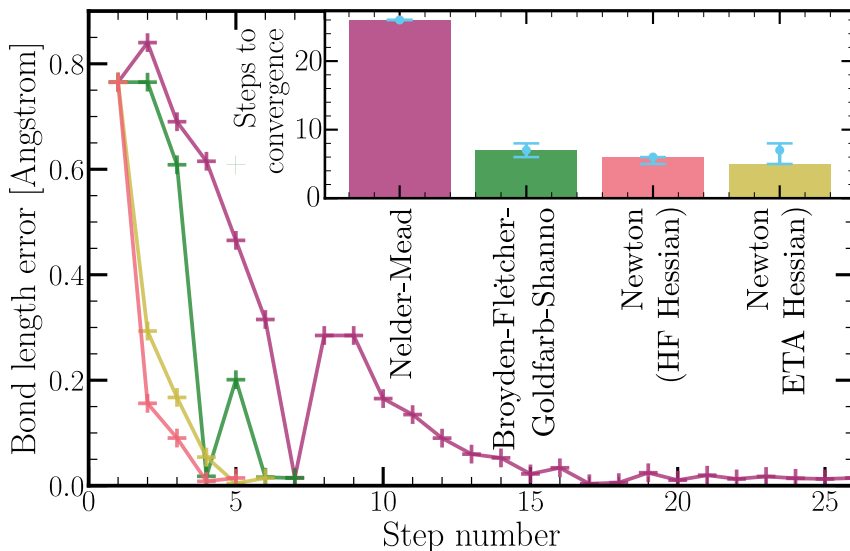


Figure 7.2: Comparison of geometry optimization via different classical optimization routines, using a quantum computer to return energies and Jacobians as required, and estimating Hessians as required either via the ETA on the experimental device, or the Hartree-Fock (HF) approximation on a classical computer. Each algorithm was run till termination with a tolerance of 10^{-3} , so as to be comparable to the final error in the system. (Inset) bar plot of the number of function evaluations of the four compared methods. Light blue points correspond to median N_{fev} from 100 density-matrix simulations (Sec. 7.C) of geometry optimization, and error bars to the interquartile ranges.

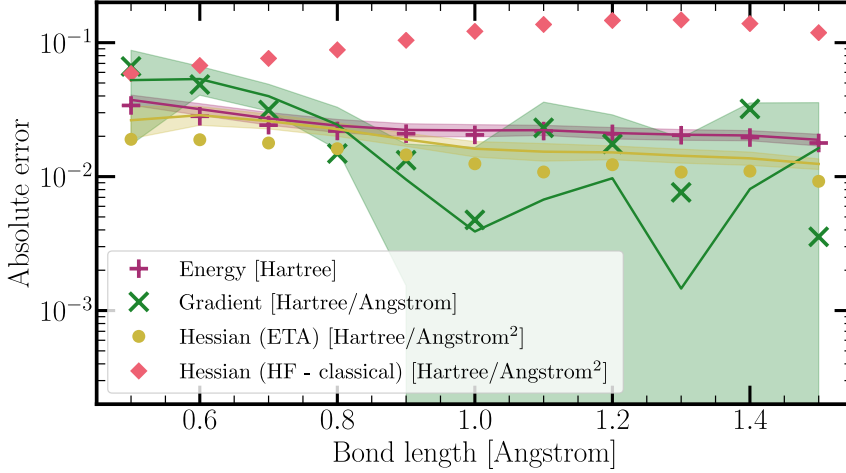


Figure 7.3: Absolute error in energies and energy derivatives from an experimental quantum computation on 11 points of the bond dissociation curve of H_2 . The error is dominated here by experimental sources (in particular qubit decay channels); error bars from sampling noise are smaller than the points themselves. Continuous lines connect the median value of 100 density matrix simulations at each points, with the shaded region corresponding to errors to the interquartile range.

mance. However, we expect the separation between gradient and Hessian-free optimization routines to become more stark at larger system sizes, as is observed typically in numerical optimization [155].

To separate the performance of the energy derivative estimation from the optimization routine, we study the error in the energy E , the Jacobian J and Hessian K given as $\epsilon_A = |A_{\text{FCI}} - A_{\text{expt}}|$, ($A = E, J, K$). In Fig. 7.3, we plot these errors for different bond distances. For comparison we additionally plot the error in the HF Hessian approximation. We observe that the ETA Hessian is significantly closer than the HF-approximated Hessian to the true value, despite the similar performance in geometry optimization. The accuracy of the ETA improves at large bond distance, where the HF approximation begins to fail, giving hope that the ETA Hessian will remain appropriate in strongly correlated systems where this occurs as well.

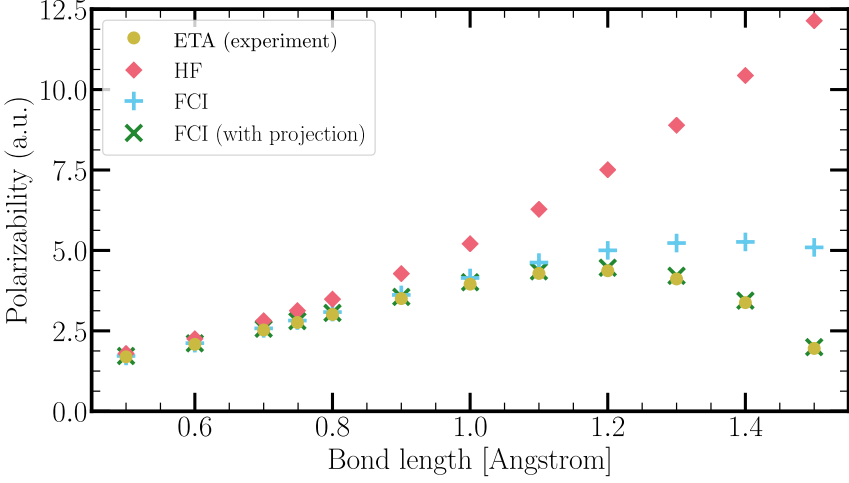


Figure 7.4: Estimated polarizability of the hydrogen molecule as a function of the bond distance, in atomic units (1 a.u. = 0.14818471 Å³).

7.6 Polarizability estimation

A key property to model in quantum chemistry is the polarizability, which describes the tendency of an atom or molecule to acquire an induced dipole moment due to a change in an external electric field \vec{F} . The polarizability tensor may be calculated as $\alpha_{i,j} = \left. \frac{\partial E(\vec{F})}{\partial F_i \partial F_j} \right|_{\vec{F}=0}$ ¹. In Fig. 7.4, we calculate the z -component of the polarizability tensor of H₂ in the ETA, and compare it to FCI and HF polarizability calculations on a classical computer. We observe good agreement to the target FCI result at low $R_{\text{H-H}}$, finding a 0.060 a.u. (2.1%) error at the equilibrium bond distance (including the inaccuracy in estimating this distance). However our predictions deviate from the exact result significantly at large bond distance ($R_{\text{H-H}} \gtrsim 1.2$ Å). We attribute this deviation to the transformation used to reduce the description of H₂ to a two-qubit device (see Sec. 7.B), which is no longer valid when adding the dipole moment operator to the Hamiltonian. To confirm this, we classically compute the FCI polarizability following the same transformation (which corresponds to projecting the larger operator onto a 2-qubit Hilbert space). We find excellent agreement between this and the result from the quantum device across the entire bond dissociation range.

¹The first-order derivative $\partial E / \partial F_i$ gives the dipole moment, which is also of interest, but is zero for the hydrogen molecule.

tion curve. This implies that simulations of H_2 on a 4-qubit device should match the FCI result within experimental error.

7.7 Conclusion

In this chapter, we have designed a new method for estimating energy gradients on a near-term quantum computer that we named eigenstate truncation approximation. We have demonstrated the use of this method on a small-scale quantum computing experiment, obtaining the equilibrium bond length of the H_2 molecule to 0.014\AA (2%) of the target Full-CI value, and estimating the polarizability at this bond length to within 0.060 a.u. (2.1%) of the same target.

Our method do not particularly target the ground state over any other eigenstate of the system, and so can be used out-of-the-box for gradient estimation for excited state chemistry. Further investigation is also required to improve some of the results drawn upon for this work, in particular reducing the number of measurements required during a VQE.

Appendix

7.A Appendix: Classical computation

The one- and two-electron integrals defining the fermionic Hamiltonian in Eq. 7.3 are obtained from a preliminary HF calculation that is assumed to be easily feasible on a classical computer. In non-relativistic theory the one-electron integrals are given by

$$h_{pq} = \int d\mathbf{r} \phi_p^*(\mathbf{r}) \left(-\frac{1}{2} \nabla^2 + V(\mathbf{r}) \right) \phi_q(\mathbf{r}), \quad (7.15)$$

where $V(\mathbf{r})$ is the electron-nuclear attraction potential from fixed nuclei at positions \mathbf{R}_i . The two-electron integrals are given by,

$$g_{pqrs} = \iint d\mathbf{r}_1 d\mathbf{r}_2 \frac{\phi_p^*(\mathbf{r}_1) \phi_q(\mathbf{r}_1) \phi_r^*(\mathbf{r}_2) \phi_s(\mathbf{r}_2)}{r_{12}}. \quad (7.16)$$

For simplicity we used a finite difference technique to compute the matrix representations of perturbations corresponding to a change in nuclear coordinates and an external electric field

$$\frac{\partial \hat{H}}{\partial \lambda} \approx \frac{\hat{H}(\lambda + \delta\lambda/2) - \hat{H}(\lambda - \delta\lambda/2)}{\delta\lambda}, \quad (7.17)$$

and

$$\frac{\partial^2 \hat{H}}{\partial \lambda^2} \approx \frac{\hat{H}(\lambda + \delta\lambda) + \hat{H}(\lambda - \delta\lambda) - 2\hat{H}(\lambda)}{\delta\lambda^2}, \quad (7.18)$$

where $\delta\lambda = 0.001$ corresponds to a small change in λ . The above (perturbed) quantum chemical Hamiltonians have been determined within the Dirac program [219] and transformed into qubit Hamiltonians using the OpenFermion [90] package. This uses the newly-developed, freely-available [220] OpenFermion-Dirac interface, allowing for the simulation of relativistic quantum chemistry calculations on a quantum computer. While a finite difference technique was sufficient for the present purpose,

such schemes are sensitive to numerical noise and have a high computational cost when applied to larger molecular systems. A consideration of the analytical calculation of energy derivatives can be found in the Supplementary Materials.

