Quantum number preserving ansätze and error mitigation studies for the variational quantum eigensolver

INAUGURAL-DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT DER UNIVERSITÄT ZU KÖLN VORGELEGT VON

Gian-Luca Romano Anselmetti

aus Basel

unter der Aufsicht von

Prof. Dr. David GROSS
Dr. Christian GOGOLIN

veröffentlicht in Köln im September 2023

Berichterstatter (Gutachter:)	Prof. Dr. David Gross
	PD Dr. Rochus Klesse

Mündliche Prüfung am 08.09.2023

Computational chemistry has advanced rapidly in the last decade on the back of the progress of increased performance in CPU and GPU based computation. The prediction of reaction properties of varying chemical compounds in silico promises to speed up development in, e.g., new catalytic processes to reduce energy demand of varying known industrial used reactions. Theoretical chemistry has found ways to approximate the complexity of the underlying intractable quantum many-body problem to various degrees to achieve chemically accurate ab initio calculations for various, experimentally verified systems. Still, in theory limited by fundamental complexity theorems accurate and reliable predictions for large and/or highly correlated systems elude computational chemists today. As solving the Schrödinger equation is one of the main use cases of quantum computation, as originally envisioned by Feynman himself, computational chemistry has emerged as one of the applications of quantum computers in industry, originally motivated by potential exponential improvements in quantum phase estimation over classical counterparts. As of today, most rigorous speed ups found in quantum algorithms are only applicable for so called error-corrected quantum computers, which are not limited by local qubit decoherence in the length of the algorithms possible. Over the last decade, the size of available quantum computing hardware has steadily increased and first proof of concepts of error-correction codes have been achieved in the last year, reducing error rates below the individual error rates of qubits comprising the code. Still, fully error-corrected quantum computers in sizes that overcome the constant factor in speed up separating classical and quantum algorithms in increasing system size are a decade or more away. Meanwhile, considerable efforts have been made to find potential quantum speed ups of non-error corrected quantum systems for various applications in the noisy intermediate-scale quantum (NISQ) era. In chemistry, the variational quantum eigensolver (VQE), a family of classical-quantum hybrid algorithms, has become a topic of interest as a way of potentially solving computational chemistry problems on current quantum hardware.

The main contributions of this work are: extending the VQE framework with two new potential ansätze, (1) a maximally dense first-order trotterized ansatz for the paired approximation of the electronic structure Hamiltonian, (2) a gate fabric with many favourable properties like conserving relevant quantum numbers, locality of individual operations and potential initialisation strategies mitigating plateaus of vanishing gradient during optimisation. (3) Contributions to one of largest and most complex VQE to date, including the aforementioned ansatz in paired approximation, benchmarking different error-mitigation techniques to achieve accurate results, extrapolating performance to give perspective on what is needed for NISQ devices having potential in competing with classical algorithms and (4) Simulations to find optimal ways of measuring Hamiltonians in this error-mitigated framework. (5) Furthermore a simulation of different purification error mitigation techniques and their combination under different noise models and a way of efficiently calibrating for coherent noise for one of them is part of this manuscript. We discuss the state of VQE after almost a decade after its introduction and give an outlook on computational chemistry on quantum computers in the near future.

Declaration

”I hereby declare that I have completed the present dissertation independently and without the use of any aids or literature other than those referred to. All passages that have been taken, either literally or in sense, from published and unpublished works, are marked as such. I declare that this dissertation has not been submitted to any other faculty or university; that - apart from the partial publications and included articles and manuscripts listed below - it has not yet been published, and that I will not publish the dissertation before completing my doctorate without the permission of the PhD Committee. I am aware of the terms of the doctoral regulations. In addition, I hereby declare that I am aware of the “Regulations for Safeguarding Good Scientific Practice and Dealing with Scientific Misconduct” of the University of Cologne, and that I have observed them during the work on the thesis project and the written doctoral thesis. I hereby commit myself to observe and implement the guidelines mentioned there in all scientific activities. I assure that the submitted electronic version is identical to the submitted printed version”.

Partial publications

- O’Brien, T. E., **Anselmetti, G.**, Gkritsis, F., Elfving, V. E., Polla, S., Huggins, W. J., ... Rubin, N. C. (2022, October). Purification-based quantum error mitigation of pair-correlated electron simulations. arXiv. Retrieved 2023-05-31, from `http://arxiv.org/abs/2210.10799` (arXiv:2210.10799 [quant- ph]) doi: 10.48550/arXiv.2210.1079, Accepted in Nature Physics
- **Anselmetti, G.-L. R.**, Wierichs, D., Gogolin, C., Parrish, R. M. (2021, November). Local, expressive, quantum-number-preserving VQE ans atze for fermionic systems. New Journal of Physics, 23 (11), 113010. Retrieved 2023-04-06, from https://dx.doi.org/10.1088/1367-2630/ac2cb3 (Publisher: IOP Publishing) doi: 10.1088/1367-2630/ac2cb
- Polla, S., **Anselmetti, G.-L. R.**, O’Brien, T. E (2022, July). Optimizing the information extracted by a single qubit measuremen. 10.48550/arXiv.2207.094. arXiv:2207.0947.

Gian-Luca Romano Anselmetti, Cologne 19th of June 2023

Acknowledgments

First, I'm extremely grateful to Christian Gogolin for taking me on the promised 'wild ride' of the last four years through quantum chemistry and quantum computing, supervising my work. The countless things he taught me from professional software development over quantum many-body physics all the way down to managing oneself as an independent researcher have allowed me to grow on so many aspects as person and as a scientist. I very much enjoyed the last years of work we got to spent together. On the academic side, a big thank you to David Gross for supervising my endeavours from the University of Cologne. He has been tremendously helpful, without this assistance this work would not have been possible. Maximilian Scheurer shall also not go unmentioned here as his contribution to my understanding of chemistry cannot be understated, I am very thankful to have been working with you. Furthermore, I would like to thank my collaborators from Google Quantum AI and QCWare for the many discussions and insights they gave and the possibility of collaboration on these state of the art experiments and methods. My PhD colleagues Oumarou Oumarou and Fotios Gkritsis on the Covestro side have also made my time memorable, thank you for the support.

As not only my work and academic pears have influenced this work and myself, I would like to express my deepest gratitude to my partner for putting up with me especially towards the end of this PhD and the continuing support over the years. The support from my friends and family during the last years has also been very much appreciated, thank you all for walking this path with me.

Contents

1 Classical quantum chemistry

Since the discovery of quantum mechanics and the proclamation of the Schrödinger equation, the fundamental mechanisms of the interaction of nuclei and electrons have been understood and its complexity has been captured in the *electronic structure problem.* Even having discovered the equations governing the nature of chemistry, the exponential scaling of complexity of the resulting quantum many-body problem limits the extent of what is computable and in this sense knowable about chemical properties of molecular systems. The art of finding reasonable approximations to the intractable original problem, easing restrictions in demand of computational resources while still remaining accuracy to predict chemically accurate behaviours is at the heart of the field of (classical) quantum chemistry over the last 90 years [1]. Ab-initio chemistry methods that only take physical constants as inputs have grown with the available computational hardware from mean-field methods like Hartree-Fock [2] restricted to approximations without any quantum mechanical correlation to more sophisticated methods being able to calculate relevant properties like absolute and relative energies of given molecules, electronic charge density distributions, dipoles and higher multipole moments, vibrational frequencies or other spectroscopic quantities to high precision. These methods enable quantum chemists to make predictions that have come to rival experimental investigations into these systems [3]. Algorithms like density-functional-theory (DFT) [4] have enabled large scale simulations of chemically relevant systems.

The first section of this thesis will briefly cover the formulation of the electronic structure problem and the traditional approximations and treatments developed to tackle solutions, as these methods developed over the decades have heavily influenced the theory developed of simulating these systems on quantum computing hardware instead. For more detailed introductions of the foundations refer to [5, 6].

1.1 Quick introduction into quantum chemistry

As this work borders quantum chemistry and quantum many body physics, one has to be careful has both fields have developed different notations and terminologies for identical mathematical concepts. This section contains a brief introduction to notation and the foundation of the theory and concepts used throughout this work, lending from both fields and try to bridge them in some cohesive form.

Computational chemistry is aiming to resolve energies with an error ΔE up to 1.6×10^{-3} Hartree in accuracy, which is often referred to as *chemical accuracy.* This is chosen as a threshold because it is the energy landscape where one can model energy reactions at room temperature and estimate reaction rates $\propto e^{-\Delta E/k_B T}$ with the temperature T and the Boltzmann constant k_B up to an order of magnitude of precision using Eyrings equation [7, 8]. As one is interested in energy *differences*, individual energies often do not have to be calculated with chemical accuracy but as long as a fortuitous cancellation of error occurs, one can recover energy barriers with the needed precision.

The electronic Schrödinger equation in first quantization after the Born-Oppenheimer approximation with electron coordinates $\mathbf{r}$ and nuclear coordinates R is described by

$$\left[\underbrace{\underbrace{-\frac{1}{2}\sum_i \nabla_i^2}_{\hat{T}_e(\mathbf{r})} \underbrace{-\sum_{A,i}\frac{Z_A}{r_{Ai}}}_{\hat{V}_{eN}(\mathbf{r},\mathbf{R})}}_{\sum_i \hat{h}_i} + \underbrace{\sum_{A>B}\frac{Z_A Z_B}{R_{AB}}}_{\hat{V}_{NN}(\mathbf{R})} + \underbrace{\sum_{i>j}\frac{1}{r_{ij}}}_{\hat{V}_{ee}(\mathbf{r})} \right]\Psi(\mathbf{r};\mathbf{R}) = E_{el}\Psi(\mathbf{r};\mathbf{R}) \tag{1}$$

with the distances between electrons and nuclei $r_{Ai} = |r_i - R_A|$, between nuclei $R_{AB} = |R_A - R_B|$, between electrons $r_{ij} = |r_i - r_j|$ and the atomic number of the nucleus A Z_A. Short hand notations are introduced for the kinetic part of the electron Hamiltonian $\hat{T}_e$, the three Coulombic potential terms for electron-nuclear interaction electronic $\hat{V}_{eN}$, nuclear-nuclear interaction $\hat{V}_{NN}$ and electron-electron interaction $\hat{V}_{ee}$, The electron nuclei coulombic term is usually combined with the kinetic term to form the single electron operator $\hat{h}_i$, the nuclear Coulomb interaction in the Born-Oppenheimer approximation amounts to a constant which is usually referred to as the *core energy* and most of the complexity is captured in the electron-electron interaction term due to the fermionic nature of the electron.

In the usual introduction, one might first try to approximate the problem of non-interacting fermions, ignoring the two-body term of the Hamiltonian. In this case the wave function is separable into a product state of spatial orbitals $\phi_i(r)$. One can think of these as spatial orbitals as the solutions to the one-electron problem of the local Hamiltonian. Further, there are four degrees of freedom of these spatial orbitals, three continuous degrees of space $\mathbf{r}$ and one discrete degree of spin w, $\mathbf{x} \to \{\mathbf{r}, w\}$ explicitly to introduce the preferred basis of spin orbitals $\phi(\mathbf{r})\alpha(w) = \chi(\mathbf{x})$ where $\alpha(w)$ is either the spin-up or -down (alpha or beta in chemists notation) component of the spatial orbital. The separable wave function can then be written as the product

$$\Psi_{HP}\left(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\right) = \chi_1\left(\mathbf{x}_1\right)\chi_2\left(\mathbf{x}_2\right)\cdots\chi_N\left(\mathbf{x}_N\right), \tag{2}$$

which is also referred to as the *Hartree product* [5].

One quickly sees that this in general does not obey the anticommutation needed for a fermionic wave function $\Psi(x_1, x_2) = -\Psi(x_2, x_1)$, so one has to take care of antisymmetricity to make this a valid fermionic wave function. One way to do this is to use the natural anticommutation of a matrix determinant to write the wave function as so called Slater determinant [5]

$$\Psi_{\mathrm{SD}}(\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_N}) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \chi_1\left(\mathbf{x}_1\right) & \chi_2\left(\mathbf{x}_1\right) & \cdots & \chi_N\left(\mathbf{x}_1\right) \\ \chi_1\left(\mathbf{x}_2\right) & \chi_2\left(\mathbf{x}_2\right) & \cdots & \chi_N\left(\mathbf{x}_2\right) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1\left(\mathbf{x}_N\right) & \chi_2\left(\mathbf{x}_N\right) & \cdots & \chi_N\left(\mathbf{x}_N\right) \end{vmatrix}, \tag{3}$$

which antisymmetrizes the product in $\mathcal{O}(N^3)$ steps. Interchanging two particles corresponds to exchanging two rows in the determinant which gives back the correct sign of the wave function. As it is quite lengthy to write out the entire matrix determinant every time one wants to work with Slater determinants one has introduced a shorter notation which only specifies the Trace of the matrix and implies the normalisation factor

$$\Psi_{\mathrm{SD}}(\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_N}) = \left|\chi_i\left(\mathbf{x}_1\right)\chi_j\left(\mathbf{x}_2\right)\cdots\chi_k\left(\mathbf{x}_N\right)\right\rangle = \left|\chi_i\chi_j\cdots\chi_k\right\rangle, \tag{4}$$

where in the second part of the equation an order of $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_N}$ as been assumed. The ordering in the shorthand notation takes care of the antisymmetry as a change of rows in the determinant changes the order of the labels in the shorthand notation

$$\left|\cdots\chi_i\cdots\chi_j\cdots\right\rangle = -\left|\cdots\chi_j\cdots\chi_i\cdots\right\rangle. \tag{5}$$

In second quantization one finds another representation of these Slater determinants by introducing fermionic creation and annihilation operators $a_i^\dagger, a_i$ which create a electron in the spin orbital χ_i. One postulates the fermion vacuum state $|0\rangle$ and defines the operators by the following relations [5]

Fermionic operators

$$c_\alpha|0\rangle = 0 \quad \{a_i, a_j\} = 0 = \left\{a_i^\dagger, a_j^\dagger\right\} \tag{6}$$

$$\langle 0|0\rangle = 1 \quad \left\{a_i, a_j^\dagger\right\} = \langle\chi_i \mid \chi_j\rangle = \delta_{i,j} \tag{7}$$

where $\{,\}$ is the anticommutator and $\langle\chi_i|\chi_j\rangle = \delta_{i,j}$ holds as long as one has chosen an orthonormal set of spin orbitals. When pairing electrons like e.g. in Bardeen-Cooper-Schriefer theory of superconductivity [9], the paired approximation used in Section 4 as a stepping stone to the full electronic structure Hamiltonian or just Spin-systems like qubits in general, one deals with excitations of hard-core bosonic nature, meaning they cannot occupy the same mode more than once but are distinguishable in comparison to the original fermionic excitations [10].