7.B Appendix: Experimental methods

The experimental implementation of the geometry optimization algorithm was performed using two of three transmon qubits in a circuit QED quantum processor. This is the same device used in Ref. [44] (raw data is the same as in Fig.1(e) of this paper, but with heavy subsequent processing). The two qubits have individual microwave lines for single-qubit gating and flux-bias lines for frequency control, and dedicated readout resonators with a common feedline. Individual qubits are addressed in readout via frequency multiplexing. The two qubits are connected via a common bus resonator that is used to achieve an exchange gate,

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & i \sin(\theta) & 0 \\ 0 & i \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (7.19)$$

via a flux pulse on the high-frequency qubit, with an uncontrolled additional single-qubit phase that was cancelled out in post-processing. The exchange angle θ may be fixed to a $\pi/6000$ resolution by using the pulse duration (with a 1 ns duration) as a rough knob and fine-tuning with the pulse amplitude. Repeat preparation and measurement of the state generated by exciting to $|01\rangle$ and exchanging through one of 41 different choices of θ resulted in the estimation of 41 two-qubit density matrices ρ_i via linear inversion tomography of 10^4 single-shot measurements per pre-rotation [111]. All circuits were executed in eQASM [221] code compiled with the QuTech OpenQL compiler, with measurements performed using the qCoDeS [222] and PycQED [223] packages.

To use the experimental data to perform geometry optimization for H_2 , the ground state was estimated via a VQE [39, 87]. The Hamiltonian at a given H-H bond distance $R_{\text{H-H}}$ was calculated in the STO-3G basis using the Dirac package [219], and converted to a qubit representation using the Bravyi-Kitaev transformation, and reduced to two qubits via exact block-diagonalization [41] using the Openfermion package [90] and the Openfermion-Dirac interface [220]. With the Hamiltonian $\hat{H}(R_{\text{H-H}})$ fixed, the ground state was chosen variationally: $\rho(R_{\text{H-H}}) =$

$\min_{\rho_i} \text{Trace}[\hat{H}(R_{\text{H-H}})\rho_i]$. The gradient and Hessian were then calculated from $\rho(R_{\text{H-H}})$ using the Hellmann–Feynman theorem and ETA respectively. For the ETA, we generated eigenstates using the QSE, with the Pauli operator XY as a single excitation. This acts within the number conserving subspace of the two-qubit Hilbert space, and, being imaginary, will not have the real-valued H_2 ground state as an eigenstate. (This in turn guarantees the generated excited state is linearly independent of the ground state.) For future work, one would want to optimize the choice of θ at each distance $R_{\text{H-H}}$, however this was not performed due to time constraints. We have also not implemented the error mitigation strategies studied in Ref. [44] for the sake of simplicity.

7.C Appendix: Simulation methods

Classical simulations of the quantum device were performed in the full-density-matrix simulator (quantumsim) [92]. A realistic error model of the device was built using experimentally calibrated parameters to account for qubit decay (T_1), pure dephasing (T_2^*), residual excitations of both qubits, and additional dephasing of qubits fluxed away from the sweet spot (which reduces T_2^* to $T_2^{*,red}$ for the duration of the flux pulse). This error model further accounted for differences in the observed noise model on the individual qubits, as well as fluctuations in coherence times and residual excitation numbers. Further details of the error model may be found in Ref. [44] (with device parameters in Tab.S1 of this reference).

With the error model given, 100 simulated experiments were performed at each of the 41 experimental angles given. Each experiment used unique coherence time and residual excitation values (drawn from a distribution of the observed experimental fluctuations), and had appropriate levels of sampling noise added. These density matrices were then resampled 100 times for each simulation.

7.D Appendix: Numerical optimization and approximate Hessian calculations

Numerous numerical methods for geometry optimization exist, some gradient-free, some requiring only gradient calculations, and some making use of both gradients and Hessian data [155, 195]. As sampling noise from a quantum computer is typically far larger than the fixed point error on a classical computer, optimization techniques are required to be stable

in the presence of this noise. In particular, common implementations of algorithms that numerically estimate gradients tend to construct approximate derivatives by difference approximations, which (as we investigated above) dramatically enhance sampling noise unless care is taken. The Nelder–Mead gradient-free algorithm [224] is a common choice for optimization in quantum algorithms for this reason; as it does not rely on such an approximation, and implementations in `scipy` [159] prove relatively stable. Gradient- and Hessian-requiring algorithms do not tend to suffer from such instability as gradient-free methods.

In this work, our geometry optimization was reduced to a one-dimensional problem, removing some of the complexity of the task. With more atoms, one need to choose both the direction and the distance to step towards the minima of the energy landscape. Both the CG and Newton’s methods are adjustments to the steepest descent algorithm (which aims to go solely in the direction of the derivative) to account for local curvature. In the absence of any higher order derivatives to assist adjustment, the non-linear CG algorithm weights each direction against traveling in previously-explored directions, and then performs a line-search in this direction (absent additional information that allows an estimation of how far to initially travel). Newton’s method, by comparison, benefits from access to the Hessian, allowing us to choose

$$\delta\mathbf{R} = \left[\frac{\partial^2 E_0}{\partial \mathbf{R}^2} \right]^{-1} \frac{\partial E_0}{\partial \mathbf{R}}, \quad (7.20)$$

for the direction. One must compensate here for the fact that we wish to minimize, and not maximise, the energy. For our one-dimensional problem this is achieved by taking the absolute value of $\partial^2 E_0 / \partial R_{\text{H-H}}^2$; for a higher-dimensional problem this is slightly more involved [155]. Regardless, such modified Newton’s methods tend to provide a far more optimal method for estimating higher dimensional functions than Hessian-free methods [155, 195]. We are further able to bound the minimum bond length in our geometry optimization (in particular to $R_{\text{H-H}} > 0.3 \text{ \AA}$), which can be of importance for stability as classical methods tend to fail when atoms are unrealistically close together.

For large systems when low accuracy is needed (e.g. at the start of a geometry optimization calculation), one may consider calculating the Hessian via the HF Hamiltonian for the same geometry as a low-cost alternative to explicit calculation on the quantum computer. This is a standard technique for geometry optimization in computational chemistry [225]. As the Hessian is not used to determine convergence (which depends instead

on the size of the gradient), the approximation only affects the convergence rate and stability, rather than the final result. This is even more so for quasi-Newton methods, as the Hessian is updated during the geometry optimization by the estimated gradients, which are more accurate. Calculating the HF Hessian is a standard procedure in most computational programs; for further mathematical information, we refer the reader to Ref. [226].

Bibliography

- [1] D. Deutsch, *Quantum theory, the Church-Turing principle and the universal quantum computer*, Proceedings of the Royal Society of London A **400**, 97-117 (1985).
- [2] D. Deutsch, and R. Jozsa, *Rapid solution of problems by quantum computation*, Proceedings of the Royal Society of London A **439**, 553-558 (1992).
- [3] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comp. **26** (5), 1484-1509 (1997).
- [4] E. Bernstein, and U. Vazirani, *Quantum complexity theory*, SIAM Journal on Computing **26**, 1411-1473 (1997).
- [5] A. Montanaro, *Quantum algorithms: an overview*, npj Quantum Information **2**, 15023 (2016).
- [6] A. P. Lund, M. J. Bremner, and T. C. Ralph, *Quantum sampling problems, BosonSampling and quantum supremacy*, npj Quantum Information **3**, 15 (2017).
- [7] A. W. Harrow, and A. Montanaro, *Quantum computational supremacy*, Nature **549**, 203-209 (2017).
- [8] C. L. Degen, F. Reinhard and P. Cappellaro, *Quantum sensing*, Rev. Mod. Phys. **89**, 035002 (2017).
- [9] C. Zalka, *Shor's algorithm with fewer (pure) qubits*, ArXiv:quant-ph/0601097 (2013).
- [10] J. Smolin, G. Smith, and A. Vargo, *Oversimplifying quantum factoring*, Nature **499**, 163-165 (2013).
- [11] M. Born, *Zur Quantenmechanik der Stoßvorgänge*, Z. Physik **37**, 863-867 (1926).

Bibliography

- [12] L. E. Ballentine, *Quantum Mechanics: a modern development (2nd edition)*, (World Scientific Publishing Company (2014)).
- [13] M. Troyer, and U.-J. Wiese, *Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations*, Phys. Rev. Lett. **94** (17), 170201 (2005).
- [14] G. Ortiz, J. Gubernatis, E. Knill, and R. Laflamme, *Quantum algorithms for fermionic simulations*, Phys. Rev. A **64**, 22319 (2001).
- [15] D. Poulin, M. B. Hastings, D. Wecker, N. Wiebe, A. C. Doherty, and M. Troyer, *The Trotter Step Size Required for Accurate Quantum Simulation of Quantum Chemistry*, Quantum Inf. and Comput. **15**, 361-384 (2015).
- [16] K. Sugisaki, S. Yamamoto, S. Nakazawa, K. Toyota, K. Sato, D. Shiomi, and T. Takui, *Quantum chemistry on quantum computers: a polynomial-time quantum algorithm for constructing the wave functions of open-shell molecules*, J. Phys. Chem. A **120**, 6459-6466 (2016).
- [17] F. Motzoi, M. P. Kaicher, and F. K. Wilhelm, *Linear and logarithmic time compositions of quantum many-body operators*, Phys. Rev. Lett. **119**, 160503 (2017).
- [18] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, *Quantum computational chemistry*, Chem. Rev. **119** (19), 10856-10915 (2019).
- [19] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, *Quantum chemistry in the age of quantum computing*, Rev. Mod. Phys. **92**, 015003 (2020).
- [20] M. M. Wilde, *Quantum Information and Entropy, In Quantum Information Theory (pp. 300-346)*, (Cambridge University Press (2017)).
- [21] W. J. Huggins, K. Wan, J. McClean, T. E. O'Brien, N. Wiebe, and R. Babbush, *Nearly optimal quantum algorithm for estimating multiple expectation values*, ArXiv:2111.09283 (2021).
- [22] N. Hansen, A. Auger, R. Ros, O. Mersmann, T. Tusar, and D. Brockhoff, *COCO: a platform for comparing continuous optimizers in a black-box setting*, Optim. Methods Softw. **36**, 114-144 (2021).

- [23] C. Doerr, F. Ye, N. Horesh, H. Wang, O. M. Shir, and T. Bäck, *Benchmarking discrete optimization heuristics with IOHprofiler*, Appl. Soft Comput. **88**, 106027 (2020).
- [24] T. Elsken, J. H. Metzen, and F. Hutter, *Neural Architecture Search: A Survey*, J. Mach. Learn. Res. **20**, 55:1-55:21 (2019).
- [25] H. Xiong, S. Shi, D. Ren, and J. Hu, *A survey of job shop scheduling problem: The types and models*, Comput. Oper. Res. **142**, 105731 (2022).
- [26] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, *Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms*, Journal of Magnetic Resonance **172**, 296-305 (2005).
- [27] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, *Comparing, optimizing, and benchmarking quantum-control algorithms in a unifying programming framework*, Phys. Rev. A **84**, 022305 (2011).
- [28] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm, *Tunable, Flexible, and Efficient Optimization of Control Pulses for Practical Qubits*, Phys. Rev. Lett. **120**, 150401 (2018).
- [29] F. Motzoi, *Controlling Quantum Information Devices*, PhD Thesis, U. Waterloo (2012).
- [30] M. A. Rol, *Control for programmable superconducting quantum systems*, PhD Thesis, TU Delft (2020).
- [31] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, *Simulation of electronic structure Hamiltonians using quantum computers*, Molecular Physics **109** (5), 735-750 (2011).
- [32] J. Hubbard, *Electron correlations in narrow energy bands*, Proceedings of the Royal Society of London A **276**, 1365 (1963).
- [33] J.P.F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. Kin-Lic Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozik, X.-W. Liu, A. J. Millis, N. V. Prokof'ev, M. Qin, G. E. Scuseria, H. Shi, B.V. Svistunov, L. F. Tocchio, I.S. Tupitsyn, S. R. White, S. Zhang, B.-X. Zheng, Z. Zhu, and E. Gull, *Solutions of the two-Dimensional Hubbard model: benchmarks and*

- results from a wide range of numerical algorithms*, Phys. Rev. X **5**, 041041 (2015).
- [34] E. P. Wigner, and P. Jordan, *Über das Paulische Äquivalenzverbot*, Z. Phys. **47**, 631 (1928).
- [35] S. B. Bravyi, and A. Y Kitaev, *Fermionic quantum computation*, Ann. Phys. **298** (1), 210-266 (2002).
- [36] J. T. Seeley, M. J. Richard, and P. J. Love, *The Bravyi-Kitaev transformation for quantum computation of electronic structure*, J. Chem. Phys. **137** (22), 224109 (2012).
- [37] M. Steudtner, and S. Wehner, *Fermion-to-qubit mappings with varying resource requirements for quantum simulation*, New Journal of Physics **20** (6), 063010 (2018).
- [38] K. Setia, and J. D. Whitfield, *Bravyi-Kitaev Superfast simulation of electronic structure on a quantum computer*, J. Chem. Phys. **148**, 164104 (2018).
- [39] A. Peruzzo, J. McClean, P. Shadbolt, M-H. Yung, X-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, *A variational eigenvalue solver on a photonic quantum processor*, Nat. Comm. **5**, 4213 (2014).
- [40] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, *Variational quantum algorithms*, Nature Reviews Physics **3**, 625-644 (2021).
- [41] 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. Lov, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, *Scalable quantum simulation of molecular energies*, Phys. Rev. X **6**, 031007 (2016).
- [42] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets*, Nature **549**, 242-246 (2017).

- [43] M. A. Nielsen, and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000).
- [44] 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*, Phys. Rev. A **100**, 010302(R) (2019).
- [45] M. Ganzhorn, D.J. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp, *Gate-Efficient Simulation of Molecular Eigenstates on a Quantum Computer*, Phys. Rev. Applied **11**, 044092 (2019).
- [46] Google AI Quantum, and collaborators, *Hartree-Fock on a superconducting qubit quantum computer* Science **369** (6507), 1084-1089 (2020).
- [47] J. Preskill, *Quantum Computing in the NISQ era and beyond*, Quantum **2**, 79 (2018).
- [48] K. Bharti, A. Cervera-Lierta, T. Ha Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W-K. Mok, S. Sim, L-C. Kwek, and A. Aspuru-Guzik *Noisy intermediate-scale quantum algorithms*, Rev. Mod. Phys. **94**, 015004 (2022).
- [49] Y. Li, and S. C. Benjamin, *Efficient Variational Quantum Simulator Incorporating Active Error Minimization*, Phys. Rev. X **7**, 021050 (2017).
- [50] K. Temme, S. Bravyi, and J. M. Gambetta, *Error Mitigation for Short-Depth Quantum Circuits*, Phys. Rev. Lett. **119**, 180509 (2017).
- [51] S. Endo, S. C. Benjamin, and Y. Li, *Practical Quantum Error Mitigation for Near-Future Applications*, Phys. Rev. X **8**, 031027 (2018).
- [52] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien *Low-cost error mitigation by symmetry verification*, Phys. Rev. A **98**, 062339 (2018).
- [53] S. McArdle, X. Yuan, and S. C. Benjamin, *Error-Mitigated Digital Quantum Simulation*, Phys. Rev. Lett. **122**, 180501 (2019).

Bibliography

- [54] Z. Cai, *Quantum Error Mitigation using Symmetry Expansion*, Quantum **5**, 548 (2021).
- [55] W. J. Huggins, S. McArdle, T. E. O’Brien, J. Lee, N. C. Rubin, S. Boixo, K. Birgitta-Whaley, R. Babbush, and J. R. McClean, *Virtual Distillation for Quantum Error Mitigation*, Phys. Rev. X **11**, 041036 (2021).
- [56] B. Koczor, *Exponential Error Suppression for Near-Term Quantum Devices*, Phys. Rev. X **11**, 031057 (2021).
- [57] P. W. Shor, *Scheme for reducing decoherence in quantum computer memory*, Phys. Rev. A **52**, R2493 (1995).
- [58] A. M. Steane, *Error correcting codes in quantum theory*, Phys. Rev. Lett. **77**, 793 (1996).
- [59] D. Gottesman, *Class of quantum error-correcting codes saturating the quantum Hamming bound*, Phys. Rev. A **54**, 1862 (1996).
- [60] D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, Ph.D. thesis, Caltech (1997).
- [61] S. B. Bravyi, and A. Y. Kitaev, *Quantum codes on a lattice with boundary*, ArXiv:quant-ph/9811052 (1998).
- [62] C. C. Bultink, T. E. O’Brien, R. Vollmer, N. Muthusubramanian, M. W. Beekman, M. A. Rol, X. Fu, B. Tarasinski, V. Ostroukh, B. Varbanov, A. Bruno, and L. DiCarlo, *Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements*, Science Advances **6** (12), eaay3050 (2020).
- [63] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff *Repeated quantum error detection in a surface code*, Nat. Phys. **16**, 875-880 (2020).
- [64] Y. Zhao, Y. Ye, H-L. Huang, Y. Zhang, D. Wu, H. Guan, Q. Zhu, Z. Wei, T. He, S. Cao, F. Chen, T-H. Chung, H. Deng, D. Fan, M. Gong, C. Guo, S. Guo, L. Han, N. Li, S. Li, Y. Li, F. Liang, J. Lin, H. Qian, H. Rong, H. Su, L. Sun, S. Wang, Y. Wu, Y. Xu, C. Ying, J. Yu, C. Zha, K. Zhang, Y-H. Huo, C-Y. Lu, C-Z. Peng, X. Zhu, and J-W. Pan, *Realization of an Error-Correcting Surface Code with Superconducting Qubits*, Arxiv:2112.13505 (2021).

- [65] H.-Y. Huang, R. Kueng, and J. Preskill, *Predicting many properties of a quantum system from very few measurements*, Nat. Phys. **16**, 1050-1057 (2020).
- [66] H.-Y. Huang, R. Kueng, and J. Preskill, *Efficient estimation of pauli Observables by derandomization*, Phys. Rev. Lett. **127**, (2021).
- [67] A. Zhao, N. C. Rubin, and A. Miyake, *Fermionic partial tomography via classical shadows*, Phys. Rev. Lett. **127**, (2021).
- [68] K. Wan, W. J. Huggins, J. Lee, and R. Babbush, *Matchgate shadows for fermionic quantum simulation*, ArXiv:2207.13723 (2021).
- [69] D. Ristè, S. Poletto, M.-Z. Huang, A. Bruno, V. Vesterinen, O.-P. Saira, and L. DiCarlo, *Detecting bit-flip errors in a logical qubit using stabilizer measurements*, Nat. Comm. **6**, 6983 (2015).
- [70] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, *Superconducting quantum circuits at the surface code threshold for fault tolerance*, Nature **508**, 441-445 (2014).
- [71] S. Debnath, N. M. Linke, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, *Demonstration of a small programmable quantum computer with atomic qubits*, Nature **536**, 63-66 (2016).
- [72] T. Monz, D. Nigg, E. A. Martinez, M. F. Brandl, P. Schindler, R. Rines, S. X. Wang, I. L. Chuang, and R. Blatt, *Realization of a scalable Shor algorithm*, Science **351** (6277), 1068-1070 (2016).
- [73] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, *Extending the lifetime of a quantum bit with error correction in superconducting circuits*, Nature **536**, 441 (2016).
- [74] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, *Quantum optimization using variational algorithms on near-term quantum devices*, Quantum Science and Technology **3**, 3 (2018).