Hardcore bosonic operators

$$\begin{aligned} \left[\hat{b}_p, \hat{b}_q^\dagger\right] = \left[\hat{b}_p^\dagger, \hat{b}_q^\dagger\right] = \left[\hat{b}_p, \hat{b}_q\right] &= 0 \quad (p \neq q) \\ \left\{\hat{b}_p\dagger, \hat{b}_p^\dagger\right\} = \left\{\hat{b}_p, \hat{b}_p\right\} &= 0 \\ \left\{\hat{b}_p, \hat{b}_p^\dagger\right\} &= 1 \end{aligned} \tag{8}$$

One can specify Slater determinants by creating them from the true vacuum

$$|\chi_i \chi_j \cdots \chi_k\rangle = a_i^\dagger a_j^\dagger \cdots a_k^\dagger |0\rangle \quad \forall \quad i < j < k, \tag{9}$$

where the anticommutation now has been taken care of the anticommuting properties of the creation operators.

The first quantized Hamiltonian in second quantization takes the following form [11]

$$H = V_{NN} + \sum_{pq} h_{pq} a_p^\dagger a_q + \sum_{pqrs} V_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \tag{10}$$

where the matrix elements are defined by the one and two body electron integrals

$$h_{ij} = \int dr \chi_i^*(r) \hat{h}_i \chi_j(r) = \int dr \chi_i^*(r) \left(-\nabla^2(r) + \sum_A \frac{Z_A}{r_{Ai}} \right) \chi_j(r), \tag{11}$$

$$V_{ijkl} = \frac{1}{2} \iint dr_1 dr_2 \chi_i^* \left(r_1\right) \chi_j^* \left(r_2\right) \frac{1}{r_{ij}} \chi_k \left(r_2\right) \chi_l \left(r_1\right). \tag{12}$$

Chemists have introduced notation to keep track of electronic integrals in a convenient way [5]. For **spin orbitals**, the following notation has been introduced:

$$\begin{aligned}
[i|h|j] =& \langle i|h|j\rangle = \int d\mathbf{x}_1 \chi_l^* \left(\mathbf{x}_1\right) h \left(\mathbf{r}_1\right) \ell_J \left(\mathbf{x}_1\right) \\
\langle ij|kl\rangle =& \langle \chi_i \chi_J | \chi_k \chi_l \rangle = \int d\mathbf{x}_1 d\mathbf{x}_2 \chi_l^* \left(\mathbf{x}_1\right) \chi_J^* \left(\mathbf{x}_2\right) r_{12}^{-1} \chi_k \left(\mathbf{x}_1\right) \chi_l \left(\mathbf{x}_2\right) = [ik \mid jl] \\
\langle ij||kl\rangle =& \langle ij|kl\rangle - \langle ij|kl\rangle = \int d\mathbf{x}_1 d\mathbf{x}_2 \chi_l^* \left(\mathbf{x}_1\right) \alpha_J^* \left(\mathbf{x}_2\right) r_{12}^{-1} \left(1 - P_{12}\right) \chi_k \left(\mathbf{x}_1\right) \chi_l \left(\mathbf{x}_2\right)
\end{aligned} \tag{13}$$

$\langle ij||kl\rangle$ is an already antisymmetrized two-electron integral, a quite useful construction because of its frequent appearance in derivations, where the operator P_{12} exchanges coordinates of electrons one and two. As the ordering in these cases is arbitrary and agreed on, another ordering and therefore another notation is also popular, often encountered in Hartree-Fock theory which is referred to as the **chemists notation** $[ij|kl] = \langle ik|jl\rangle$. This is not used in this thesis but one should be aware of its existence. For **spatial orbitals** different brackets are used to note the different set of orbitals [5]

$$\begin{aligned}
(i|h|j) = h_{ij} = \left(\Psi_l|h|\Psi_J\right) = \int d\mathbf{r}_1 \Psi_l^* \left(\mathbf{r}_1\right) h \left(\mathbf{r}_1\right) \Psi_J \left(\mathbf{r}_1\right), \\
(ij|kl) = \left(\Psi_l \Psi_J | \Psi_k \Psi_l\right) = \int d\mathbf{r}_1 d\mathbf{r}_2 \Psi_l^* \left(\mathbf{r}_1\right) \Psi_J \left(\mathbf{r}_1\right) r_{12}^{-1} \Psi_k^* \left(\mathbf{r}_2\right) \Psi_l \left(\mathbf{r}_2\right),
\end{aligned} \tag{14}$$

where $J_{ij} \equiv (ii|jj)$ is referred to as the *Coulomb integral*, related to electronic repulsion and the *exchange integral* $K_{ij} \equiv (ij|ij)$ enforcing the Pauli-exclusion principle. The overlap matrix S_{ij} is defined as

$$S_{ij} = \langle \chi_i | \chi_j \rangle \tag{15}$$

as the basis set $\{\chi_j\}$ is in general not orthonormal, e.g. throughout optimisation procedures, and having nonzero overlap. The energy under this Hamiltonian from Eq. 10 is therefore

$$E_\psi = \langle\psi| H |\psi\rangle = \sum_{pq} h_{pq} {}^1D_q^p + \sum_{pqrs} V_{pqrs} {}^2D_{rs}^{pq}, \tag{16}$$

with the one particle reduced density matrix ${}^1D_j^i$ (1-RDM) and the two-particle reduced density matrix ${}^2D_{lk}^{ij}$ (2-RDM) defined by [11]

$$\begin{aligned}
{}^1D_p^q &= \langle\psi| a_p^\dagger a_q |\psi\rangle, \\
{}^2D_{rs}^{pq} &= \langle\psi| a_p^\dagger a_q^\dagger a_s a_r |\psi\rangle,
\end{aligned} \tag{17}$$

and for completeness sake the definition of a k-RDM [12]

$$ {}^{k}D^{\vec{p}}_{\vec{q}} = \langle\psi| a^{\dagger}_{p_1} \cdots a^{\dagger}_{p_k} a_{q_k} \cdots a_{q_1} |\psi\rangle . \tag{18} $$

The formulation in RDMs allows to define observables and other functions of given wave function by their order, so every 2-order operator can be calculated by knowledge of the 1- and 2-RDM.

A special case arises for the k-RDMs of pure Slater determinants, as they do not have contributions from electron correlation, any k-RDM is purely a function of the 1-RDM [13]

$$ {}^{k}D^{\vec{p}}_{\vec{q}} = \langle\Psi_{\mathrm{SD}}| a^{\dagger}_{p_1} \cdots a^{\dagger}_{p_k} a_{q_k} \cdots a_{q_1} |\Psi_{\mathrm{SD}}\rangle = \begin{vmatrix} {}^{1}D^{q_1}_{p_1} & \cdots & {}^{1}D^{q_k}_{p_1} \\ \vdots & \ddots & \vdots \\ {}^{1}D^{q_1}_{p_k} & \cdots & {}^{1}D^{q_k}_{p_k} \end{vmatrix} . \tag{19} $$

1.2 The Hartree-Fock method and common expansions

As a result of second quantization, the original problem has been split into two parts - finding a set of orbitals for the given molecule and solving the second quantized Hamiltonian for this configuration. A set of orbitals can be found classically using the Hartree-Fock formalism, while the solutions to of the second quantized Hamiltonian can now be treated on a quantum computer, either by variational eigensolvers discussed in the present work (Section 2) or solved in polynomial time on a fault-tolerant quantum computer. The Hartree-Fock formalism to that is used here to find a good set of orbitals is one of the cornerstones of quantum chemistry. As an efficient method of calculating the energy of a molecule with only mean-field electron correlation, referred to as the Hartree-Fock energy E_{HF} which seeks to find a single Slater determinant, or the set of molecular orbitals that minimizes the energy of the electronic structure Hamiltonian. This energy and set of orbitals is used as a starting point for so called *post-Hartree-Fock methods* that include correlation energy (which is defined as the difference from Hartree-Fock to the true ground state received from diagonalising the Hamiltonian $E_{\mathrm{FCI}} = E_{\mathrm{HF}} + E_{\mathrm{corr}}$ as it is in the present work. A trial wave function $|\psi\rangle = \sum_i c_i |\chi_i\rangle$ is formed from a linear combination of the molecular orbitals (see Section 1.3). The task becomes to minimize the energy [5]

$$ E_{\mathrm{HF}} = \langle\psi| H |\psi\rangle = \sum_{ij} c_i^* c_j \langle\chi_i| H |\chi_j\rangle , \tag{20} $$

with the constraint that the wave function stays normalized [5]

$$ \langle\psi|\psi\rangle - 1 = \sum_{ij} c_i^* c_j \langle\chi_i| H |\chi_j\rangle - 1 = 0. \tag{21} $$

After reforming the optimization problem as a Lagrangian and including the constraint with a Lagrangian multiplier, one can recover a generalized eigenvalue problem [5]

$$ \hat{f} |\Psi\rangle = S_{ij} E |\Psi\rangle , \tag{22} $$

where one has defined the *Fock operator* $\hat{f}$ with the Coulomb operator $\hat{J}$ and the exchange operator $\hat{K}$

$$ \hat{f} = \sum_i \hat{h}_i + \sum_k^{N/2} \hat{J}_k - \hat{K}_k . \tag{23} $$

As the $\hat{f}$ depends on the chosen set of orbitals as it includes the mean field action of all electrons towards a single electron, the usual way of solving this generalized problem is in a *self consistent field* (SCF) approach. By dividing the problem into two subproblems - finding the optimal Slater determinant for a given $\hat{f}_i$ and then calculating the new $\hat{f}_{i+1}$ for which a new optimal Slater determinant can be determined until the procedure has converged. One recovers the dominant Slater determinant $|\Psi_{\mathrm{HF}}\rangle$ and a set of molecular orbitals $\{\Psi_j\}$ with associated energies ϵ_j.

As this only gives a first approximation without any correlation energy, one can then further refine the energy and trial wave function by expanding around this dominant determinant. Configuration interaction expands the wave function around the Hartree-Fock state as a sum-expansion [5, 6]

$$ |\Psi_{\mathrm{FCI}}\rangle \equiv \left(I + \sum_{pq} C_{pq} a^{\dagger}_p a_q + \sum_{pqrs} C_{pqrs} a^{\dagger}_p a^{\dagger}_q a_r a_s + \dots \right) |\Psi_{\mathrm{HF}}\rangle . \tag{24} $$

This is usually referenced in its entirety as the **Full configuration interaction (FCI)** as the analytical, although computationally intractable, solution of the wave function. As the truncations of this expansion converge slowly in energy due to the linearity of the expansion and do not obey *size extensivity* (the separability of wave function in the non-interacting setting of two different parts of Hilbert space $|AB\rangle = |A\rangle |B\rangle$ for $H_{AB} = H_A \bigoplus H_B$[14]) they are seldomly used in computational chemistry [14].

Coupled cluster (CC) expansion is a product expansion around the Hartree-Fock state, consisting of cluster operators $\hat{T}_n$ of order n, the first two orders are explicitly defined here by $\hat{T}_1 \equiv \sum_{pq} t_{pq} a_p^\dagger a_q$ for $p \in I_{\text{virt}}, q \in I_{\text{occ}}$ and $\hat{T}_2 \equiv \sum_{pqrs} t_{pqrs} a_p^\dagger a_q^\dagger a_r a_s$ for $p, q \in I_{\text{virt}}$, $r, s \in I_{\text{occ}}$ for the index sets I_{virt}, containing all indices of virtual orbitals, I_{occ} containing all indices of occupied orbitals [6, 5]:

$$|\Psi_{\text{CC}}\rangle \equiv \exp\left(\hat{T}\right)|\Psi_{\text{HF}}\rangle = \exp\left(\hat{T}_1 + \hat{T}_2 + \ldots + \hat{T}_n\right)|\Psi_{\text{HF}}\rangle\,. \tag{25}$$

When including all $\hat{T}_n$, this representation becomes exact and recovers the FCI wave function described in . As this is computationally intractable, one terminates the cluster expansion at a fixed n, and labels the relevant expansion with the included level of T_n, e.g. when only including $\hat{T}_1$ and $\hat{T}_2$ one refers to the method as CCSD for *coupled cluster singles doubles* and with perturbatively included triple excitations one uses the term CCSD(T) [15], which is also often referred to as the gold standard of quantum chemistry as it recovers wave functions with low approximation error. Unfortunately, this comes at a prohibitive cost for large scale simulations of $\mathcal{O}(N^7)$ [14].

1.3 Atomic orbitals and Gaussian approximations

As previously mentioned, one starts from a so called basis set which is a list of orbitals associated with each element of the periodic table, and a One possible basis set are Slater type orbitals (STO) introduced by John Slater [1] which model the basis functions after solutions to the Schrödinger equation of the hydrogen atom one centered around the coordinates of the nucleus [5]

$$\phi_{nlm}^{STO}(r,\theta,\phi) = R_{n,l}(r)Y_{l,m}(\theta,\phi) = |\mathbf{r}-\mathbf{R}|^{n-1}e^{-\zeta|\mathbf{r}-\mathbf{R}|}Y_\ell^m(\theta,\phi), \tag{26}$$

where $n \in \mathbb{N}$ is the principal quantum number, N is a normalisation constant, $|r - R|$ is the relative distance to the position of the atomic nucleus R the orbital is centered around, ζ is related to the effective nuclear charge and the spherical harmonics $Y_\ell^m(|\mathbf{r}-\mathbf{R}|)$ take care of the radial part of the orbital. As previously mentioned the molecular orbitals χ are formed from linear combinations of these atomic orbitals ϕ so to evaluate the one- and two-body electron integrals $h_{i,j}$ and $V_{l,k}^{i,j}$ one has to integrate over these types of orbitals. Unfortunately the integrals over the product of two Slater-type orbitals of this form is not computationally efficient, so one has to find other means of evaluating these integrals. The way this is done today is by approximating the STOs with a linear combinations of Gaussian functions ϕ_n^{GF} [5]

$$\phi^{\text{STO}} \approx \phi^{\text{STO-NG}} = \sum_n^N c_n \phi_n^{GF}, \tag{27}$$

where the individual Gaussians follow this functional form.

$$\phi_n^{GF} = \left(\frac{2\alpha_n}{\pi}\right)^{3/4} e^{-\alpha_n|r-R|^2}. \tag{28}$$

These approximations are found by maximizing the overlap between the two functions for a given N and comprise the readily available basis sets online. As one might recall,

$$e^{-\alpha_n(\mathbf{r}-\mathbf{R}_n)^2}e^{-\alpha_m(\mathbf{r}-\mathbf{R}_m)^2} = e^{-(\alpha_n+\alpha_m)\left(\mathbf{r}-\mathbf{R}'\right)^2}e^{-\alpha'(\mathbf{R}_n-\mathbf{R}_m)^2}, \tag{29}$$

where one has introduced

$$R' = \frac{\alpha_n\mathbf{R_n} + \alpha_m\mathbf{R_n}}{\alpha_n+\alpha_m} \quad \text{and} \quad \alpha' = \frac{\alpha_n\alpha_m}{\alpha_n+\alpha_m}. \tag{30}$$

So the product of two Gaussian functions is a another Gaussian centered around R'. This way, even the two-electron integrals over four basis functions can be written as integrals over at most as two center Gaussian function.