Bibliography

- [75] C. Neill, P. Roushan, K. Kechedzhi, S. Boixo, S. V. Isakov, V. Smelyanskiy, A. Megrant, B. Chiaro, A. Dunsworth, K. Arya, R. Barends, B. Burkett, Y. Chen, Z. Chen, A. Fowler, B. Foxen, M. Giustina, R. Graff, E. Jeffrey, T. Huang, J. Kelly, P. Klimov, E. Lucero, J. Mutus, M. Neeley, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, H. Neven, J. M. Martinis, *A blueprint for demonstrating quantum supremacy with superconducting qubits*, Science **360**, 195-199 (2018).
- [76] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K-L. Chan, *Low-Depth Quantum Simulation of Materials*, Phys. Rev. X **8**, 011044 (2018).
- [77] D. Poulin, A. Kitaev, D. S. Steiger, M. B. Hastings, and M. Troyer, *Quantum Algorithm for Spectral Measurement with a Lower Gate Count*, Phys. Rev. Lett. **121**, 010501 (2018).
- [78] D. W. Berry, M. Kieferová, A. Scherer, Y. R. Sanders, G. H. Low, N. Wiebe, C. Gidney, and R. Babbush, *Improved techniques for preparing eigenstates of fermionic Hamiltonians*, npj Quantum Information **4**, 22 (2018).
- [79] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K-L. Chan, and R. Babbush, Ryan *Quantum Simulation of Electronic Structure with Linear Depth and Connectivity*, Phys. Rev. Lett. **120**, 110501 (2018).
- [80] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, *Error mitigation extends the computational reach of a noisy quantum processor*, Nature **567**, 491-495 (2019).
- [81] M. Otten, and S. K. Gray, *Recovering noise-free quantum observables*, Phys. Rev. A **99**, 012338 (2019).
- [82] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, *Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states*, Phys. Rev. A **95**, 042308 (2017).
- [83] 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*, Phys. Rev. X **8**, 011021 (2018).

- [84] D. Gottesman, *An introduction to quantum error correction and fault-tolerant quantum computation*, Proc. Sympos. Appl. Math. **68**, 13 (2010).
- [85] B. M. Terhal, *Quantum error correction for quantum memories*, Rev. Mod. Phys. **87**, 307-346 (2015).
- [86] A. Y. Kitaev, *Quantum measurements and the Abelian Stabilizer Problem*, ArXiv:quant-ph/9511026 (1995).
- [87] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, *The theory of variational hybrid quantum-classical algorithms*, New Journal of Physics **18**, 023023 (2016).
- [88] E. Farhi, J. Goldstone, and S. Gutmann, *A Quantum Approximate Optimization Algorithm*, ArXiv:1411.4028 (2014).
- [89] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill *Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability*, Journal of Chemical Theory and Computation **13**, 3185-3197 (2017).
- [90] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, Tarini Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. P. D. Sawaya, B. Senjean, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, *OpenFermion: the electronic structure package for quantum computers*, Quantum Science and Technology **5**, 034014 (2020).
- [91] D. Gottesman, *The Heisenberg Representation of Quantum Computers*, Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics, 32-43 (1999).
- [92] T. E. O'Brien, B. Tarasinski, and L. DiCarlo, *Density-matrix simulation of small surface codes under current and projected experimental noise*, npj Quantum Information **3**, 39 (2017).

Bibliography

- [93] S. Filipp, P. Maurer, P. J. Leek, M. Baur, R. Bianchetti, J. M. Fink, M. Göppl, L. Steffen, J. M. Gambetta, A. Blais, and A. Wallraff, *Two-Qubit state tomography using a joint dispersive readout*, Phys. Rev. Lett. **102**, 200402 (2009).
- [94] J. M. Chow, L. DiCarlo, J. M. Gambetta, A. Nunnenkamp, Lev S. Bishop, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf *Detecting highly entangled states with a joint qubit readout*, Phys. Rev. A **81**, 062325 (2010).
- [95] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, *Improving quantum algorithms for quantum chemistry*, Quant. Inf. Comput. **15**, (2015).
- [96] I. M. Georgescu, S. Ashhab, and F. Nori, *Quantum simulation*, Rev. Mod. Phys. **86**, 153-185 (2014).
- [97] M. Suzuki, *Generalized Trotter's formula and systematic approximations of exponential operators and inner derivations with applications to many-body problems*, Comm. Math. Phys. **51**, 183-190 (1976).
- [98] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, *Quantum chemistry calculations on a trapped-ion quantum simulator*, Phys. Rev. X **8**, 031022 (2018).
- [99] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, *Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure*, Phys. Rev. A **95**, 020501(R) (2017).
- [100] R. Santagati, J. Wang, A. Gentile, S. Paesani, N. Wiebe, J. McClean, S. Morley-Short, P. Shadbolt, D. Bonneau, J. Silverstone, D. Tew, X. Zhou, J. O'Brien, and M. Thompson, *Witnessing eigenstates for quantum simulation of Hamiltonian spectra*, Science Advances **4** (2018).
- [101] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. Joshi, P. Jurcevic, C. Muschik, P. Silvi, R. Blatt, C. Roos, and P. Zoller, *Self-verifying variational quantum simulation of lattice models*, Nature **569**, 355-360 (2019).

- [102] M. Huo and Y. Li, *Self-consistent tomography of temporally correlated errors*, ArXiv:1811.02734 (2018).
- [103] J. McClean, S. Boixo, V. Smelyanskiy, R. Babbush, and H. Neven, *Barren plateaus in quantum neural network training landscapes*, Nature Communications **9**, 4812 (2018).
- [104] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *Demonstration of two-qubit algorithms with a superconducting quantum processor*, Nature **460**, 240-244 (2009).
- [105] J. Majer, J. M. Chow, J. M. Gambetta, B. R. Johnson, J. A. Schreier, L. Frunzio, D. I. Schuster, A. A. Houck, A. Wallraff, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, *Coupling superconducting qubits via a cavity bus*, Nature **449**, 443 (2007).
- [106] D. Brod and A. Childs, *The computational power of matchgates and the XY interaction on arbitrary graphs*, Quantum Information & Computation **14**, 901 (2014).
- [107] N. Hansen, *Benchmarking the Nelder-Mead downhill simplex algorithm with many local restarts*, in Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, 2403-2408 (2009).
- [108] N. C. Rubin, R. Babbush, and J. McClean, *Application of fermionic marginal constraints to hybrid quantum algorithms*, New J. Phys **20**, 053020 (2018).
- [109] R. Blume-Kohout, *Optimal, reliable estimation of quantum states*, New J. Phys **12**, 043034 (2010).
- [110] Y.-K. Liu, M. Christandl, and F. Verstraete, *Quantum Computational Complexity of the N-Representability Problem: QMA Complete*, Phys. Rev. Lett. **98**, 110503 (2007).
- [111] O.-P. Saira, J. P. Groen, J. Cramer, M. Meretska, G. de Lange, and L. DiCarlo, *Entanglement Genesis by Ancilla-Based Parity Measurement in 2D Circuit QED*, Phys. Rev. Lett. **112**, 070502 (2014).
- [112] C. Dickel, J. J. Wesdorp, N. K. Langford, S. Peiter, R. Sagastizabal, A. Bruno, B. Criger, F. Motzoi, and L. DiCarlo, *Chip-to-chip entanglement of transmon qubits using engineered measurement fields*, Phys. Rev. B **97**, 064508 (2018).

- [113] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, *Universal Quantum Gate Set Approaching Fault-Tolerant Thresholds with Superconducting Qubits*, Phys. Rev. Lett. **109**, 060501 (2012).
- [114] D. Greenbaum, *Introduction to Quantum Gate Set Tomography*, ArXiv:1509.02921 (2015).
- [115] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, *Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz*, Quantum Sci. Technol. **4**, 014008 (2018).
- [116] S. McArdle, A. Mayorov, X. Shan, S. Benjamin, and X. Yuan, *Quantum Computation of Molecular Vibrations*, Chem. Sci. **10**, 5725 (2019).
- [117] T. E. O’Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, *Calculating Energy Derivatives for Quantum Chemistry on a Quantum Computer*, npj Quantum Information **5**, 113 (2019).
- [118] Y. Nam, J-S. Chen, N. C. Pienti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, K. M. Beck, A. Blinov, V. Chaplin, M. Chmielewski, C. Collins, S. Debnath, K. M. Hudek, A. M. Ducore, M. Keesan, S. M. Kreikemeier, J. Mizrahi, P. Solomon, M. Williams, J. D. Wong-Campos, D. Moehring, C. Monroe, and J. Kim, *Ground-State Energy Estimation of the Water Molecule on a Trapped Ion Quantum Computer*, npj Quantum Information **6**, 33 (2019).
- [119] D. Wecker, M. B. Hastings, and M. Troyer, *Progress Towards Practical Quantum Variational Algorithms*, Phys. Rev. A **92**, 042303 (2015).
- [120] C. Overy, G. H. Booth, N. S. Blunt, J. J. Shepherd, D. Cleland, and A. Alavi, *Unbiased Reduced Density Matrices and Electronic Properties from Full Configuration Interaction Quantum Monte Carlo*, J. Chem. Phys. **141**, 244117 (2014).
- [121] G. Gidofalvi, and D. A. Mazziotti, *Molecular Properties from Variational Reduced-Density-Matrix Theory with Three-Particle N-representability Conditions*, J. Chem. Phys. **126**, 024105 (2007).