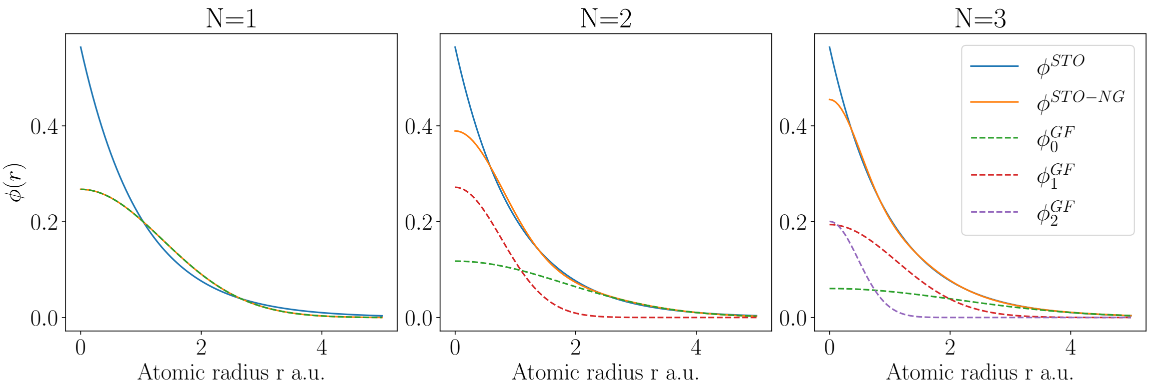


Figure 1: Approximations of the Slater type orbital in STO-NG with N=1,2,3 and the individual Gaussian functions comprising the approximation, plots generated with code starting from [16].

These STO-NG basis sets are very preliminary and are only used in first calculations and approximations of the molecular orbitals of molecules as they do not provide the level of detail to achieve chemically accurate results in post-Hartree-Fock scenarios. Here more elaborate basis sets are used, also in Gaussian form for the aforementioned reasons of classical tractability of electron integrals. These basis sets like 6-31G [17] or cc-pVDZ [18] which are split-valence band basis sets. These split valence band orbitals into multiple functions with different decay coefficients to allow more precise interaction at different length scales, e.g. in bond dissociation scenarios.

2 Quantum computation

Quantum computing has matured from a theoretical endeavor, first proclaimed by Richard Feynmann, then pioneered by Deutsch and Josza first showing that quantum computers could achieve calculations in less steps than classical computers could [19]. Later, Peter Shor showed exponential speedups in factoring [20] and Grover polynomial speedups in unsorted searches [21], increasing interest in potential applications and algorithms outside the Turing model. *Hamiltonian simulation* is one of the main applications of quantum computing, by efficiently implementing time evolution e^{iHt} and being able to efficiently estimate eigenvectors and eigenvalues from this using *quantum phase estimation* (QPE) [22], first algorithms for quantum chemistry [23, 24, 25, 26] were discovered. Further reducing quantum resource requirements [27, 28] and the introduction of new techniques like double factorization [29] and qubitization [30] have reduced costs significantly. Although providing potential exponential speedup over classical alternatives, all these algorithms require a fault-tolerant quantum computer to be realized in practice. Hardware progress has been made on ion-traps [31] and superconducting qubits [32], even showing first scalable results implementing the surface code [33], one possible error-correcting code. Still, these error corrected machines are a decade or more away of reaching the sizes needed to outperform classical algorithms.

While on the path to error-correction, one has to find uses for the current generations of quantum devices. This stage of the development of quantum computing hardware has been coined the NISQ (noisy intermediate scale quantum computation) era [34]. Instead of running entire algorithms end to end, a classical (in most cases optimization loop) accesses the quantum computer to move some of the computation to classical computers, easing the load on the quantum hardware. These **variational quantum algorithms** come in two main flavors: the quantum approximate optimization algorithm (QAOA) [35] and the variational quantum eigensolver (VQE) [36]. The former is an algorithm for combinatorial optimization which up until classical algorithms improved [37] showed better performance on certain problems and the latter is as the name implies an algorithm which finds the (lowest) eigenvalue of a problem Hamiltonian. The potential applications of VQE in chemistry are finding ground and excited state [38] energies. Both algorithms rely on heuristics and do not give any performance guarantees. Outside of the variational algorithms, there are other methods to be highlighted. Quantum Krylov methods [39, 40, 41] and quantum-classical hybrid quantum Monte-Carlo (QC-QMC) [42] as noteworthy examples. In QC-QMC a quantum computer unbiases the calculations of a auxiliary field Monte-Carlo method, not needing an optimisation loop potentially mitigating potential problems that arise from trying to optimize the noisy quantum wave function to extract exact energies. One of the contributions of this work is to propose novel VQE ansätze that have a property that is essential for applying VQE to chemistry and which can also be implemented by repeating simple, physically motivated, gate elements, which reduces the effort needed to calibrate NISQ hardware (Section 5) and benchmark error mitigation techniques (Section 3 and 4) needed to achieve accurate results from noisy devices in simulation and on hardware. The next section introduce the framework of the variational quantum eigensolver and its parts that each pose a challenge on their own.

2.1 Variational quantum eigensolver

As fully error-corrected quantum computers are still in the future (but recent progress has been made showing suppression in error by increasing code distances [33]), algorithms like quantum phase estimation and the quantum Fourier transforms are not possible on current hardware yet. Starting off from a demonstration on a quantum photonic processor [36], this framework as become one of the potential algorithms with useful applications compatible with the NISQ era.

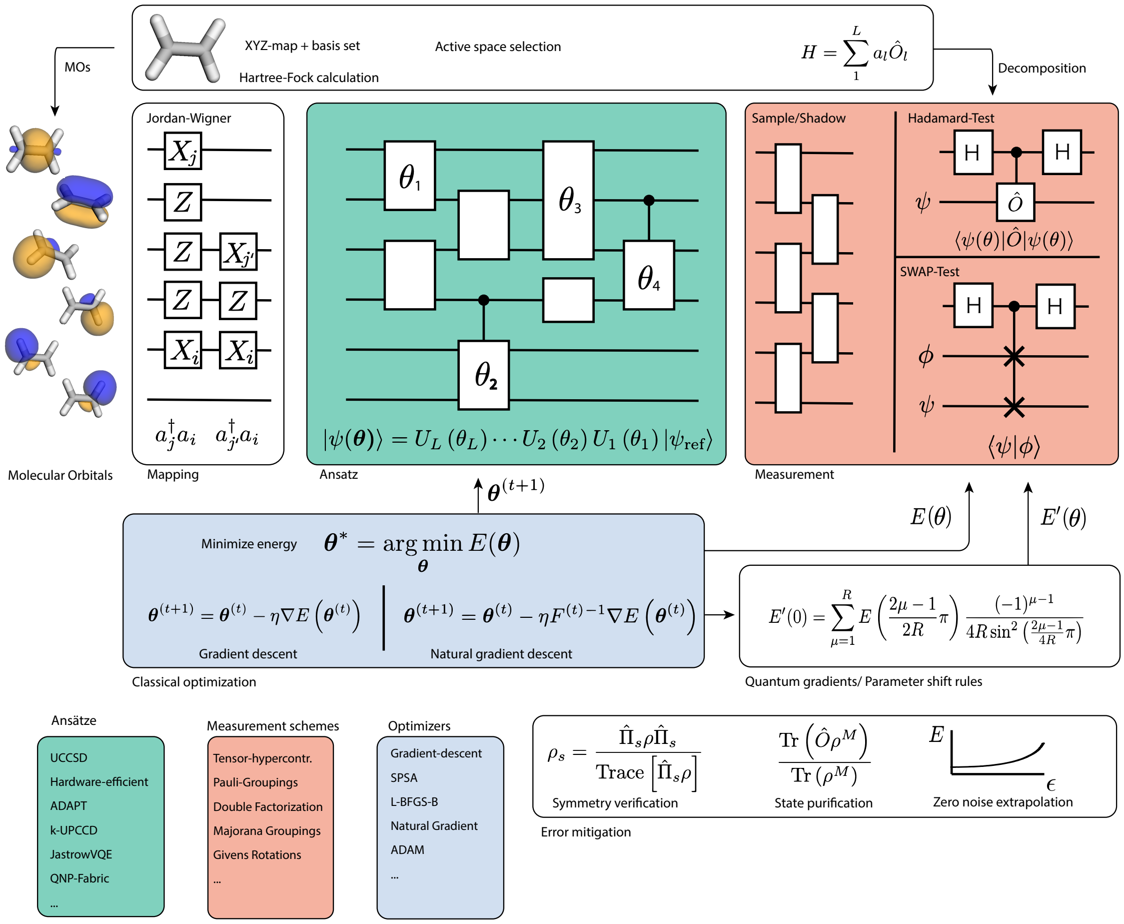


Figure 2: A representation of the VQE framework: On a chosen geometry of a molecule and a given basis-set determining the accuracy of the calculation, a restricted Hartree-Fock calculation calculates a set of molecular orbitals and their respective energies. An active space is chosen (a nontrivial matter, but of no concern here) around the Fermi-energy of the molecule to calculate the correlation energy in the given active space, leaving the orbitals outside the active space in the mean-field treatment of Hartree-Fock. An ansatz is chosen (green), which determines the parametrization given by the quantum circuit. Starting from a random or determined (mostly $|\psi_{HF}\rangle$) state, estimates of the expectation value of the Hamiltonian/ energy are taken by some measurement scheme (red) which tries to optimally estimate the $\mathcal{O}(N^4)$ Pauli terms given by the mapping of the fermionic operators to qubit operators. A classical optimization loop running outside the quantum computer is then querying the quantum circuit at individual parameter configurations and can also measure more complicate properties of the state by the means of the Hadamard or SWAP-test (e.g. in the natural gradient descent) to then suggest a new set of parameters with a lower energy than before. Guided by the variational principle, at the end of the optimization one recovers a trial state that resembles the ground state of the Hamiltonian and possesses the ground state energy. Error mitigation can be used to mitigate errors in the procedure. Colored boxes mark popular choices of each part of the VQE framework.

Instead of running an entire quantum algorithm on the quantum hardware, the variational quantum eigensolver uses the quantum computer to prepare a trial wave function $|\psi(\boldsymbol{\theta})\rangle$, determined by a choice of ansatz (see Section 2.3) and its parametrization $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_L)$. The Hamiltonian H, in the case of chemistry usually taken to be the electronic structure Hamiltonian in second quantization, stemming from a Hartree-Fock calculation on a given geometry of the given molecule, is then mapped onto the qubit system by some fermionic to qubit mapping (Section 2.2) does not necessarily need to be implemented as a imaginary time evolution. Instead H can be decomposed into groups of jointly measurable terms $\hat{O}_m$, often associated by a short rotation of $\mathcal{O}(1) - \mathcal{O}(N)$ into the correct measurement basis, to form a measurable cost function of the system (see Section 2.4). A classical optimizer of choice (Section 2.5) is then given access to sample from this trial wave function to determine energy, possibly gradients and further readings at different choices of $\boldsymbol{\theta}$ to find the lowest configuration of the system. Protected by the variational principle and the assumption that the ground state or at least a sufficiently close state is representable by the ansatz this yields the lowest eigenvalue of the Hamiltonian of the cost function, which corresponds to the ground state of the system.

$$E(\boldsymbol{\theta}) = \frac{\langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle}{\langle \psi(\boldsymbol{\theta}) | \psi(\boldsymbol{\theta}) \rangle} = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle}, \quad \boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) \tag{31}$$

The following section will discuss the main ingredients constituting the algorithm, the challenges to overcome in each of them and possible choices present in the literature. For further details the interested reader should consult [43, 44].

2.2 Fermion to Qubit Mappings

The fermionic creation and anhilation operator give us a convenient way of representing fermionic states. The issue now becomes representing them on a quantum computer, which does not itself consist of indistinguishable fermions but of very much distinguishable qubits and their respective algebra. The Pauli algebra governing the qubit system does not obey the correct anti-commutation relations of the original fermionic algebra, but fortunately there is several mappings from one to the other enforcing the aforementioned relations. The most popular mapping is the Jordan-Wigner (JW) mapping [45], because of its intuitive construction of its operators. The Braviy-Kitaev [46] mapping interleaves qubits representing occupation with ones representing parity, leading to in general shorter representation of fermionic operators but with a more cumbersome construction of operations on this mapping.

The idea behind the JW-mapping is the following. To construct lowering or raising operators one can use the Pauli-matrices $\{\hat{\sigma}_p^X, \hat{\sigma}_p^Y, \hat{\sigma}_p^Z\}$ on qubit p in a linear combination as a first draft:

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \frac{\hat{\sigma}_p^X - i\hat{\sigma}_p^Y}{2}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \frac{\hat{\sigma}_p^X{}_p + i\hat{\sigma}_p^Y}{2} \tag{32}$$

One quickly sees that although representing raising or lowering operators, these commute instead of anticommute $\left[\frac{\hat{\sigma}_p^X - i\hat{\sigma}_p^Y}{2}, \frac{\hat{\sigma}_q^X - i\hat{\sigma}_q^Y}{2}\right] = 0$ because of the commutation relation of the Pauli matrices $\left[\hat{\sigma}_p^a, \hat{\sigma}_q^a\right] = 0$. By using the natural anticommutation relation of the Pauli matrices, $\{\hat{\sigma}_p^a, \hat{\sigma}_p^b\} = 2\delta^{ab}$, and especially $\hat{\sigma}_p^{X/Y}\hat{\sigma}_p^Z = -\hat{\sigma}_p^Z\hat{\sigma}_p^{X/Y}$. By adding so called *Wigner Strings*, chains of $\hat{\sigma}_p^X$ to the mapping to care of the anticommutation.