- [122] T. Takeshita, N. C. Rubin, Z. Jiang, E. Lee, R. Babbush, and J. R. McClean, *Increasing the Representation Accuracy of Quantum Simulations of Chemistry without Extra Quantum Resources*, Phys. Rev. X **10**, 011004 (2020).
- [123] J. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity*, (Oxford University Press, New York, 2006).
- [124] J. Cotler, and F. Wilczek, *Quantum Overlapping Tomography*, Phys. Rev. Lett. **124**, 100401 (2020).
- [125] V. Verteletskyi, T.-C. Yen, and A. Izmaylov, *Measurement Optimization in the Variational Quantum Eigensolver Using a Minimum Clique Cover*, J. Chem. Phys. **152**, 124114 (2020).
- [126] R. M. Karp in *Complexity of Computer Computations*, edited by R. E. Miller, J. W. Thatcher, and J. D. Bohlinger (Springer, Boston, 1972).
- [127] A. Jena, S. Genin, and M. Mosca, *Pauli Partitioning with Respect to Gate Sets*, arXiv:1907.07859 (2019).
- [128] T.-C. Yen, V. Verteletskyi, and A. F. Izmaylov, *Measuring All Compatible Operators in One Series of a Single-Qubit Measurements Using Unitary Transformations*, J. Chem. Theory Comput. **16**, 4 (2020).
- [129] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, *Minimizing State Preparations in Variational Quantum Eigensolver by Partitioning into Commuting Families*, arXiv:1907.13623 (2019).
- [130] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, *Unitary Partitioning Approach to the Measurement Problem in the Variational Quantum Eigensolver Method*, J. Chem. Theory Comput. **16**, 1 (2020).
- [131] A. F. Izmaylov, T.-C. Yen, and I. G. Ryabinkin, *Revising Measurement Process in the Variational Quantum Eigensolver: Is It Possible to Reduce the Number of Separately Measured Operators?*, Chem. Sci. **10**, 3746 (2019)
- [132] W. J. Huggins, J. McClean, N. Rubin, Z. Jiang, N. Wiebe, B. Whaley, and R. Babbush, *Efficient and noise resilient measurements for*

- quantum chemistry on near-term quantum computers*, npj Quantum Information **7**, 23 (2021).
- [133] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, *Efficient quantum measurement of Pauli operators in the presence of finite sampling error*, Quantum **5**, 385 (2021).
 - [134] A. Zhao, A. Tranter, W. M. Kirby, S. F. Ung, A. Miyake, and P. J. Love, *Measurement Reduction in Variational Quantum Algorithms*, Phys. Rev. A **101**, 062322 (2020).
 - [135] A. N. Haberman, Carnegie Mellon University Technical Report No. AD-759-248, 1972
 - [136] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, and et al., *Quantum supremacy using a programmable superconducting processor*, Nature **574**, 505-510 (2019).
 - [137] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li, S. Li, Y. Li, F. Liang, C. Lin, J. Lin, H. Qian, D. Qiao, H. Rong, H. Su, L. Sun, L. Wang, S. Wang, D. Wu, Y. Wu, Y. Xu, K. Yan, W. Yang, Y. Yang, Y. Ye, J. Yin, C. Ying, J. Yu, C. Zha, C. Zhang, H. Zhang, K. Zhang, Y. Zhang, H. Zhao, Y. Zhao, L. Zhou, C.-Y. Lu, C.-Z. Peng, X. Zhu, and J.-W. Pan, *Quantum Computational Advantage via 60-Qubit 24-Cycle Random Circuit Sampling*, ArXiv:2109.03494 (2021).
 - [138] K. M. Nakanishi, K. Fujii, and S. Todo, *Sequential minimal optimization for quantum-classical hybrid algorithms*, Phys. Rev. Research **2**, 043158 (2020).
 - [139] M. Ostaszewski, E. Grant, and M. Benedetti, *Structure optimization for parameterized quantum circuits*, Quantum **5**, 391 (2021).
 - [140] M. Wilson, R. Stromswold, F. Wudarski, S. Hadfield, N. M. Tubman, and E. G. Rieffel, *Optimizing quantum heuristics with meta-learning*, Quantum Machine Intelligence **3**, 13 (2021).
 - [141] K. J. Sung, J. Yao, M. P. Harrigan, N. C. Rubin, Z. Jiang, L. Lin, R. Babbush, and J. R. McClean, *Using models to improve optimizers for variational quantum algorithms*, Quantum Science and Technology **5**, 044008 (2020).

- [142] X. Bonet-Monroig, R. Babbush and T. E. O'Brien, *Nearly optimal measurement scheduling for partial tomography of quantum state*, Phys. Rev. X **10**, 031064(2020).
- [143] J. Wu, W. Hu, H. Xiong, J. Huan, V. Braverman, and Z. Zhu, *On the Noisy Gradient Descent that Generalizes as SGD*, 37th International Conference on Machine Learning, PMLR 119:10367-10376, (2020).
- [144] L. Yu, K. Balasubramanian, S. Volgushev, and M. A. Erdogdu, *An Analysis of Constant Step Size SGD in the Non-convex Regime: Asymptotic Normality and Bias*, CoRR (2020).
- [145] Z. Zhu, J. Wu, B. Yu, L. Wu, and J. Ma, *The Anisotropic Noise in Stochastic Gradient Descent: Its Behavior of Escaping from Sharp Minima and Regularization Effects*, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, Proceedings of Machine Learning Research **97**, 7654-7663 (2019).
- [146] C. Cade, L. Mineh, A. Montanaro, and S. Stanisic, *Strategies for solving the Fermi-Hubbard model on near-term quantum computers*, Phys. Rev. B **102**, 235122 (2020).
- [147] C. Doerr, F. Ye, N. Horesh, H. Wang, O. M. Shir, and T. Bäck, *Benchmarking discrete optimization heuristics with IOHprofiler*, App. Soft Computing **88**, 106027 (2020).
- [148] N. Hansen, D. Brockhoff, O. Mersmann, T. Tusar, D. Tusar, O. A. ElHara, P. R. Sampaio, A. Atamna, K. Varelas, U. Batu, D. M. Nguyen, F. Matzner, and A. Auger, *COMparing Continuous Optimizers: numbb/COCO on Github (v2.3)*, <https://doi.org/10.5281/zenodo.2594848>, Zenodo (2019)
- [149] M. López-Ibáñez, J. Dubois-Lacoste, L. Pérez Cáceres, M. Birattari, and T. Stützle, *The irace package: Iterated racing for automatic algorithm configuration*, Operations Research Perspectives **3**, 43-58 (2016).
- [150] J. C. Spall, *Multivariate stochastic approximation using a simultaneous perturbation gradient approximation*, IEEE Transactions on Automatic Control **37**, 332 (1992).
- [151] J. C. Spall, *John Hopkins apl technical digest* 19, 482 (1998).

- [152] M. J. D. Powell, *Direct search algorithms for optimization calculations*, Acta Numerica **7**, 287-336 (1998).
- [153] D. Kraft, *A software package for sequential quadratic programming*, Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt Köln: Forschungsbericht (Wiss. Berichtswesen d. DFVLR, 1988).
- [154] D. Kraft, *Algorithm 733: TOMP-Fortran modules for optimal control calculations*, ACM Transactions on Mathematical Software **20**, 262-281 (1994).
- [155] *Sequential quadratic programming* in Numerical Optimization (Springer New York, New York, NY, 2006) pp. 529-562.
- [156] N. Hansen, and A. Ostermeier, *Completely Derandomized Self-Adaptation in Evolution Strategies*, Evolutionary Computation **9**, 159-195 (2001).
- [157] X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Back, V. Dunjko, and T. E. O'Brien, *Performance comparison of optimization methods on variational quantum algorithms (v0.1)*, <https://doi.org/10.5281/zenodo.5721349>, Zenodo (2021).
- [158] Cirq Developers, *Cirq 0.12.0*, <https://doi.org/10.5281/zenodo.5182845>, Zenodo (2021).
- [159] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, *SciPy 1.0: fundamental algorithms for scientific computing in Python*, Nature Methods **17**, 261-272 (2020).
- [160] N. Hansen, Y. Akimoto, and P. Baudis, *CMA-ES/pycma: r3.2.2*, <https://doi.org/10.5281/zenodo.2559634>, Zenodo (2019).
- [161] A. Mayer and S. O. Aas, *andim/noisyopt*, <https://doi.org/10.5281/zenodo.580120>, Zenodo (2017).
- [162] J. de Nobel, D. Vermetten, H. Wang, C. Doerr, and T. Bäck, *Tuning as a means of assessing the benefits of new ideas in interplay with*

- existing algorithmic modules*, in GECCO '21: Genetic and Evolutionary Computation Conference, Companion Volume, Lille, France, July 10-14, 2021 , edited by K. Krawiec (ACM) pp. 1375-1384 (2021).
- [163] D. Vermetten, H. Wang, T. Bäck, and C. Doerr, *Towards dynamic algorithm selection for numerical black-box optimization: investigating BBOB as a use case* in GECCO '20: Genetic and Evolutionary Computation Conference, Cancún Mexico, July 8-12, 2020 , edited by C. A. C. Coello (ACM) pp. 654-662 (2020).
 - [164] I. W. Bulik, T. M. Henderson, and G. E. Scuseria, *Can single-reference coupled cluster theory describe static correlation?*, J. Chem. Theory Comput. **11**, 3171-3179 (2015).
 - [165] R. J. Bartlett, and M. Musiał, *Coupled-cluster theory in quantum chemistry*, Rev. Mod. Phys. **79**, 291 (2007).
 - [166] D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, *Multireference nature of chemistry: The coupled-cluster view*, Chem. Rev. **112**, 182-243 (2012).
 - [167] F. A. Evangelista, G. K.-L. Chan, and G. E. Scuseria, *Exact parameterization of fermionic wave functions via unitary coupled cluster theory*, J. Chem. Phys. **151**, 244112 (2019).
 - [168] G. Greene-Diniz, and D. Muñoz Ramo, *Generalized unitary coupled cluster excitations for multireference molecular states optimized by the variational quantum eigensolver*, Int. J. Quantum Chem. **121**, e26352 (2021).
 - [169] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, *Generalized unitary coupled cluster wave functions for quantum computation*, J. Chem. Theory Comput. **15**, 311-324 (2018).
 - [170] W. Mizukami, K. Mitarai, Y. O. Nakagawa, T. Yamamoto, T. Yan, and Y.-y. Ohnishi, *Orbital optimized unitary coupled cluster theory for quantum computer*, Phys. Rev. Research **2**, 033421 (2020).
 - [171] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, *Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents?*, J. Chem. Theory Comput. **152**, 124107 (2020).