Jordan Wigner Mapping

$$a_p^\dagger \mapsto \bigotimes_{i=0}^{p-1} \hat{\sigma}_i^Z \otimes \frac{\hat{\sigma}_p^X - i\hat{\sigma}_p^Y}{2}, \quad a_p \mapsto \bigotimes_{i=0}^{p-1} \hat{\sigma}_i^Z \otimes \frac{\hat{\sigma}_p^X + i\hat{\sigma}_p^Y}{2}, \tag{33}$$

To then transform the Hamiltonian in second quantization one replaces the fermionic operators with their JW counterpart, leading to $\mathcal{O}(N^4)$ Pauli terms. One can now write the different parts of the 1- and 2-RDM needed to calculate any operator up to order 2 like the energy of the system explicitly in qubit observables:

$$
\begin{aligned}
{}^1D_p^p &= \frac{1}{2}\left(1 - \hat{\sigma}_p^Z\right) \\
{}^1D_p^q + h.c. &= \frac{1}{2}\left(\bigotimes_{j=q+1}^{p-1} \hat{\sigma}_j^Z\right) \otimes \left(\hat{\sigma}_p^X \hat{\sigma}_q^X + \hat{\sigma}_p^Y \hat{\sigma}_q^Y\right) \\
{}^2D_{pq}^{pq} &= \frac{1}{4}\left(1 - \hat{\sigma}_p^Z - \hat{\sigma}_q^Z + \hat{\sigma}_p^Z \hat{\sigma}_q^Z\right) \\
{}^2D_{pq}^{qr} + h.c. &= \frac{1}{2}\left(\bigotimes_{j=r+1}^{p-1} \hat{\sigma}_j^Z\right) \otimes \left(\hat{\sigma}_p^X \hat{\sigma}_r^X + \hat{\sigma}_p^Y \hat{\sigma}_r^Y\right) \otimes \left(\frac{1 - \hat{\sigma}_q^Z}{2}\right) \\
{}^2D_{pq}^{rs} + h.c. &= \frac{1}{8} \bigotimes_{j=s+1}^{r-1} \hat{\sigma}_j^Z \bigotimes_{k=q+1}^{p-1} \hat{\sigma}_k^Z \otimes (\hat{\sigma}_p^X \hat{\sigma}_q^X \hat{\sigma}_r^X \hat{\sigma}_s^X - \hat{\sigma}_p^X \hat{\sigma}_q^X \hat{\sigma}_r^Y \hat{\sigma}_s^Y + \hat{\sigma}_p^X \hat{\sigma}_q^Y \hat{\sigma}_r^X \hat{\sigma}_s^Y \\
&\quad + \hat{\sigma}_p^Y \hat{\sigma}_q^X \hat{\sigma}_r^X \hat{\sigma}_s^Y + \hat{\sigma}_p^Y \hat{\sigma}_q^X \hat{\sigma}_r^Y \hat{\sigma}_s^X - \hat{\sigma}_p^Y \hat{\sigma}_q^Y \hat{\sigma}_r^X \hat{\sigma}_s^X + \hat{\sigma}_p^X \hat{\sigma}_q^Y \hat{\sigma}_r^Y \hat{\sigma}_s^X + \hat{\sigma}_p^Y \hat{\sigma}_q^Y \hat{\sigma}_r^Y \hat{\sigma}_s^Y)
\end{aligned}
\tag{34}
$$

As the k-RDMs of single Slater determinants are functions of the 1-RDM (see Eq. 19), the energy under the Hartree-Fock approximation can be expressed solely as a function of the 1-RDM, as used in [47]

$$
E_{\mathrm{HF}} = \sum_{pq} h_{pq} {}^1D_p^q + \sum_{pqrs} V_{pqrs} \begin{vmatrix} {}^1D_p^r & {}^1D_p^s \\ {}^1D_q^r & {}^1D_q^s \end{vmatrix}. \tag{35}
$$

In the doubly-occupied/seniority zero (DOCI) approximation (used in Section 4), electrons are only considered in pairs and the excitations follow hardcore bosonic exchange statistics from Eq. 8. This means one does not have to carry Wigner-strings in the measurement and in the trotterized ansatz, easing strain on the quantum computational hardware. The full second quantized Hamiltonian from Eq. 10 can then be written explicitly in parts of the RDMs as:

$$
\begin{aligned}
E_{\mathrm{DOCI}} &= C + \sum_p \frac{h_{pp} + V_{pppp}}{2} {}^1D_p^p + \sum_{p \neq q} \frac{V_{ppqq}}{4} \left({}^1D_p^q + h.c.\right) \\
&\quad + \sum_{p \neq q} \frac{2V_{pqqp} - V_{pqpq}}{4} {}^2D_{pq}^{pq} \\
E_{\mathrm{full}} &= C + \sum_{pq} h_{pq} {}^1D_p^q + \sum_{pqrs} V_{pqrs} {}^2D_{pq}^{rs}
\end{aligned}
\tag{36}
$$

For more efficient measurements of the full Hamiltonian, refer to Section 2.4.

2.3 Ansätze

One of the central elements of the VQE framework is the choice of ansatz. As the reason of this entire framework is limited quantum hardware quality, one has to be very mindful of how to spend these resources. Optimization success and actual feasibility of the algorithm depend highly on the choice of ansatz. The main trade-off these ansätze take into consideration is between cost of representation, as in how many gates each operation consists of and in expressibility of the rotations the parameters are assigned to, where expressibility is measure of which part of state space the ansatz can explore [48]. Starting from an easily prepared reference state, usually the Hartree-Fock state $|\psi_{\mathrm{ref}}\rangle = |\psi_{\mathrm{HF}}\rangle = |0\dots01\dots1\rangle$ prepared by single layer of Pauli-rotations, one parametrizes unitary rotations to build up a more complicated wave function

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|\psi_{\mathrm{ref}}\rangle = \prod_{l} U_l\,(\theta_l)\,|\psi_{\mathrm{ref}}\rangle\,. \tag{37}$$

We will give a brief overview over some of the possible choices of ansätze and their motivations to put the ansatz proposed in publication in Section 5 into context. A more detailed comparison with these ansätze to the QNP-fabric is part of the publication.

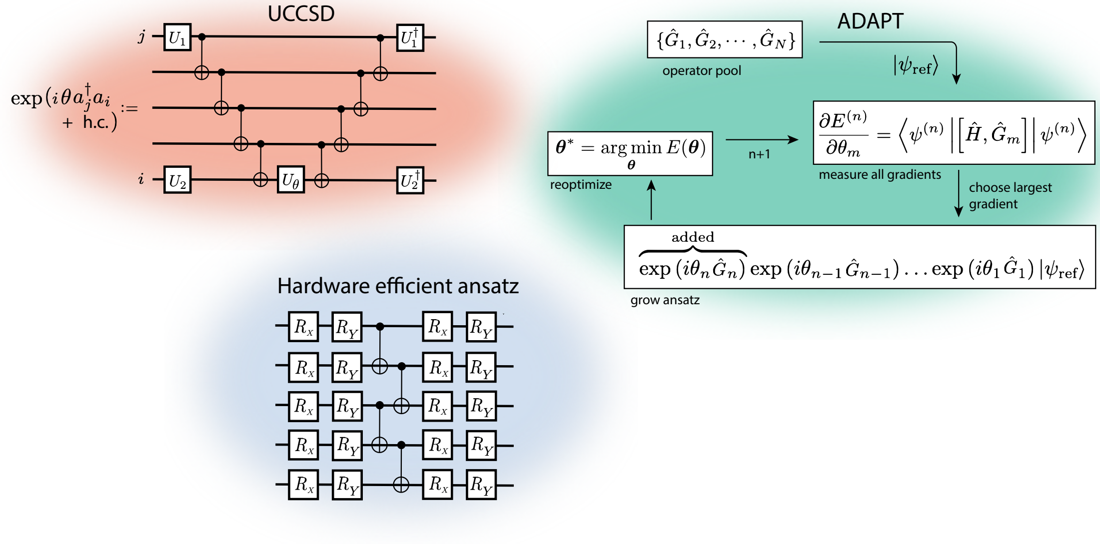


Figure 3: Three popular choices for the ansatz visualized. The UCCSD ansatz consisting of the direct mapping of trotterized fermionically generated rotations through the JW mapping leading to long CNOT chains spanning the qubit space taking care of the Wigner strings (red), the hardware efficient ansatz tiling out the operations natively available to the respective hardware to depths limited by the coherence times or error rates of the device [49] (blue) and the ADAPT ansatz/framework successively choosing the locally best operator with the highest gradient from a predetermined pool of operators, reoptimizing after every step until convergence is reached [50] (green).

The **UCCSD** ansatz from the original VQE paper [36] proposes an ansatz inspired by the classical chemistry algorithm CCSD. It replaces the first and second order cluster operators T_1 and T_2 in Eq. 25 with anti-hermitition operators $\exp\left(T_1^u + T_2^u\right)|\psi_{\mathrm{ref}}\rangle$ where $T_1^u = \prod_{pq} \theta_{pq}(a_p^\dagger a_q - a_q^\dagger a_p)$ and $T_2^u = \prod_{pqrs} \theta_{pqrs}(a_p^\dagger a_q^\dagger a_r a_s - a_r^\dagger a_s^\dagger a_p a_q)$. In principle these rotations can be decomposed into one- and two- qubit rotations as any unitary can be, although these decompositions would be too long for NISQ devices. One instead *trotterizes* $\exp(T_1^u + T_2^u)$ rotations generated under this Hamiltonian using the Trotter Suzuki-Formula

$$e^{Ht} = e^{\sum_{j=1}^{m} O_j t} = \left(\prod_{j=1}^{m} e^{O_j t/r}\right)^{r} + O\left(m^2 t^2/r\right)\,. \tag{38}$$

This becomes exact for $\lim_{r\to\infty}$. UCCSD usually takes approximations of the first order where one creates an ansatz consisting of unitary operations generated by the individual fermionic creation and annihilation operators through the JW mapping $U(\theta) = \exp\left(i\theta a_p^\dagger a_q + h.c.\right)$ and similar for the two body excitations, a circuit representation after JW is shown in Fig. 3. In other approximations e.g. implementing imaginary time evolution on a quantum computer in first order trotterization the trotter error $O\left(m^2 t^2\right)$ destroys the dynamics one tries to capture and significantly higher orders/ smaller trotter steps are needed to minimize the error [51, 52],. In the VQE setting, having a variational ansatz and optimizing the parameters associated with the trotterized rotations can compensate for the Trotter error to some degree [52]. In principle this gives a chemically well motivated ansatz parametrizing sensible rotations in fermionic space. The long CNOT ladders needed to represent the Wigner strings lead to an ansatz that without further optimization in ordering of operators to reduce CNOT count leads to an ansatz that is currently not representable on current hardware.

On the other end of the expressibility-feasability tradeoff there is the **hardware-efficient ansatz** [49] (HEA). Here one tiles out the natively available operations on the used hardware (on superconducting hardware e.g. $\{R_x, R_y, R_z, \text{CNOT}\}$), leading to a very densely parametrized ansatz. This can be done on any hardware, and as the natively available gateset of all hardwares is by design universal, one can in principle build up any wave function possible if one increases depth. The problem with the ansatz is that it also parametrizes unphysical parts of the trial wave function, not contributing to any meaningful complexity and only increasing searchable space of the optimizer. Another problem is the phenomenon of *barren plateaus* that in highly expressible ansätze the cost function in vast regions of the space has vanishing gradient [53]. This makes converging this ansatz in general situations difficult.

The **ADAPT** [50] ansatz/framework is an adaptive way of growing the ansatz operator by operator to achieve minimal scaling in the number of parameters constituting the trial wave function. One begins by defining a pool of operators $\{\hat{G}_1, \hat{G}_2, \cdots, \hat{G}_N\}$ that generate rotations $\exp\left(i\theta_m \hat{G}_m\right)$. The first draft used the cluster operators from UCCSD [50], other schemes like Qubit-ADAPT [54] use a qubit-friendly set of universal operators omitting Wigner-strings. By starting from a reference state, one measures the gradient of the energy with respect to all operators at the current trial state and chooses the operator maximizing the gradient

$$\frac{\partial E^{(n)}}{\partial \theta_m} = \left\langle \psi^{(n)} \middle| \left[\hat{H}, \hat{G}_m\right] \middle| \psi^{(n)} \right\rangle . \tag{39}$$

The ansatz is grown by the chosen operator to form a new trial wave function

$$|\psi_{\text{n}}\rangle = \exp\left(i\theta_n \hat{G}_n\right) |\psi_{\text{n-1}}\rangle . \tag{40}$$

Then a global optimisation of all parameters is performed starting from the old set of parameters. This loop is then performed up until convergence with some convergence criteria, a common convergence criteria is all gradients from Eq. 39 fall below a threshold $\partial E^{(n)}/\partial \theta_m < \epsilon \, \forall m$. A representation of the algorithm is shown in Fig. 3. Numerical studies show that this way of building an adaptive ansatz is insensitive to barren plateaus and vanishing gradients [52].

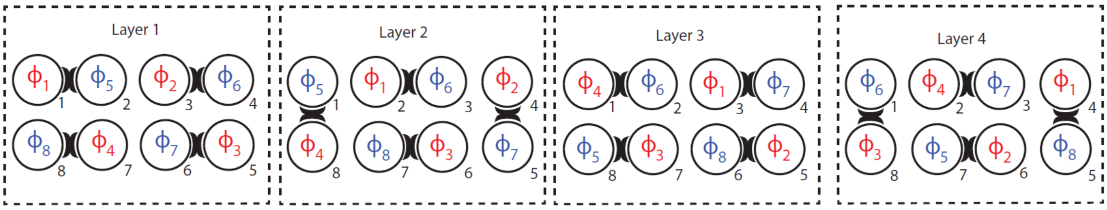


Figure 4: A SWAP-network on a ring as used as an ansatz in Section 4. After $N/2$ steps, each blue mode has been coupled to each red mode but no intercolour coupling as taken place. In the chemistry scenarios, this couples each occupied to each virtual efficiently.

As the full second quantized Hamiltonian is still prohibitive in scaling to current NISQ platforms, there exist more tailored ansätze only parametrizing the relevant parts of Hilbert space associated with a restricted part of the Hamiltonian. When doing Hartree-Fock on a quantum computer [47], one only implements the the number preserving orbital rotation U which transforms the underlying basis set and corresponds to a basis change to a new set of orbitals. This can be implemented in $\mathcal{O}(N)$ in a Givens-Network [32]. In the DOCI approximation of chemistry [10], discussed in 2.2, one can take a trotterized step in the paired approximation in depth N using a diamond shaped Givens-SWAP network. In this work in Section 4 a more compact version of this ansatz is introduced by the author, only taking $N/2$ layers to implement the fully first order trotterized step between virtual and occupied orbitals by taking advantage of the missing Wigner Strings associated with the operators, resulting in a maximally dense arrangement. This is achieved putting the operators on a ring instead of a line, which creates two modes traveling in opposite direction. The action of the SWAP-network on a ring is shown in Figure 4.

Furthermore, the author has introduced the **quantum-number preserving ansatz** (QNP), a more hardware accessible, chemistry inspired ansatz that only parametrizes the relevant subspace with the correct quantum numbers. One way to view the ansatz is the combination of orbital rotations from Hartree-Fock parametrizations of [47] and pairwise excitations from [10] in a brick-layer circuit. For further discussion please refer to Section 5.