- [172] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, *An adaptive variational algorithm for exact molecular simulations on a quantum computer*, Nature Communications **10**, 1-9 (2019).
- [173] D. Claudino, J. Wright, A. J. McCaskey, and T. S. Humble, *Benchmarking adaptive variational quantum eigensolvers*, Frontiers in Chemistry **8**, 1152 (2020).
- [174] Y. S. Yordanov, V. Armaos, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, *Qubit-excitation-based adaptive variational quantum eigensolver*, Communications Physics **4**, 228 (2021).
- [175] N. Gomes, A. Mukherjee, F. Zhang, T. Iadecola, C.-Z. Wang, K.-M. Ho, P. P. Orth, and Y.-X. Yao, *Adaptive Variational Quantum Imaginary Time Evolution Approach for Ground State Preparation*, Advanced Quantum Technologies, 2100114 (2021).
- [176] J. Liu, Z. Li, and J. Yang, *An efficient adaptive variational quantum solver of the Schrödinger equation based on reduced density matrices*, J. Chem. Phys. **154**, 244112 (2021).
- [177] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, *qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor*, Phys. Rev. X Quantum **2**, 020310 (2021).
- [178] Z.-J. Zhang, T. H. Kyaw, J. Kottmann, M. Degroote, and A. Aspuru-Guzik, *Mutual information-assisted adaptive variational quantum eigensolver*, Quantum Science and Technology **6**, 035001 (2021).
- [179] H. R. Grimsley, D. Claudino, S. E. Economou, E. Barnes, and N. J. Mayhall, *Is the trotterized uccsd ansatz chemically well-defined?*, J. Chem. Theory Comput. **16**, 1-6 (2019).
- [180] M.-A. Filip, and A. J. Thom, *The Best of Both Worlds: Optimizing Quantum Hardware Resources with Classical Stochastic Methods*, ArXiv:2108.10912 (2021).
- [181] QuTech and TNO, *Quantum Inspire Home*, <https://www.quantum-inspire.com> (2018).
- [182] T. Last, N. Smakharadze, P. Eendebak, R. Versluis, X. Xue, A. Sammak, D. Brousse, K. Loh, H. Polinder, G. Scappucci, M. Veldhorst,

- L. Vandersypen, K. Maturova, J. Veltin, and G. Albers, *Quantum Inspire - QuTech's platform for co-development and collaboration in quantum computing*, Proc. of SPIE **11324** (2020).
- [183] QuTech and TNO, *Quantum Inspire Starmon-5 fact sheet*, <https://qutech.nl/wp-content/uploads/2020/04/3.-Technical-Fact-Sheet-Quantum-Inspire-Starmon-5.pdf> (2020).
- [184] S. P. Jordan, *Fast Quantum Algorithm for Numerical Gradient Estimation*, Phys. Rev. Lett. **95**, 050501 (2005).
- [185] S. Lloyd, *Universal Quantum Simulators*, Science **273**, 1073-1078 (1996).
- [186] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, *Elucidating reaction mechanisms on quantum computers*, Proc. Natl. Acad. Sci. U.S.A. **114**, 7555 (2017).
- [187] D. S. Abrams, and S. Lloyd, *Simulation of Many-Body Fermi Systems on a Universal Quantum Computer*, Phys. Rev. Lett. **79**, 2586 (1997).
- [188] C. Zalka, *Simulating quantum systems on a quantum computer*, Proc. Royal Soc. Lond. A **454**, 313 (1998).
- [189] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, *Simulated Quantum Computation of Molecular Energies*, Science **309**, 1704-1707 (2005).
- [190] R. M. Dreizler, and E. K. U. Gross, *Density Functional Theory: An Approach to the Quantum Many-Body Problem*, (Springer, Berlin, Heidelberg, 1990).
- [191] I. Shavitt, and R. J. Bartlett, *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory*, (Cambridge university press, 2009).
- [192] G. H. Booth, A. J. Thom, and A. Alavi, *Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space*, J. Chem. Phys. **131**, 054106 (2009).
- [193] F. Jensen, *Introduction to computational chemistry*, 2nd ed. (John Wiley & Sons, 2007).

Bibliography

- [194] P. Norman, K. Ruud, and T. Saue, *Principles and Practices of Molecular Properties: Theory, Modeling and Simulations*, (John Wiley & Sons, 2018).
- [195] H. B. Schlegel, *Geometry optimization*, Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**, 790 (2011).
- [196] D. Marx, and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advance Methods*, (Cambridge University Press, 2009).
- [197] J. C. Tully, *Molecular dynamics with electronic transitions*, J. Chem. Phys. **93**, 1061 (1990).
- [198] J. Behler, *Perspective: Machine learning potentials for atomistic simulations*, J. Chem. Phys. **145**, 170901 (2016).
- [199] Z. Li, and W. Liu, *First-order nonadiabatic coupling matrix elements between excited states: A Lagrangian formulation at the CIS, RPA, TD-HF, and TD-DFT levels*, J. Chem. Phys. **141**, 014110 (2014).
- [200] B. F. Curchod, U. Rothlisberger, and I. Tavernelli, *Trajectory-Based Nonadiabatic Dynamics with Time-Dependent Density Functional Theory*, ChemPhysChem, **14**, 1314-1340 (2013).
- [201] S. Faraji, S. Matsika, and A. I. Krylov, *Calculations of non-adiabatic couplings within equation-of-motion coupled-cluster framework: Theory, implementation, and validation against multi-reference methods*, J. Chem. Phys. **148**, 044103 (2018).
- [202] A. Gilyen, S. Arunachalam, and N. Wiebe, *Optimizing quantum optimization algorithms via faster quantum gradient computation*, Proceedings of the 2019 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 1425-1444 (2019).
- [203] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, *Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer*, Quantum Sci. Technol. **4**, 045005 (2019).
- [204] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, *Evaluating analytic gradients on quantum hardware*, Phys. Rev. A **99**, 032331 (2019).

- [205] A. Harrow, and J. Napp, *Low-Depth Gradient Measurements Can Improve Convergence in Variational Hybrid Quantum-Classical Algorithms*, Phys. Rev. Lett. **126**, 140502 (2021).
- [206] G. G. Guerreschi, and M. Smelyanskiy, *Practical optimization for hybrid quantum-classical algorithms*, ArXiv:1701.01450, (2017).
- [207] I. Kassal, and A. Aspuru-Guzik, *Quantum algorithm for molecular properties and geometry optimization*, J. Chem. Phys. **131**, 224102 (2009).
- [208] A. Roggero, and J. Carlson, *Dynamic linear response quantum algorithm*, Phys. Rev. C **100**, 034610 (2019).
- [209] L. Visscher, *The Dirac equation in quantum chemistry: Strategies to overcome the current computational problems*, J. Comp. Chem. **23**, 759 (2002).
- [210] E. Knill, G. Ortiz, and R. D. Somma, *Optimal quantum measurements of expectation values of observables*, Phys. Rev. A **75**, 012328 (2007).
- [211] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, *Solving strongly correlated electron models on a quantum computer*, Phys. Rev. A **92**, 062318 (2015).
- [212] P.-L. Dallaire-Demers, and F. K. Wilhelm, *Method to efficiently simulate the thermodynamic properties of the Fermi-Hubbard model on a quantum computer*, Phys. Rev. A **93**, 032303 (2016).
- [213] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, *Hybrid Quantum-Classical Approach to Correlated Materials*, Phys. Rev. X **6**, 031045 (2016).
- [214] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martinez, *Quantum Computation of Electronic Transitions Using a Variational Quantum Eigensolver*, Phys. Rev. Lett. **112**, 230401 (2019).
- [215] O. Higgott, D. Wang, and S. Brierley, *Variational Quantum Computation of Excited States*, Quantum **3**, 156 (2019).
- [216] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, *Variational quantum algorithms for discovering Hamiltonian spectra*, Phys. Rev. A **99**, 062304 (2019).

- [217] R. J. Bartlett, *Coupled-cluster theory and its equation-of-motion extensions*, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 126 (2011).
- [218] I. Kassal, S. P. Jordan, P. J. Love, M. Mohseni, and A. Aspuru-Guzik, *Polynomial-time quantum algorithm for the simulation of chemical dynamics*, Proc. Natl. Acad. Sci. U.S.A., 18681 (2008).
- [219] T. Saue, L. Visscher, H. J. Aa. Jensen, and R. Bast, with contributions from V. Bakken, K. G. Dyall, S. Dubillard, U. Ekström, E. Eliav, T. Enevoldsen, E. Faßhauer, T. Fleig, O. Fossgaard, A. S. P. Gomes, E. D. Hedegård, T. Helgaker, J. Henriksson, M. Iliaš, Ch. R. Jacob, S. Knecht, S. Komorovský, O. Kullie, J. K. Lærdahl, C. V. Larsen, Y. S. Lee, H. S. Nataraj, M. K. Nayak, P. Norman, G. Olejniczak, J. Olsen, J. M. H. Olsen, Y. C. Park, J. K. Pedersen, M. Pernpointner, R. di Remigio, K. Ruud, P. Salek, B. Schimmelpennig, A. Shee, J. Sikkema, A. J. Thorvaldsen, J. Thyssen, J. van Stralen, S. Villaume, O. Visser, T. Winther, and S. Yamamoto, *DIRAC18 (v18.0)*, <https://doi.org/10.5281/zenodo.2253986>, Zenodo (2018).
- [220] B. Senjean, *Openfermion-Dirac*, (2022)
- [221] X. Fu, L. Rieseboos, M. A. Rol, J. van Straten, J. van Someren, N. Khammassi, I. Ashraf, R. F. L. Vermeulen, V. Newsum, K. K. L. Loh, J. C. de Sterke, W. J. Vlothuizen, R. N. Schouten, C. G. Almudever, L. DiCarlo, and K. Bertels, *eQASM: An Executable Quantum Instruction Set Architecture*, Proceedings of the 25th International Symposium on High-Performance Computer Architecture (HPCA'19), (2019).
- [222] A. Johnson, G. Ungaretti, and *et al.*, *QCoDeS*, (2016).
- [223] M.A. Rol, C. Dickel, S.Asaad, N.K. Langford, C.C. Bultink, R. Sagastizabal, N.K. Langford, G. de Lange, X. Fu, S.R. de Jong, F. Luthi, and W. Vlothuizen, *PycQED*, <https://doi.org/10.5281/zenodo.160327>, Zenodo (2016).
- [224] J. Nelder, and R. Mead, *A Simplex Method for Function Minimization*, Comp. J. **7**, (308) (1965).
- [225] T. H. Fischer, and J. Almöf, *General methods for geometry and wave function optimization*, J. Phys. Chem. **96**, 9768-9774 (1992).

- [226] Y. Yamaguchi, Y. Osamura, J. D. Goddard, and H. F. Schaefer III, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*, (Oxford University Press: New York, Oxford, 1994).

Samenvatting

Een quantumcomputer is een apparaat dat de wetten van de quantummechanica gebruikt om berekeningen uit te voeren. De interesse in quantumberekeningen komt voort uit de verwachting dat je daarmee sommige problemen veel sneller zou kunnen oplossen dan met klassieke computers. Men spreekt van quantumversnelling. In tegenstelling tot wat men wellicht zou denken, ligt het niet voor de hand dat quantumcomputers de klassieke computers ooit helemaal zullen vervangen. Hoogstwaarschijnlijk zullen ze geïntegreerd worden in grote supercomputers. Ze zullen worden gebruikt voor heel specifieke taken waarbij de quantumversnelling kan worden benut. In het afgelopen decennium is opmerkelijke vooruitgang geboekt bij het bouwen van prototypes van quantumcomputers, al zijn deze klein, kwetsbaar en foutgevoelig. Hoewel een grootschalige, ruisvrije quantumcomputer nog vele jaren van ons verwijderd is, verwachten we spoedig toegang te hebben tot quantumhardware die groot genoeg is om de beperkingen van de klassieke supercomputer te doorbreken. Dit proefschrift behandelt een aantal specten die van invloed zijn op kleine quantumcomputers, als zij gecombineerd worden met klassieke supercomputers.

Een van de meest uitdagende aspecten van bestaande quantumcomputers is het feit dat ze extreem foutgevoelig zijn. Als we ernaar streven nauwkeurige berekeningen van dergelijke apparaten te maken, is het belangrijk om dergelijke fouten zoveel mogelijk te verwijderen of te beperken. In hoofdstuk 2 ontwikkelen we de theorie van een strategie om fouten te beperken, gebruikmakend van het feit dat problemen in de natuurkunde en scheikunde eigenschappen hebben die onveranderd moeten blijven (bijv. aantal deeltjes, aantal spin-up/down-deeltjes, pariteit). Door een of meer van dergelijke eigenschappen te meten en te verifiëren dat ze ongewijzigd blijven ten opzichte van hun bekende waarde, kunnen we vaststellen wanneer er fouten zijn opgetreden. De uiteindelijke berekening wordt verbeterd door de berekeningen te verwijderen waarin de symmetrieën zijn gewijzigd. In hoofdstuk 3 zetten we de theorie van symmetrieverificatie aan het werk op een fysiek twee-qubit quantumapparaat voor de taak om de laagste energie van het waterstofmolecuul te berekenen. De voorgestelde foutbeperkingsstrategie laat een tienvoudige verbetering zien in de

nauwkeurigheid van de berekening.

Zelfs voor een middelgrote quantumcomputer met meer dan 50 qubits is de hoeveelheid geheugen die op een klassieke computer nodig zou zijn om de volledige beschrijving op te slaan astronomisch groot. Een quantumalgoritme moet daarom slechts een deel van de informatie gebruiken die de quantumtoestand beschrijft. In hoofdstuk 4 onderzoeken we wat en hoe we de relevante informatie van een quantumcomputer kunnen bepalen voor het berekenen van eigenschappen van fysische en chemische systemen. Onze belangrijkste bijdrage is een bewijs van het minimale aantal circuits dat hiervoor nodig is. Daarnaast bieden we een recept om circuits te construeren die exact overeenkomen met het analytisch berekende minimumaantal.

In vervolg werd de prestaties van optimalisatiealgoritmen die gebruik maken van een variatieprincipe analyseert. Een optimalisatie-algoritme is een stukje (klassieke) software die probeert de set parameters te vinden die een wiskundige functie minimaliseren (of maximaliseren). Computerwetenschappers hebben tientallen jaren van onderzoek besteed aan het ontwikkelen, bestuderen en benchmarken van optimalisatiemethoden voor een breed scala aan (klassieke) problemen. Een systematische studie van hun prestaties bij gebruik in combinatie met quantumhardware ontbreekt echter nog. Hoofdstuk 5 is een poging om deze leemte op te vullen door enkele van de meest gebruikte optimalisatiealgoritmen te vergelijken in een subset van modellen uit de natuur en scheikunde.

De afgelopen jaren hebben publieke en private quantumcomputerlaboratoria kleine quantumcomputers toegankelijk gemaakt via internetdiensten. Dankzij deze quantumcomputers “in de cloud” heeft het onderzoek naar quantumberekening een grote vlucht genomen, met honderden artikelen. In hoofdstuk 6 doen we een klein experiment met de Nederlandse quantumcomputer in de cloud, genaamd Quantum Inspire. We vinden dat de communicatie via het internet gepaard gaat met grote vertragingen, die op hun beurt de rekenprestaties beïnvloeden vanwege de korte levensduur van de qubits, zelfs voor een heel klein probleem. Daarom verwachten we dat het uitvoeren van grote berekeningen op cloudgebaseerde quantumcomputers moeilijk haalbaar zal zijn.

In het laatste hoofdstuk van het proefschrift onderzoeken we een toepassing van quantumcomputers in de quantumchemie, de berekening van de afgeleide van de energiefunctie. We beschrijven een methode om de afgeleide te berekenen in bestaande quantumhardware. De methode wordt vervolgens gebruikt om de geometrie van het model van het waterstofmolecuul en de polariseerbaarheid ervan te optimaliseren met een realistisch twee-qubit-apparaat. Ten slotte verifiëren we de resultaten van het experi-

ment met uitgebreide simulaties, waarbij we een heel goede overeenkomst vinden in zowel de simulatie van de chemie als de simulatie van het experiment.

Summary

A quantum processing unit or quantum computer is a device that uses the laws of quantum mechanics to perform calculations. The interest in quantum computation and quantum algorithms arises from the fact that these machines can solve some problems much faster than classical computers, what is commonly known as quantum speed-up. Contrary to what one might think, quantum computers will never replace classical computers. They will, most likely, become integrated within large supercomputers and will be used for purpose-specific tasks where the quantum speed-up can be exploited. In the past decade outstanding progress has been achieved in building prototypes of quantum computers, albeit small, fragile and error prone. While a large-scale, noise-free quantum computer is still years away, we expect to have access to quantum hardware sufficiently large to challenge the limits of classical supercomputer. This thesis covers most of the aspects that affect noisy quantum computers when combined with classical computers.

One of the most challenging aspects of existing quantum computers is the fact that they are extremely error-prone. If we aim at providing accurate calculations from such devices it is important to remove or mitigate errors as much as possible. In chapter 2 we develop the theory of a strategy to mitigate errors using the fact that problems in physics and chemistry have known properties that must remain unchanged (e.g. number of particles, number of spin-up/down particles, parity, amongst others). By measuring one or more of such properties and verifying that they remain unchanged from their known value we are able to identify when errors have occurred. The final computation is improved by removing the calculations in which the symmetries have been changed. In chapter 3 we put the theory of symmetry verification to work on a real two-qubit quantum device for the task of calculating the lowest energy of the hydrogen molecule. The proposed error mitigation strategy shows a ten-fold improvement in the accuracy of the calculation.

Even for a moderate size quantum computer of ≥ 50 qubits, the amount of classical memory required to store its full description is astronomical. As such it is unfeasible to extract all the information of a quantum computer. A quantum algorithm must, therefore, use only partial information

Summary

of the quantum state. In chapter 4 we explore what and how to extract the relevant information of a quantum computer for calculating properties of physical and chemical systems. Our main contribution is a proof on the minimum number of circuits required to extract properties beyond the energy for chemistry and physics problem from their reduced density matrices. Additionally, we provide a recipe to construct measurement circuits that exactly match the minimum number analytically calculated for 2-reduced density matrices.

The journey continues with the analysis of the performance of optimization algorithms in the context of variational quantum algorithms. An optimization algorithm is a piece of (classical) software that attempts at finding the set of parameters that minimize (or maximize) a mathematical function. Computer scientist have devoted decades of research to develop, study and benchmark optimization methods for a wide variety of (classical) problems. However, a systematic study of their performances when used in combination with quantum hardware is still lacking. Chapter 5 is an attempt of filling up this gap by comparing some of the most used optimization algorithms in a subset of toy models in physics and chemistry.

In recent years public and private quantum computer laboratories have made small quantum computers accessible via internet services. Thanks to these cloud-based quantum devices research in quantum computation has democratized, with hundreds of articles being published using them. In chapter 6 we run a toy experiment with the Dutch public quantum computer infrastructure Quantum Inspire. We find that internet communication comes with large delays, which in turn affect the performance of computation because of the short lifetime of the hardware, even for such a very small problem. Therefore we anticipate that performing large calculation on cloud-based quantum computers will be problematic.

In the last chapter of the thesis we explore an application of quantum computers in quantum chemistry, the calculation of energy derivatives. We describe a method to compute energy derivatives in near-term quantum hardware. The method is then used to optimize the geometry of the model of the hydrogen molecule as well as its polarizability with a real two-qubit device. Finally we verify the results of the experiment with extensive simulations finding an really good agreement in both, the simulation of the chemistry and the simulation of the experiment.