2.4 Measurement

As previously discussed, variational quantum eigensolver requires many circuit repetitions (*shots*) which one has traded for shorter circuit length in comparison to quantum phase estimation. One of the challenges is to managing this shot count overhead. Mapping the second quantized Hamiltonian onto qubit operators leaves $\mathcal{O}(N^4)$ to be estimated per energy evaluation of the state. The total number of distinct measurements as well as the individual variances of each term constitute the total cost of estimating the energy to finite precision. What most these techniques share is that they *decompose* the Hamiltonian into a weighted sum of *fast forwardable* observables $\hat{O}_i$

$$H = \sum_i b_i \hat{O}_i. \tag{41}$$

Fast-forwardable in this case means that the spectrum of the observable $\hat{O}$ is known/can be classically calculated by means of diagonalization, so time-evolution under these observables can be *fast forwarded* classically by evolving under the known spectrum. One way to reduce the number of measurements needed is to group the Pauli words into cliques of jointly measurable groups of Paulis [55], where the distribution into the groups is a nontrivial task When allowing for deeper circuits of $\mathcal{O}(N)$ rotating into the measurement basis, one can reduce the cost of measurement further:

One possibility is by means of different factorization techniques, e.g. tensor hypercontraction [56, 57] (mostly used in Qubitization approaches), single factorizations [58] and **double factorization** [29] (named after a second factorization step over the single factorized approach) techniques. In double factorization approaches, the second quantized electronic structure Hamiltonian from Eq. 10 in slightly modified form where some of the two body contribution has been included in the single body term $h^{\mathrm{mod}}_{pq} = h_{pq} + \sum_i [2(pq|ii) - (pi|qi)] + 1/2 \sum_r (pr|qr)$ with $i \in I_{core}$, where I_{core} is the set of indices of frozen core orbitals, not participating in correlation and being treated at an HF level of theory. The modified single electron $(p|h^{\mathrm{mod}}_{pq}|q)$ and the two electron tensor $(pq|rs)$ are then eigendecomposed and the individual *leaves* resulting from the first factorization of two-body term are then eigendecomposed again, resulting in [59].

$$E \equiv \mathcal{E} + \sum_k \mathcal{F}_k^{\varnothing} \omega_k^{\varnothing} + \sum_{t=1}^{n_t} \sum_{kl} Z_{kl}^t \omega_{kl}^t. \tag{42}$$

where n_t is the number of leaves from first factorization of the two-body term, the modified one body (two body) integrals $\mathrm{F}_k^{\varnothing}(\mathrm{Z}_{kl}^t)$ being a result from the (nested) factorized one body (two body) electron integrals and the doubly factorized reduced density matrices are $\omega_k^{\varnothing}$ and ω_{kl}^t defined by [59]

$$\begin{aligned} \omega_k^{\varnothing} \equiv& \frac{1}{2} \left\langle \Psi(\varnothing) \middle| \hat{Z}_k + \hat{Z}_{\bar{k}} \middle| \Psi(\varnothing) \right\rangle, \\ \omega_{kl}^t \equiv& \frac{1}{8} \langle \Psi(t) | \hat{Z}_k \hat{Z}_l - \hat{I}\delta_{kl} + \hat{Z}_k \hat{Z}_{\bar{l}} \\ &+ \hat{Z}_{\bar{k}} \hat{Z}_l + \hat{Z}_{\bar{k}} \hat{Z}_{\bar{l}} - \hat{I}\delta_{\bar{k}\bar{l}} | \Psi(t) \rangle. \end{aligned} \tag{43}$$

where k and $\bar{k}$ are indexing orbitals of different spin and the modified states (or better, states in the rotated measurement basis) $|\Psi(\varnothing)\rangle \equiv U^{\varnothing}|\psi\rangle$ and $|\Psi(t)\rangle = U^t|\psi\rangle$ can be measured by implementing a rotation by unitaries $U^{\varnothing}$ and U^t determined by the factorizations and efficiently implemented by a Givens-network (orbital rotation) of depth N. As one only has to measure $\hat{Z}_p$ and $\hat{Z}_p\hat{Z}_q$ observables in each of these basis, which all commute, one only has to measure in $n_t + 1$ basis to estimate the energy of the given Hamiltonian. Truncating the full number of leaves $n_t = N(N+1)/2$ in an appropriate ordering makes the solution not exact anymore but reduces the measurement overhead further. Even more elaborate means of compressing the factorization, compressed double facorization (C-DF) have been developed [60, 61].

By randomizing over different measurement basis **classical shadows** can estimate a linear $\mathcal{O}(N)$ amount of observables with only $\mathcal{O}(\log(N))$ shots [62]. To create a classical shadow representation of a unknown quantum state ρ measure the quantum state by applying a random unitary U which transforms $\rho \rightarrow U\rho U^\dagger$ and storing the resulting bitstring $|b\rangle$ The quantum channel $\mathcal{M}$ that constitutes the average over choise of random unitary and measurement outcome constitutes the following equation.

$$\mathbb{E}\left[U^\dagger|\hat{b}\rangle\langle\hat{b}|U\right] = \mathcal{M}(\rho) \Longrightarrow \rho_{\text{shadow}} = \mathbb{E}\left[\mathcal{M}^{-1}\left(U^\dagger|\hat{b}\rangle\langle\hat{b}|U\right)\right] \tag{44}$$

The inverse of the measurement channel is not necessarily physical but one can still apply it [62] and find that the then estimated shadow density matrix mirrors the original density matrix in expectation value $\mathbb{E}[\hat{\rho}] = \rho$ in average so by building up an array of these shadows one can then predict certain properties from the purely classical estimates.

First introduced randomizing over local or global Clifford measurements [62], the local Clifford variant had an unfavorable exponential scaling in the size of the support of the observable one tries to estimate from the classical shadow. There is a geometrical argument that supports this finding: as Clifford measurement constitutes a discrete set and one is trying to estimate a Pauli-word $P = \sum_s \sigma_s^{x/y/z(s)}$ with support s and all Pauliwords form a orthogonal basis, only measurements that *contain* the given Pauli-word contribute to the estimation. With growing s, its exponentially unlikely that a random measurement will contain the given pauli-word. As qubitized fermionic k-RDMs have large support, these original methods did not achieve competitive scaling for fermionic. Different protocols adapted for fermions randomizing over the unison of Clifford circuits and Gaussian unitaries [63], over the entire Gaussian unitary ensembe [64] and over the particle-number preserving matchgates [12], randomizes over the same measurement basis as the double-factorized measurement methods, have been introduced in the last year achieving scaling for estimating all elements of the k-RDM in $\mathcal{O}\left(\eta^k/\epsilon^2\right)$ where η is the number of fermions present, independent of the number of orbitals/modes. A major hurdle to use these favorable scalings is that current hardware can not support these randomized single-shot measurements as reloading a new cirucit associated with a new random measurement basis takes orders of magnitutes more time than sampling a fixed circuit many times [65].

2.5 Quantum gradients and optimization

As the name suggests, variational quantum algorithms draw some of their appeal from the connection to a classical optimizer, having shorter quantum circuits to evaluate leading to more loose hardware needs and protecting against coherent noise e.g. overrotation. Methods that only require evaluations of the cost function but no gradients or higher order derivatives like a Simplex or COBYLA can be used and are used in practice, as in classical optimization these are more robust against noise, shot and device noise all together. However, with growing size these optimization problems become very challenging and in general have been shown to be NP-hard [66]. To increase performance, gradient based optimization or even higher derivative methods are used [67]. The challenge then becomes evaluating the gradient of the parameters in the circuit, but it turns out there is an analytical way of recovering the gradient of a quantum gate without using finite difference.

The foundation of all shift rules and further analysis of quantum cost functions is that cost functions arising from quantum gates are trigonometric polynomials with order equal to the number of eigenvalues of the generator of the quantum gate. Fig. 5 shows the associated cost functions of two gates at varying parameter settings. By (taken from [67]) starting with a generated gate $U(\theta) = \exp(i\theta\hat{G})$ with the Generator $\hat{G}$ with eigenvalues $\{\exp i\omega_j x\}$, real valued $\{\omega_j\}$ in non-

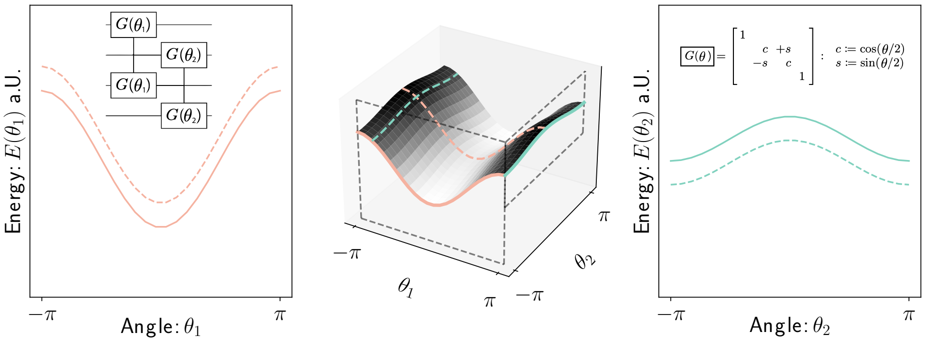


Figure 5: Visualizations of quantum cost functions of two Givens gates (inset on the right shows the matrix form of the individual gates) on a random state. Both individual gates have single sinuses as cost functions as they have two eigenvalues and form a trigonometric polynomial of higher order when viewed together. Cuts along individual parameter values with the other parameter fixed take the form of the original single frequency cost function with different coefficients.

decreasing order. One absorbs gates before and after the gate U into the state $|\psi\rangle$ and into the observable $\hat{O}$ [67]

$$
\begin{aligned}
E(\theta) &:= \left\langle \psi \middle| U^\dagger(\theta)\hat{O}U(\theta) \middle| \psi \right\rangle \\
&= \sum_{j,k=1}^{d} \overline{\psi_j e^{i\omega_j\theta}} b_{jk}\psi_k e^{i\omega_k\theta} \\
&= \underbrace{\sum_{j=1}^{d} |\psi_j|^2 b_{jj}}_{a_0} + \sum_{\substack{j,k=1 \\ j<k}}^{d} \left[\underbrace{\overline{\psi_j} b_{jk}\psi_k}_{c_\ell = c_{\ell(j,k)}} e^{i\overbrace{(\omega_k - \omega_j)}^{\Omega_\ell}\theta} + \text{h.c.} \right] \\
&= a_0 + \sum_{\ell=1}^{R} c_\ell e^{i\Omega_\ell\theta} + \text{h.c.} \\
&= a_0 + \sum_{\ell=1}^{R} a_\ell \cos(\Omega_\ell\theta) + b_\ell \sin(\Omega_\ell\theta),
\end{aligned}
\tag{45}
$$

where [67] introduces a new index ℓ through a mapping of the old indices, the new coefficients trough $c_\ell = \frac{1}{2}(a_\ell - ib_\ell)$ and R unique positive differences $\{\Omega_\ell\}_{\ell\in[R]} := \{\omega_k - \omega_j \mid j,k \in [d], \omega_k > \omega_l\}$.

First discovered for single-qubit gates [68] (which are always generated by matrices of two eigenvalues),

$$
\frac{\partial}{\partial\theta}E(\theta) = \frac{1}{2\sin(\frac{\pi}{2})}\left[E\left(\theta + \frac{\pi}{2}\right) - E\left(\theta - \frac{\pi}{2}\right)\right], \tag{46}
$$

then generalized to arbitrarily generated gates [67] with equidistant spacing of R equidistant eigenvalues

General parameter-shift rule

$$
E'(0) = \sum_{\mu=1}^{2R} E\left(\frac{2\mu-1}{2R}\pi\right) \frac{(-1)^{\mu-1}}{4R\sin^2\left(\frac{2\mu-1}{4R}\pi\right)}. \tag{47}
$$

these parameter-shift rules provide an analytical form (disregarding shot noise) to evaluate gradients of quantum gates. Having access to the gradients of the cost function allows for the use of gradient based optimization methods like Gradient-descent, ADAM [69] and L-BFGS-B [70].

The insight into the trigonometric nature of the cost function from Eq. 45 has far reaching consequences and is at the core of many discoveries in the last years. By adding an additional measurement point to the parameter-shift rules one can recover the the functional form of the cost as a function of one parameter with all other parameters fixed [71, 72, 67]. When recovering the full cost function one can go beyond taking only a short step along the gradient but can classically determine the minimum of the cost function and setting the parameter locally optimal, which is often referred to as the *Rotosolve* optimizer [72]. For small R there exist closed forms of the minima of the function [72, 67], and for functions with higher R where no closed form exist one can still classically optimize on the extracted function without spending quantum hardware resources [67]. Although usually limited to local updates of parameters, prohibiting collective movement in parameter space, for certain applications like Section 3.7 where no collective movement is needed one can converge to an optimum in very few steps, as each parameter is set optimally locally at each step.

3 Quantum error mitigation

The absence of error correction protecting against device noise in the NISQ era makes non-scalable approaches of estimating quantities with reduced dependency on error processes a vital task to make these devices useful for relevant applications in the near and mid term scope. Especially in chemical algorithms that aim to extract quantities with *chemical accuracy* from quantum hardware like ground and excited states as described above this is necessary to be competitive with classical algorithms, as the level of precision one can estimate these quantities to is one of the key factors different algorithms are judged by. The definition of quantum error mitigation is not well agreed on [73], usually one refers to procedures that reduce dependency on error that in principle scale exponentially in dependency of shots or system size but have potential windows of usefulness as the quality of hardware improves in error rates and coherence times to bridge the gap until fully error corrected quantum computers arrive. Another way of viewing the task of error mitigation is building estimators, whose inherent bias is reduced under common noise models [73]. Common techniques to mitigate noise are artificially injecting noise at varying levels into the given circuit to extrapolate the performance under no noise, commonly referred to as zero-noise extrapolation or Richardson extrapolation [74, 75]. By learning the particular noise models of a given device one can insert additional gates to compensate/mitigate noise processes [74, 76]. In the following sections a more detailed explanation of the error mitigation techniques used in Section 4 is given, a more detailed review of methods is found in [73].

3.1 Noise in quantum circuits

Before talking about mitigating noise, a quick introduction into the reason these techniques are necessary - **noise**. The imperfection of quantum hardware is a major challenge of quantum computation in stark contrast to classical hardware, where digitization of analog signals and error corrected memory has made errors inside processors almost nonexistent and error correction codes are mostly used when dealing with communicating information through imperfect channels. As amplitudes of different parts of the prepared wave function still remain as a continuous spectrum, this is not the case for quantum computers. Noise in quantum circuits can be categorized into two main categories:

Coherent noise is noise that does not change the purity $\mathrm{Tr}(\rho)$ of a quantum state. One kind of coherent noise present in current quantum devices are over/under-rotations stemming from faulty calibration of quantum gates or drift of experimental parameters after calibration. They add a small offset γ to the parameter associated with the unitary operation $U(\theta) \mapsto U(\theta \pm \gamma)$. In the original formulation of variational settings this does effect measurement outcomes as the variational principle protects against this offset. The minimum after optimization will be found at a perturbed set of parameters $\boldsymbol{\theta}'$ with the same energy as in the unperturbed setting $\min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}'} E(\boldsymbol{\theta}')$.