Resum

Una unitat de processament quàntica o ordinador quàntic és un aparell de computació que utilitza les lleis de la mecànica quàntica per a executar càlculs. L'interés en la computació i algorísmia quàntica és deu al fet que els ordinadors quàntics poden resoldre alguns problemes molt més ràpidament que els ordinadors clàssics, fet que es coneix com acceleració quàntica. Contràriament al que es puga pensar, els ordinadors quàntics mai reemplaçaran els ordinadors clàssics. Probablement, les unitats de processament quàntiques seran integrades dins de grans supercomputadores, i seran utilitzades solament per a tasques específiques on l'acceleració quàntica puga ser explotada. L'última dècada s'ha caracteritzat per un progrés inaudit en la construcció de prototips d'ordinadors quàntics, encara que menuts (en nombre de qubits), fràgils i erronis. Tot i que encara que falten varies dècades per a veure l'aparició d'un ordinador quàntic a gran escala sense errors, creiem que en els pròxims anys podrem tenir a l'abast ordinadors quàntics amb un nombre de qubits suficient de tal forma que els superordinadors clàssics més potents no puguin simular-los. Aquesta tesi tracta gran part dels aspectes dels ordinadors quàntics amb errors quan es combinen amb ordinadors clàssics.

Un dels reptes més grans per a l'ús dels ordinadors quàntics existents és el fet que aquestes màquines són extremadament errònies. Si volem fer computacions utilitzant aquesta tecnologia és de vital importància eliminar, o al menys, reduir el nombre d'errors que ocorren durant l'execució dels algorismes. Al segon capítol desenvolupem la teoria per a mitigar errors utilitzant el fet que els problemes en física i química tenen propietats conegudes que han de ser constants (per exemple el nombre de partícules, la direcció del seu spin (up o down) o la seua paritat, entre d'altres). L'estratègia de reducció d'errors funciona mesurant una o més d'aquestes propietats, i verificant que el seu estat roman inalterat al final del càlcul. El resultat de la computació és millorat quan els resultats que senyalen que ha hagut un error són eliminats al final del procés. Al tercer capítol utilitzem la mitigació d'errors per verificació de simetries en un experiment amb un ordinador quàntic real de dos qubits, amb l'objectiu de calcular l'energia mínima de la molècula d'hidrogen. L'estratègia de mitigació d'errors per verificació de simetries fa que el resultat final siga

10 voltes més precis que sense la redució d'errors.

Un ordinador quàntic d'uns ≥ 50 qubits pot produir un estat quàntic la descripció del qual requereix un nombre astronòmic d'espai en la memòria d'un ordinador clàssic. Per tant, extraure tota la informació d'un estat quàntic d'aquesta mida és impossible. En un algorisme quàntic és necessari extraure informació de l'estat quàntic, però aquesta informació sols pot ser parcial. Al quart capítol explorem com i quina és la informació rellevant de l'estat quàntic per a calcular propietats de sistemes físics i químics. La principal contribució d'aquest capítol és una demostració analítica del mínim nombre de circuits de mesura necessaris per a extraure les propietats de sistemes físics i químics més enllà de les energies, utilitzant les propietats de les matrius de densitat reduïdes. A més a més, hi descrivim un mètode pel qual arribem al mínim nombre de circuits per al cas de les matrius de densitat reduïdes d'ordre 2.

La tesi continua amb una anàlisi del rendiment d'alguns algorismes d'optimització en el context d'algorismes quàntics variacionals. Un algorisme d'optimització és un software clàssic que té com a finalitat trobar els paràmetres que minimitzen (o maximitzen) una funció matemàtica de forma numèrica. Els enginyers informàtics porten dècades desenvolupant, estudiant i caracteritzant el rendiment de mètodes d'optimització numèrica en una gran varietat de problemes clàssics. Tanmateix, encara no existeix un estudi sistemàtic del rediment dels optimitzadors numèrics en algorismes quàntics variacionals. El cinquè capítol és un intent d'omplir aquest buit, on hi comparem alguns dels optimitzadors més populars en un conjunt de problemes modelics en física i química.

En els últims anys hem vist una expansió dels servicis de computació quàntica a través de la xarxa, gràcies als laboratoris públics i privats que han ficat les seues unitats de processament quàntiques al servici de tothom. Això ha produït una democratització de la recerca en computació quàntica, reflectida amb la publicació de centenars d'articles en aquest camp. Al sisè capítol utilitzem el servici públic de computació quàntica dels Països Baixos, Quantum Inspire, per a fer un xicotet experiment. Els nostres resultats demostren que la interacció entre un ordinador quàntic en el núvol i un ordinador clàssic local a través de la xarxa té un sobrecoast de comunicació que fa que els resultats siguin molt dolents, ja que la vida de l'ordinador és molt curta. Per tant, anticipem que executar grans càlculs en un ordinador quàntic al núvol serà problemàtic.

En l'últim capítol de la tesi explorem aplicacions dels ordinadors quàntics en química quàntica, en particular, el càlcul de derivades de l'energia. Ací describim un mètode per a calcular aquestes derivades en un ordinador quàntic que pot existir en un futur no molt llunyà. A més, utilitzem

aquest mètode per a calcular la geometria i la polaritzabilitat del model de la molècula d'hidrogen en un ordinador quàntic real de dos qubits. Finalment, fem ús de simulacions detallades per a verificar els resultats obtinguts a l'experiment, trobant una gran concordança tant amb la química del sistema, com entre les simulacions i l'experiment.

Curriculum Vitæ

I was born on the 29th of March 1991 in Xàtiva, València. I attended primary school in Manuel (València), and received my high school education at the institut d'educació pública Sivera Font in Canals (València). In parallel to my high school education I received music education at Conservatori Professional de Música Mestre Vert in Carcaixent (València), completing the professional degree of music with honors in the specialization of trombone.

In 2009, I started the degree of physics at Universitat de València. I attended the Universiteit Groningen during the final year of my bachelor studies through the Erasmus exchange programme. There, I completed my bachelor's research thesis under the supervision of Prof. Dr. ir. Caspar van der Wal. I received my bachelor of science degree in 2014.

Before starting my master's degree I spent two years in Ireland where I worked as information and technology support specialist.

In 2016 I started the master's program of physics at Universiteit Leiden. I completed my Master's thesis "*Comparison of error mitigation strategies in a hydrogen molecule quantum simulation*" under the supervision of Dr. T. E. O'Brien. I obtained my Master's degree cum laude in 2018.

In April 2018, I started my Ph.D. studies under the supervision of Dr. T. E. O'Brien and Prof. dr. C. W. J. Beenakker at the Instituut-Lorentz of Universiteit Leiden. Part of my research was done in the laboratory of Prof. dr. L. DiCarlo at TU Delft. My research focused on the practical implementation of quantum algorithms on real quantum processing units. This thesis covers most of the issues that existing quantum computers have: quantum noise, quantum state tomography, optimization of parametrized quantum circuits, and experimental applications in small quantum devices. During my Ph.D. studies I was teaching assistant to the Master's degree courses of Quantum Information, Quantum Mechanics and Applied Quantum Algorithms at Universiteit Leiden. I was also teaching assistant to the Master's degree course Fundamentals of Quantum Information at TU Delft. Additionally, I supervised bachelor and master students together with Dr. B. Senjean, Dr. H. Wang and Dr. J. Tura. I attended several international schools and conferences, presenting my work in the USA, France, Spain and The Netherlands. I

Curriculum Vitæ

was invited to give a talk at the workshop “*New Trends in Computational Chemistry*” held in Barcelona in September 2022.

List of publications

- X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Bäck, V. Dunjko and T. E. O'Brien, *Performance comparison of optimization methods on variational quantum algorithms* ArXiv:2111.13454 (2021) submitted to Physical Review A. [Chapter 5]
- X. Bonet-Monroig, R. Babbush and T. E. O'Brien, *Nearly optimal measurement scheduling for partial tomography of quantum states*, Physical Review X **10**, 031064 (2020). [Chapter 4]
- T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo and L. Visscher, *Calculating energy derivatives for quantum chemistry on a quantum computer* npj Quantum Information **5**, 113 (2019). [Chapter 7]
- 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*, Physical Review A **100**, 010302(R) (2019). [Chapter 3]
- X. Bonet-Monroig, R. Sagastizabal, M. Singh and T. E. O'Brien, *Low-cost error mitigation by symmetry verification*, Physical Review A **98**, 062339 (2018). [Chapter 2]
- J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. Schuyler-Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. E. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. P. D. Sawaya, B. Senjean, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang and R. Babbush, *OpenFermion: the electronic structure package for quantum computers*, Quantum Science and Technology **5**, 034014 (2020)

Stellingen

behorende bij het proefschrift

Optimization of quantum algorithms for near-term quantum computers

1. Symmetry verification provides the minimum amount of error mitigation at the lowest computational cost possible.

Chapter 2, 3

2. A binary partition strategy allows one to sample all k-local qubit operators in an N-qubit system in poly-log(N) time.

Chapter 4

3. Surrogate models can improve the sampling noise floor in variational quantum algorithms.

Chapter 5

4. Cloud-based quantum computers might be successful under the edge computing paradigm.

Chapter 6

5. The claim in Phys. Rev. Research **3**, 023092 (2021) that a symmetry-breaking ansatz in a variational quantum algorithm provides a faster convergence is not true in general.

6. The accuracy required for quantum chemistry calculations can not be achieved with variational quantum algorithms.

D. Wecker, M. B. Hastings, and M. Troyer,
Phys. Rev. A **2**, 042303 (2015).

7. There exist a heuristic separation between classical and quantum cost functions.

8. Quantum computing is already providing a practical computational speed-up by motivating the discovery of better classical algorithms.

9. Academic ability is just one of many forms of intelligence, yet it is not the most important one.

Xavier Bonet-Monroig

2 november 2022