On the other hand, **incoherent noise** does scatter the originally pure state into an incoherent mixture of states ρ_{noise}, resulting in a stochastic mixture of states. These noise processes are modelled with a quantum channel Φ, which is a linear, completely positive, and trace-preserving map. A convenient way of representing these channels are *Kraus operators* $\{\hat{K}_i\}$ associated with the given quantum channel with the property $\sum_i \hat{K}_i^\dagger \hat{K} = I$. Their action on a given ρ is then [77]

$$\Phi(\rho) = \sum_i \hat{K}_i \rho \hat{K}_i^\dagger. \tag{48}$$

One can see that after the application of the quantum channel it has applied the Kraus operator K_i with probablity $p_i = \mathrm{Tr}(\hat{K}_i \rho \hat{K}_i^\dagger)$, resulting in a statistical mixture described above. Furthermore noise channels that are described by one Kraus operator, or in other words quantum channels of rank 1, describe pure channels. One can now find Kraus operators for typical noise channels present in the different hardware settings. Typical noise channels present in most devices are the *depolarizing* channel, *amplitude damping* channel and the *phase damping* channel.

The depolarizing channel applies the three Pauli matrices with a given probability p [77]

$$\hat{K}_{\text{dp},0} = \sqrt{1-p}\, I, \quad \hat{K}_{\text{dp},1} = \sqrt{p/3}\, \hat{\sigma}^X, \quad \hat{K}_{\text{dp},2} = \sqrt{p/3}\, \hat{\sigma}^Y, \quad \hat{K}_{\text{dp},3} = \sqrt{p/3}\, \hat{\sigma}^Z. \tag{49}$$

The amplitude damping channel projects onto the zero state with probablity p and is associated with the following Kraus operators

$$\hat{K}_{\mathrm{ad},0} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad \hat{K}_{\mathrm{ad},1} = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}. \tag{50}$$

and the phase damping channel destroys the information stored in the phase of a quantum state without altering excitation levels [77]

$$\hat{K}_{\mathrm{pd},0} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad \hat{K}_{\mathrm{pd},1} = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{pmatrix}. \tag{51}$$

As density matrix simulators have to simulate a $2^n \times 2^n$ density matrix in contrast of keeping only a 2^n sized state vector in memory, they are much more computationally demanding to simulate. This increase in size and the unknown nature of the noise processes for devices in general makes current quantum hardware competitive when simulating actual error mitigation techniques, see Section 4. The noise channels above constitute single qubit noise channels, noise on real devices generally is not limited to single qubit noise due to cross talk.

3.2 Postselection

Symmetries $\hat{S}$ in the Hamiltonian one is investigating can be used in varies way, by 'tapering' off qubits to reduce Hilbert space size in simulations [78, 79] or by projecting the noisy state into the symmetry subspace to help mitigate errors occurring during the computation, which is often just referred to as *Postselection* [80]. This can be viewed as purifying the noisy state into one with less noise by projecting out parts of the state that can only be the result of noise processes, hence the name of the publication as the following methods can also be viewed from this perspective. Postselection is often regarded as a 'zero cost' error mitigation technique [43], as it does not demand an increase in qubit number or circuit depth, only a higher repetition of shots. This is not quite true as the managing shot budgets is a major task in VQE. These few requirements allow it to be combined with more advanced techniques to further increase fidelity, e.g. here one can use Postselection on a subset of the measurements of virtual distillation (see 3.5).

The introduction here follows [80]. Symmetries are observables that commute with your Hamiltonian $\left[H, \hat{S}\right] = 0$ and its decomposition into fast forwardable terms $\left[\hat{O}_i, \hat{S}\right] = 0$. The measurements are defined as the projector-valued measurement $\{\hat{\Pi}_i\}$ (PVM) for the measurement outcomes i with the following properties $\sum_i \hat{\Pi}_i = \mathbb{I}, \hat{\Pi}_i^2 = \hat{\Pi}_i^\dagger = \hat{\Pi}_i$. The outcomes i occur with the probabilities p_i following the Born-Rule

$$p_i = \langle\psi| \hat{\Pi}_i |\psi\rangle, \tag{52}$$

and projects the state onto $p_i = \hat{\Pi}_i |\psi\rangle / \sqrt{p_i}$. The PVM $\{\hat{\Pi}_S\}$ that projects onto the target eigenspace $\mathcal{S}$ can then be constructed. As long as the symmetry projection leaves the pure state in the absence of noise unchanged $\hat{\Pi}_s|\psi\rangle = |\psi\rangle$ the measurement of ρ and the following projection/postselection under $\{\Pi_s\}$ results in the projected density matrix

$$\rho_s := \frac{\hat{\Pi}_s \rho \hat{\Pi}_s}{\mathrm{Tr}\left[\hat{\Pi}_s \rho\right]} \tag{53}$$

and one can see that

$$\mathrm{Tr}\left[\rho_s |\psi\rangle\langle\psi|\right] = \frac{\mathrm{Tr}[\rho|\psi\rangle\langle\psi|]}{\mathrm{Tr}\left[\hat{\Pi}_s \rho\right]} \geq \mathrm{Tr}[\rho|\psi\rangle\langle\psi|], \tag{54}$$

the projected density matrix ρ_s has even or greater overlap with the pure state before noise $|\psi\rangle$ as the noisy density matrix [80]. When simulating fermions, in most cases the number operator $\hat{N} = \sum_j \hat{a}_j^\dagger \hat{a}_j$ commutes with the given observable one tries to measure in the VQE setting and is conserved. Furthermore when one is in a spin-orbital mapping the number of alpha and beta electrons is also convserved, lending itself to the postselection scheme. In practice one looks at the shot register and throws out every shot that does not have the correct excitation number in total or in the separate spin sectors and only calculates expectation values from the subset attained. With the cost C associated with postselection is

$$C = \frac{1}{\mathrm{Tr}\left[\Pi_s \rho\right]}, \tag{55}$$

as this marks the probability of a single shot not passing the criteria.

3.3 Echo verification

Instead of verifying that the state one has prepared obeys the correct symmetries of the Hamiltonian, one rather would like to verify that the state one has prepared is the indeed the one that was intended. This would reach the information theoretical limit of mitigating all errors, although having a success probability that scales with the inverse of the circuit fidelity [81]. The task behind echo verification is estimating $\langle H \rangle$, where one decomposes H into a sum of fast fordwardable terms Eq. 41. A quick introduction into estimating expectation values of arbitrary, fast forwardable observables with the Hadamard test follows, as introduced in [81]. The state that was prepared by a unitary $|\psi\rangle_s = U_{\mathrm{total}}(\boldsymbol{\theta})\,|0\rangle_s$ where $|\rangle_s$ marks the register associated with the system and one has added an ancilla qubit prepared in the plus state $|+\rangle_a$.

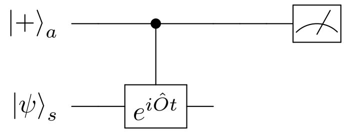


Figure 6: A depiction of the SWAP-test with an ancilla qubit.

which encodes the state $\frac{1}{\sqrt{2}}(|0\rangle_a|\psi\rangle_s + |1\rangle_a e^{i\hat{O}t}|\psi\rangle_s)$. More explicitly, one divides the unitary preparing the state $U_{\mathrm{total}}(\boldsymbol{\theta}) = U_a(\boldsymbol{\theta})S_{\mathrm{prep}}$ where S_{prep} is the *state preparation* of the starting state of the ansatz $S_{\mathrm{prep}}\,|0\rangle = |\psi_{\mathrm{ref}}\rangle$ which does not change in e.g. optimization steps of the VQE and $U_a(\boldsymbol{\theta})$ is one of the ansätze discussed in 2.3 and explicitly parametrized. As S_{prep} starts of from the all-zero state $|0\rangle_s^{\otimes n}$, it does not conserve any quantum numbers while U_a typically in fermionic simulations conserves at least the number operator $\left[U_a, \hat{N}\right] = 0$. One can expand this into the eigenstates of the observable $|E_j\rangle_s$ [81]

$$|\psi\rangle_s = \sum_j m_j\,|E_j\rangle_s \quad \text{and} \quad |\Psi(t)\rangle = \sum_j \frac{1}{\sqrt{2}}\left(|0\rangle_a + e^{iE_j t}|1\rangle_a\right)|E_j\rangle_s\,. \tag{56}$$

Tracing over the system register yields [81]

$$\rho_{\mathrm{a}}(t) = \mathrm{Tr}_{\mathrm{s}}[|\Psi(t)\rangle\langle\Psi(t)|] \tag{57}$$

$$= \frac{1}{2}\begin{pmatrix} 1 & g(t) \\ g^*(t) & 1 \end{pmatrix}, \tag{58}$$

with $g(t) = \sum_j M_j e^{iE_j t}$ and $M_j = |m_j|^2$. The definition of the phase function can be written as $g(t) = \mathrm{Tr}\left[\rho_{\mathrm{a}}|0\rangle_a\langle 1|_a\right]$. To extract the expectation value $\langle H \rangle = \sum_j M_j E_j$ from the phase function there is a multitude of ways, here the derivative of the phase function is used:

$$g'(t) = \sum_j iM_j E_j e^{iE_j t} \quad \longrightarrow \quad \mathrm{Im}(g'(0)) = \sum_j M_j E_j = \langle H \rangle \tag{59}$$

So the task of estimating expectation values in the Hadamard test can be mapped to estimating the gradient of the phase function of the controlled evolution via the ancilla qubit. One can use all the tools discussed in Section 2.5, e.g. parameter-shift rules in finding the times t on when to evaluate the phase function to extract the gradient. The optimal way of estimating expectation values in this setting is part of Section 6 and includes a comparison to the tools that the gradient picture of the phase function provides.

Now one wants to extend the Hadamard test to include verification, projecting onto the subspace of $U^\dagger(\boldsymbol{\theta})\left|0\right\rangle_s$, undoing the computation. The framework is shown in Fig. 7. As this evolves the system, even under no noise the system does not return to the all zero state $\left|0\right\rangle_s$ and fails verification. One would expect this to change the phasefunction $g(t)$ in the estimation, but one can show that this leaves the phase function $g(t)$ unchanged [81]. The physical argument from [81] on why this is the case is the following. Explicitly dividing the density matrix of the ancilla qubit ρ_a under no noise into a part that passes and one that does not pass verification gains [81]

$$\rho_a = \rho_{\mathrm{a}}^{(\mathrm{v})} + \rho_{\mathrm{a}}^{(\mathrm{f})}. \tag{60}$$

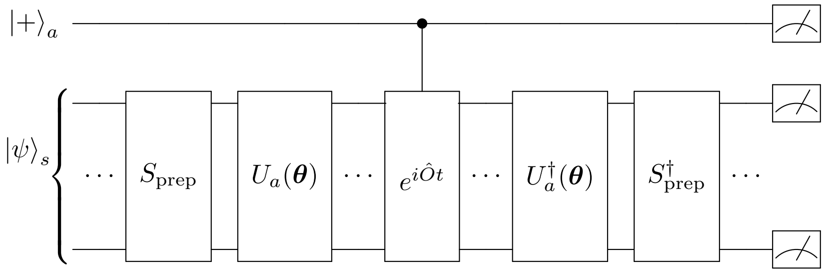


Figure 7: Controlled echo verification. Expectation value gets encoded as a phase on the additional ancilla qubit by implementing controlled time evolution on the system qubit.

When the ancilla qubit is in the $\left|0\right\rangle_a$ state at the end of the computation, no evolution has taken place, the system will always return to $\left|0\right\rangle_s$ and pass verification. This means that not passing verification always projects the ancilla into the $\left|1\right\rangle_s$ state and therefore $\rho_{\mathrm{a}}^{(\mathrm{f})} = |1\rangle_a\langle 1|_a$, contributing 0 on average to the estimation [81].

$$\mathrm{Tr}\left[\rho_{\mathrm{a}}|0\rangle_a\langle 1|_a\right] = \mathrm{Tr}\left[\rho_{\mathrm{a}}^{(\mathrm{v})}|0\rangle_a\langle 1|_a + \underbrace{\rho_{\mathrm{a}}^{(\mathrm{f})}|0\rangle_a\langle 1|_a}_{0}\right] = g(t) \tag{61}$$

Now including noise, one can decompose the noisy estimate of the phase function [81] into

$$g_{\mathrm{noise}}(t) = p_{\mathrm{ne}}g(t) + \mathcal{O}\left(p_{\mathrm{err}}(t)\right), \tag{62}$$

where it is assumed that p_{ne} does not depend on t. Noise then acts as a rescaling factor of the noisy wave function and its amplitudes M_j. The task of estimating the error corrected expectation value of $\hat{O}$ becomes

Echo verification

$$\langle\hat{O}\rangle_{\mathrm{corrected}} = \frac{\mathrm{Im}(g'(0))}{p_{\mathrm{ne}}}. \tag{63}$$

There is a more NISQ friendly way of doing the Hadamard test in this setting which is referred to as **control-free echo verification** which omits the ancilla qubit and the controlled time evolution in specific cases in exchange for a slightly more involved preparation of the reference state, making this error mitigation technique more NISQ approachable.

Choosing an eigenstate of the observable as a *phase reference* state $\hat{O}\left|\psi_O\right\rangle = E_O\left|\psi_O\right\rangle$ which is a simultaneous eigenstate of the unitary ansatz $U_a\left|\psi_O\right\rangle = E_a\left|\psi_O\right\rangle$ one can encode the phase function in the phase difference of the original reference state $\left|\psi_{\mathrm{ref}}\right\rangle_s$ and the phase reference state [81]. As long as one can efficiently compute the eigenvalue associated with the phase reference with the observable (the phase the reference state gathers from the ansatz is reversed when undoing the unitary), one can recover the expectation value of the original state in post-processing or compensate for the phase the phase reference gathers by additionally applying a phase in state preparation. In the case of number conserving fermionic simulations (which is the case for chemistry calculations)

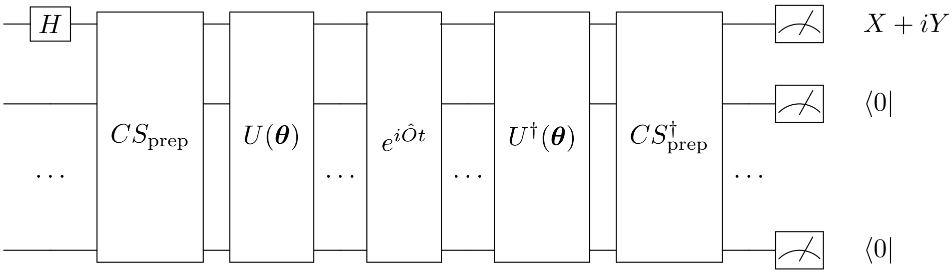


Figure 8: Control-free echo verification. One prepares an equal superposition of a reference state (often $|0\rangle$) and a starting state, the ansatz $U(\boldsymbol{\theta})$ and imaginary time evolution $e^{i\hat{O}t}$ are only allowed to change the reference state by a (classically efficiently determinable) phase.

the fermionic vacuum $|0\dots0\rangle$ is a valid reference state [81]. The procedure is depicted in Fig. 8 on the right and involves the following steps [81].

One starts of by preparing a superposition of $1/\sqrt{2}(|0\dots0\rangle_s + |10\dots0\rangle_s$ with a single Hadamard gate on one of the qubits. Then a slightly involved state preparation CS_{prep} conditioned on the chosen qubit is applied preparing $1/\sqrt{2}(|0\dots0\rangle_s + |\psi_{\text{ref}}\rangle_s$. Applying the original ansatz unitary U_a prepares the state mentioned above. Time evolution under the rotation generated by the observable and then applying $CS^\dagger_{\text{prep}}U_a^\dagger$ leaves the first qubit in the state [81]

$$\frac{1}{\sqrt{2}}(e^{iE_O}|0\rangle_1 + e^{i\hat{O}t}|1\rangle_1), \tag{64}$$

and the rest of the system register can be used for verification.

3.4 Control-free verified SWAP test

At this point this error mitigation scheme is applied to the SWAP test. The relevancy of the SWAP test to these purification methods will become apparent in the latter Section 3.5, here the concern is to formulate the SWAP test under echo verification which allows a) omitting the ancilla qubit and b) verifying at the end of the procedure to mitigate errors in the process.

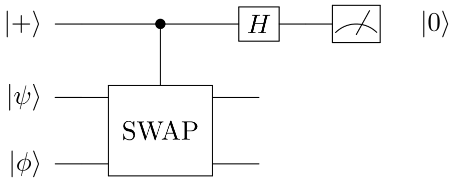


Figure 9: Diagrammatic representation of the SWAP test, which calculates the overlap between the quantum states $|\psi\rangle$ and $|\phi\rangle$ using an ancilla qubit in the $|+\rangle$ state.

The original SWAP test as introduced by [82], depicted in Fig. 9 uses one ancilla qubit to calculate the overlap between two quantum states $|\psi\rangle$ and $|\phi\rangle$ by applying a controlled SWAP gate (CSWAP) onto each pair of qubits controlled on the ancilla qubit. One can show that the probability of the ancilla qubit being in the 0-state is $1/2 + 1/2\,|\langle\psi|\phi\rangle|^2$ [82]. In the following sections the quantum states are defined by unitaries that prepare them $|\phi\rangle = U_\phi\,|0\rangle$ and $|\psi\rangle = U_\psi\,|0\rangle$.

As the SWAP-test is an integral part of measuring overlaps in quantum computing, several more efficient implementations omitting the ancilla qubit of this test have been proposed. One way to measure this the overlap is by treating the SWAP operator as an observable, and finding the unitary that diagonalizes the observable. This is referred to as the destructive SWAP test [84] and depicted in Fig. 10. Another way of implementing the SWAP test in half the qubit requirements of the original SWAP-test [83] is by directly measuring $\langle\phi|\psi\rangle = \langle 0|\,U_\phi^\dagger U_\psi\,|0\rangle$ by implementing the transposed unitary and projecting onto the all zero state $|0\rangle^{\otimes n}$ by postselecting the measurement results.

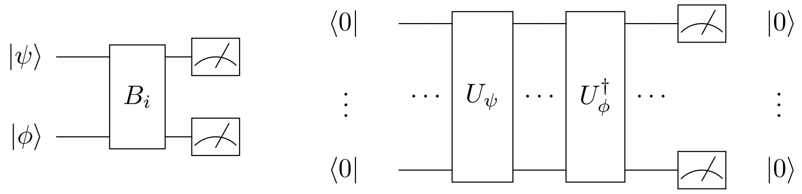


Figure 10: Different more efficient ways of implementing the SWAP test. Left: The destructive SWAP-test, diagonalizing the SWAP operator by B_i. Right: The efficient SWAP-test used in [83] which only uses half the qubit count of the original SWAP-test and the destructive SWAP test. Calculates $\langle\psi|\phi\rangle$ directly over $\langle 0| U_\phi^\dagger U_\psi |0\rangle$ by applying the transposed unitary defining $|\phi\rangle$ and projecting the measurement onto the all-zero state $|0\rangle^{\otimes n}$ by postselecting on the measurement outcomes.

Both of these methods are not compatible with the echo verification procedure above, so the verified SWAP-test is explicitly introduced, opening the SWAP-test for verification on the measurement result and correcting for incoherent errors. This can be formulated as described above also in a control-free manner doubling circuit depth, also excluding NISQ unfriendly controlled SWAP operations from the procedure, see Fig. 11.

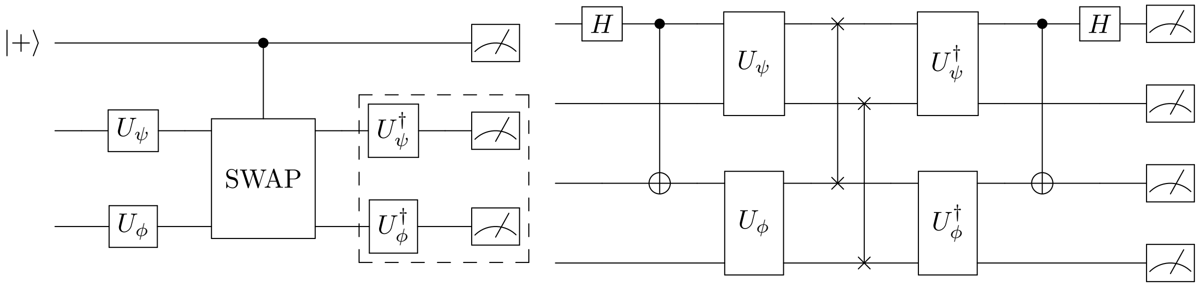


Figure 11: Left: Representation of the verified SWAP-test, allowing for verification and reducing of incoherent errors during circuit evaluation. Dashed lines mark the verification step omitted by the original SWAP-test. Right: Representation of the control-free verified SWAP-test on two qubits per system, exchanges the controlled SWAP-operation for a controlled state preparation.

Omitting the ancilla qubit is one benefit of this formulation, the robustness against decoherent noise is the main reason. In Section 3.6 the improvement over unmitigated estimations of expectation value is shown. The improvement of orders of magnitudes in precision against the unmitigated case shows that unmitigated estimations on noisy hardware suffer from large errors even in the infinite shot limit regime. As the SWAP-test naturally only aims to extract one bit of information, the restrictions that the Hadamard-test measurement setup brings in estimating larger observables discussed in Section 6 do not apply here, making the control-free verified SWAP-test an attractive measurement setting to estimate overlaps of states to high precision.

3.5 Virtual distillation

Multiple versions of this error mitigation technique were published around the same time labeled virtual distillation [85] or exponential error supression [86]. In virtual distillation (or exponential error suppression, the term virtual distillation is used for the remainder of this manuscript for both) one is concerned with a very similar task, estimating expectation value on a purified density matrix obtained by collective measurements on multiple copies of the noisy density matrix. The name virtual distilation hints here at the comparison with explicitly prepared purified states with projective measurements on multiple ancillas [82]. In comparison to echo verification discussed above one uses a redundancy in space instead of time [73], doubling the system size instead of the circuit depth. There is an equivalency when representing the echo verification process with the dual state $\bar{\rho}$. The spectral decomposition of the density matrix ρ_{noisy}, where the noise-free state $|\psi\rangle$ is

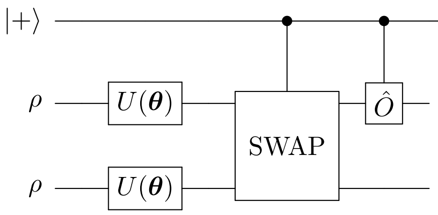


Figure 12: Representation of the virtual distillation framework in the formulation of [86] to evaluate $\mathrm{Tr}(\hat{O}\rho^2)$. $\mathrm{Tr}\left(\rho^2\right)$ can be evaluated by omitting $\hat{O}$.

still the dominant eigenvector with eigenvalue λ of the density matrix and incoherent noise process have scattered the state into eigenvectors $|\psi_k\rangle$ with smaller eigenvectors $(1-\lambda)p_k << \lambda$ is described by [86]

$$\rho = \lambda|\psi\rangle\langle\psi| + (1-\lambda)\sum_{k=2}^{2^N} p_k\,|\psi_k\rangle\langle\psi_k|\,. \tag{65}$$

The aim is to estimate $\langle\psi|\hat{O}|\psi\rangle$ to suppress contributions from $\langle\psi_k|\hat{O}|\psi_k\rangle$ in the estimate of the expectation value. To do so one prepares M copies of ρ_{noisy}, and in the case of no error one has prepared the state $|\psi, \psi \cdots, \psi\rangle$. Measurement on the first qubit in this ideal case retrieves $\left\langle \psi, \dots \psi, \psi \mid \hat{O}\psi, \psi, \dots \psi_k \right\rangle = \left\langle \psi|\hat{O}|\psi \right\rangle$ the wanted expectation value [86]. One assumes local decoherent noise processes on the quantum hardware and expects them to act on one of the copies of rho and have prepared the a state $|\psi_k, \psi, \cdots, \psi\rangle$. and contribute the bias [86]

$$\left\langle \psi_k, \dots \psi, \psi \mid \hat{O}\psi_k, \psi, \dots \psi_k \right\rangle = \left\langle \psi_k|\hat{O}|\psi_k \right\rangle \tag{66}$$

to the estimate of the expectation value. Now if one would have exchanged the first register with one in which the error did not occur or in other words would have estimated $\hat{O}\,\mathrm{SWAP}$, this bias would cancel due to the orthonormality of the eigenvectors of the density matrix [86]

$$\left\langle \psi_k, \dots \psi, \psi \mid \hat{O}\psi, \psi, \dots \psi_k \right\rangle = \left\langle \psi_k|\hat{O}|\psi \right\rangle \langle\psi_k \mid \psi\rangle = 0. \tag{67}$$

Now instead if just permuting two of the registers, one permutes all registers with all, an operation labelled as the *derangement operator* D_n. Any permutation that maps each site to any other site will do, for simplicity the cyclical permutation of indices is chosen as the derangement operator of choice. So the task now becomes in measuring $\hat{O}\,D_n$ and the estimator becomes

Virtual distillation

$$\langle\hat{O}\rangle_{\text{corrected}} := \frac{\mathrm{Tr}\left(\hat{O}\rho^M\right)}{\mathrm{Tr}\left(\rho^M\right)}. \tag{68}$$

where one can show that the derangement operator $\mathrm{Tr}\left(O\rho^M\right) = \mathrm{Tr}\left(O^1 D_M \rho^{\otimes M}\right)$ is equal in expectation value to the explicitly purified state [85].The denominator can be estimated in the same way by omitting the observable from the measurement and just measure the expectation value of the derangement operator. Fig. 12 shows the measurement setup for two copies of ρ in the SWAP-test formulation of [86]. The exponential suppression in M can be seen from

$$\mathrm{Tr}\left(\hat{O}\rho^M\right) = \lambda^n\langle\psi|\hat{O}|\psi\rangle + (1-\lambda)^n\sum_{k=2}^{2^N} p_k^n \left\langle \psi_k|\hat{O}|\psi_k \right\rangle, \tag{69}$$

where contributions $(1-\lambda)^n$ are exponentially suppressed, hence the name of the error mitigation technique. The derangement operator can be represented as a linear ladder of SWAP operators between all registers. All the different implementation now try to measure this estimator with different variants of the SWAP-Test mentioned above, exponential error suppression [86] is formulate in terms of the original SWAP test, [85] as the destructive SWAP test and the dual state purification [87] as the compact unitary SWAP test. The dual state purification is equivalent to the echo verification process described above in the controlled case and has a different projective measurement technique to achieve a control-free setting.

In fact, one can formulate echo verification in terms of this framework as [88]

$$\frac{\mathrm{Tr}\left(O\rho\bar{\rho}\right)}{\mathrm{Tr}\left(\rho\bar{\rho}\right)}, \tag{70}$$

where $\bar{\rho}$ is the dual state to ρ. This notation takes note of the fact that one cannot implement the complex conjugate of the original unitary under the given noise processes $(U(\boldsymbol{\theta})_\epsilon)^\dagger_\epsilon \neq U^\dagger(\boldsymbol{\theta})_\epsilon$, in explicit definitions of $\rho = U^\dagger(\boldsymbol{\theta})_\epsilon \left|0\right\rangle \left\langle 0\right| (U(\boldsymbol{\theta})_\epsilon)^\dagger$ and $\bar{\rho} = U^\dagger(\boldsymbol{\theta})_\epsilon \left|0\right\rangle \left\langle 0\right| (U^\dagger(\boldsymbol{\theta})_\epsilon)^\dagger_\epsilon$.

3.6 Combining purification techniques

After having discussed exponential error suppression and echo verification and explained their connections to the SWAP test, one can combine the two techniques and study them in a controlled manner. In Figure 13 the setup which combines echo verification and virtual distillation is shown, and by including/excluding certain operations one can do either or neither. By not verifying (excluding solid lines) one recovers the virtual distillation setup with two copies of ρ from [86]. The normalizing factor of $\mathrm{Tr}(\rho^2)$ can be estimated by excluding the dashed segment of the circuit. When excluding the controlled-SWAP (excluding dotted lines) which is equivalent to the two qubit derangement operator and only measuring on the first system, omitting the second, one recovers the echo verification setup from Section 3.3. The normalizing factor of echo verification can be estimated by excluding the dashed and dotted part.

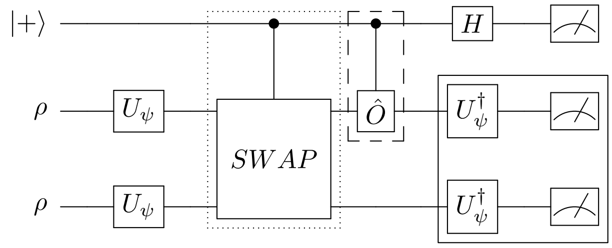


Figure 13: Combining echo verification and virtual distillation - controlled derangement/SWAP operator D (dotted), controlled unitary observable $\hat{O}$ (dashed), and the inverse ansatz and state preparation plus verification (solid).

The combination of the two techniques, also introduced here [88], can be viewed through different lenses. As all virtual distillation papers implement a different SWAP test to measure the observables needed for error mitigation, the state verified virtual distillation uses the verified SWAP test introduced in Section 3.4 to estimate state verified values for $\mathrm{Tr}_{\mathrm{ver}}\left(\hat{O}\rho^2\right)$ and $\mathrm{Tr}_{\mathrm{ver}}\left(\rho^2\right)$. The normalization factor of echo verification cancels out. This can also be expressed as the following equation in similar fashion to Eq. 70, as covered by [88]

$$\frac{\mathrm{Tr}\left(O(\rho\bar{\rho})^2\right)}{\mathrm{Tr}\left((\rho\bar{\rho})^2\right)}. \tag{71}$$

The studied behaviour of the different error mitigation schemes under different noise models is depicted in Fig. 14, plotting the absolute error of the estimation in the infinite shot limit in dependency on the parameter of the noise channel applied after every gate. Simulations are done using Pennylane [89], ρ is spanning 4 qubits and determined by a randomly initialized QNP-fabric with 25 gates. For echo verification, virtual distillation and the combination of the two the actual circuits are implemented and measured to estimate the error corrected observables from. For the virtual distillation curve with 4 copies $\mathrm{Tr}(\hat{O}\rho^4)/\,\mathrm{Tr}\left(\rho^4\right)$ the noisy estimate of ρ is exponentiated classically from a single estimate of ρ. Noise channels are applied after every gate with probability p. Estimations are taken over 50 random initializations of the circuit. For each random configurations of parameter, a random observable consisting of Z_i or Z_iZ_j are chosen to be estimated. This corresponds to individual terms needed in the double-factorized energy (Section 2.4), as the diagonalizing unitary can be absorbed into the ansatz. In Fig. 15 the fidelities of the state verifying methods are plotted, giving an idea of the shot count overhead at varying noise levels as a result of the verification on the system qubits.

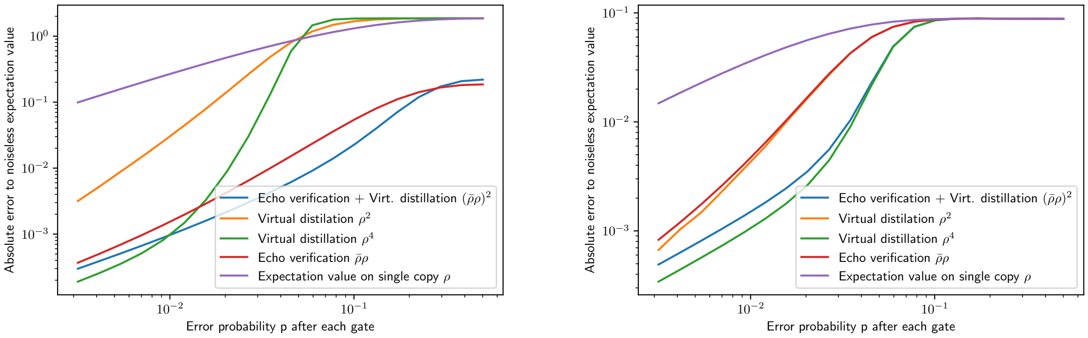


Figure 14: Left: Performance of the different studied error mitigation technique under dephasing and amplitude dampening noise. Right: Performance of the different studied error mitigation technique under depolarizing noise.

All error mitigation techniques improve over the standard evaluation of the unmitigated observable. Above error probabilities of 0.06 the dominant eigenvector of ρ is no longer the noiseless state, resulting in the purified methods converging onto a wrong state. At low noise levels below $\approx 10^{-3}$ all tested purification methods converge to the limit of $\operatorname{Tr}(\hat{O}\rho^M)/\operatorname{Tr}(\rho^M)$ where $\lim_{M\to\infty}$, as discussed in [86, 85]. This sets the lower boundary of error these methods can achieve. Under depolarizing and dephasing noise the state verified methods seem to outperform the pure distillation methods. At noise levels of $p \approx 0.08$ the better behaviour of the state verifying methods comes at an unfeasible exponential cost in shot count, see Fig. 15, which one has to keep in mind when viewing these plots as the simulations in Fig. 14 are taken without shot noise which corresponds to the infinite shot limit. Under depolarizing noise (Fig. 14, right) the methods with the same powers of ρ perform very similar, no matter if including state verification or not.

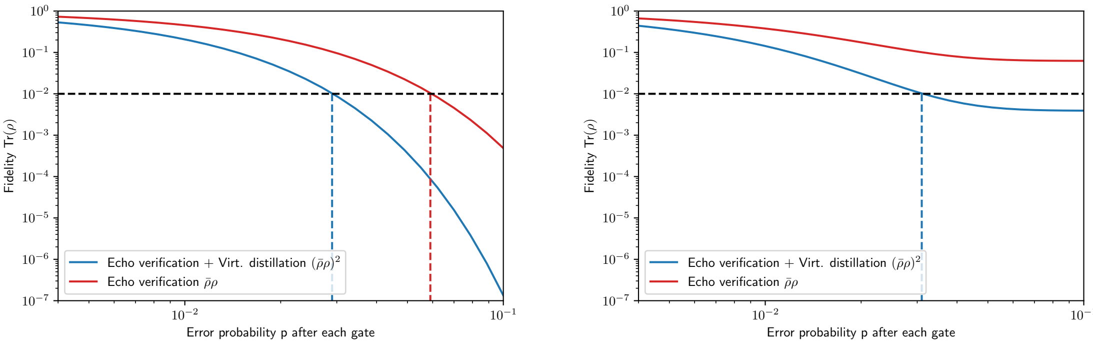


Figure 15: Left: The fidelity and therefore the shot overhead of the error mitigation methods for phase and amplitude damping noise Right: The fidelity and therefore the shot overhead of the error mitigation methods for depolarizing noise, dashed lines mark the error probability p where the mitigation methods multiplies the original needed shot count by a factor of 100.

If one has characterized device noise for the hardware of choice, different regimes of error profiles seem to have different optimal choices of error mitigation for the scenario investigated here: When mostly amplitude damping and dephasing contribute to the noise processes the echo verification protocol seems to be the method of choice in terms overall hardware cost and in reduction of contribution of incoherent error. If mostly depolarizing noise is present on the device, the combination of virtual distillation and echo verification implements a purification on four copies of ρ with only doubling circuit length and qubit count, both reasonable overall costs as shown in Section 4 and an increase in shot count depending on the overall fidelity of the circuit. At low error probabilities where one is not in the region where different error mitigation techniques converge to the lower limit $\operatorname{Tr}(\hat{O}\rho^M)/\operatorname{Tr}(\rho^M)$ with $\lim_{M\to\infty}$ at different rates, the choice of purification error mitigation technique matters less as all are bound by the same limit.

3.7 Coherent errors in purification error mitigation

Purification methods allow to correct for most of the incoherent noise occurring in a quantum circuit as discussed above. This comes at the cost of the variational principle in the individual parameters that protected against coherent noise like over/under rotations. Now these offsets between unique occurrences of the gate associated with the parameter could have different offsets which are not variationally optimized by default. In echo verification one reuses the same qubit register in the uncomputation of the state, making this is of lesser concern, as both parts of the gate $U_i(\theta_i + \gamma_i)$ and $U_i^\dagger(\theta_i + \gamma_i)$ are usually associated with the same qubit(s) and therefore the same calibration and share a coherent offset. As in virtual distillation with two copies, both copies are prepared on two different qubit registers and therefore subject to slight offsets in calibration as $U_{i,1}(\theta + \gamma_{i,1})$ and $U_{i,2}(\theta + \gamma_{i,2})$. As noted by [86], one can mitigate these effects by slightly adjusting the parameters of the two halves to maximize $\operatorname{Tr}(\rho_1\rho_2)$ to compensate for the offset $\gamma_{i,1} - \gamma_{i,2}$. We describe an efficient way of adjusting these parameters when subject to coherent noise and investigate if this can recover the original purity without coherent noise even in the presence of incoherent noise, which might interfere with finding the coherent offset.

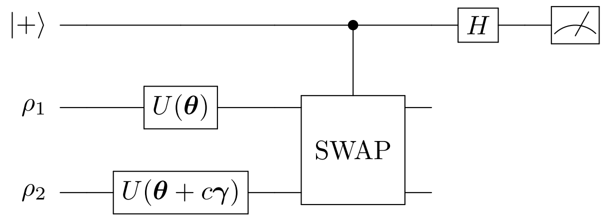


Figure 16: Representation of the investigated measurement configuration to compensate for coherent errors in virtual distillation to measure and maximize $\operatorname{Tr}(\rho_1\rho_2)$.

The procedure involves running a single step of Rotosolve [72] on one of the qubit registers and maximizing the measured purity $\operatorname{Tr}(\rho_1\rho_2)$ which scales linearly $\mathcal{O}(N_\theta)$ in the number of parameters N_θ in the ansatz. As the local cost function is maximized when the individual parameters match, Rotosolve efficiently finds the offset in a single step. The virtual distillation formulation of Section 13 is used and the exact measurement setup is shown in Fig. 16. By adding a slight offset $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{N_\theta})$ to one of the circuits, following a normal distribution scaled by a factor of c, we simulate a coherent mismatch of varying strength. By unlocking the parameters on each side from each other, a single step of Rotosolve adjusts the original parameters $\boldsymbol{\theta}$ to the perturbed ones $\boldsymbol{\theta}_{\text{opt}}$ one can extract the offset $\boldsymbol{\gamma}_{\text{opt}} = \boldsymbol{\theta}_{\text{opt}} - \boldsymbol{\theta}$. This offset can then be added to the other half of the circuit to compensate for the coherent offset in further evaluations of observables. In the simulations shown here, this offset is found for every datapoint individually as they are simulated with different $\boldsymbol{\gamma}$. In real error mitigation scenarios, this offset can be learned and used for multiple evaluations of the circuit, depending on the individual performance of the used hardware and the time scale of calibration and device drift. In virtual distilation scenarios with more than $N > 2$ copies of ρ, one can adjust the offset by running the procedure $N - 1$ times to find the offset of all copies in relation to a chosen base copy of ρ.

A numerical demonstration of this procedure is shown in Figure. 17. The ansatz from Section 13 is used with incoherent noise probabilities of p between 10^{-3} and 10^{-2} to show that even in regimes with significant incoherent noise levels the procedure can extract the coherent mismatch. Fig. 17

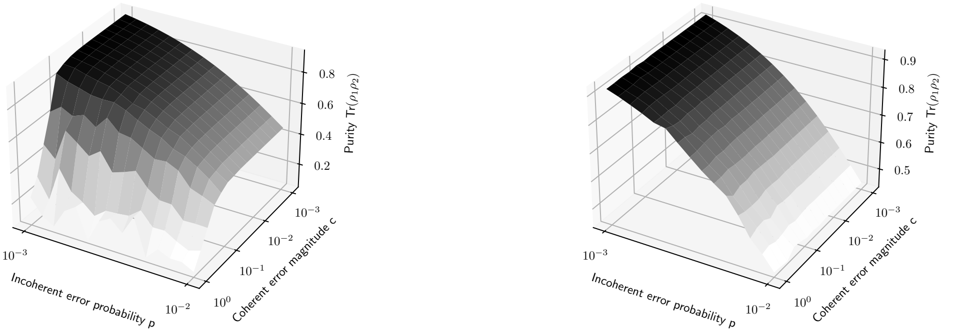


Figure 17: The measured purity $\mathrm{Tr}(\rho_1\rho_2)$ in virtual distillation in the output state when dealing with incoherent errors of probability p after every quantum gate and coherent errors taking the form of over and under rotation of magnitude c. Left: Uncorrected purity with $\rho_1 = U(\boldsymbol{\theta})_\epsilon \left|0\right\rangle \left\langle 0\right| U^\dagger(\boldsymbol{\theta})_\epsilon$, $\rho_2 = U(\boldsymbol{\theta} + c\boldsymbol{\gamma})_\epsilon \left|0\right\rangle \left\langle 0\right| U^\dagger(\boldsymbol{\theta} + c\boldsymbol{\gamma})_\epsilon$ Right: Corrected purity with $\rho_1 = U(\boldsymbol{\theta} + \boldsymbol{\gamma}_{\mathrm{opt}})_\epsilon \left|0\right\rangle \left\langle 0\right| U^\dagger(\boldsymbol{\theta} + \boldsymbol{\gamma}_{\mathrm{opt}})_\epsilon$, $\rho_2 = U(\boldsymbol{\theta} + c\boldsymbol{\gamma})_\epsilon \left|0\right\rangle \left\langle 0\right| U^\dagger(\boldsymbol{\theta} + c\boldsymbol{\gamma})_\epsilon$.

left shows the default behaviour, where for coherent errors larger than 10^{-1} the purity of the state is significantly reduced. Each point is taken over 10 random offsets $\boldsymbol{\gamma}$ and the depolarizing channel is used as a source of incoherent error. Fig. 17 right shows that after a single step of the optimizer one recovers the behaviour of low coherent noise mismatch recovering the purity with only incoherent noise. This shows that these purification techniques can be made to tolerate coherent noise in addition to incoherent noise. As these simulations are taken without shot noise, the accuracy to which one can calibrate these offsets is set by the shot budget one allocates for this optimization step and depending on calibration drift at what frequency this offset has to be reoptimized.

Further techniques to compensate for coherent errors are the quantum subspace expansion [90] and the generalized subspace expansion [91], which combines the quantum subspace expansion with virtual distillation.

4 Purification-based quantum error mitigation of pair - correlated electron simulation

Bibliographic information

O'Brien, T. E., **Anselmetti, G.**, Gkritsis, F., Elfving, V. E., Polla, S., Huggins, W. J., ... Rubin, N. C. (2022, October). Purification-based quantum error mitigation of pair-correlated electron simulations. arXiv. Retrieved 2023-05-31, from `http://arxiv.org/abs/2210.10799` (arXiv:2210.10799 [quant- ph]) doi: 10.48550/arXiv.2210.1079, Accepted in Nature Physics

Summary

We investigate the performance of different purification based error mitigation techniques on a superconducting quantum computer (Googles Sycamore chip). As quantum circuits under noise are quite costly to simulate in comparison to perfect state simulators and that noise channels that one analytically implements always will only approximate the true noise environment on a device, performance of error mitigation techniques is best evaluated on real hardware. As previously mentioned one needs accurate results (accurate up to chemical accuracy) in quantum chemistry for post-Hartree-Fock methods so in this work we investigate how precisely one can resolve energies in the DOCI approximation for quantum chemistry as well as the energy and order parameters for the Richardson-Gaudin Model, a toy model for superconductivity. We find an improvement of two orders of magnitude over post selection and a favorable scaling in performance of purification with increasing system size, and extract scaling laws. This constitutes one of the largest and most complex VQEs up to date on quantum hardware.

Contributions

The author contributed to this work: (1) the ansatz, which constitutes the shortest Trotter step of pairwise interactions between virtual and occupied orbitals by lifting the 1-dimensional constraint of the ansatz and using a SWAP-network on a ring instead of a line. (2) Setting up the infrastructure of making the experiment work, (3) building compilers which preprocess Hamiltonians into jointly-measurable terms, (4) build quantum circuits according to the used purification scheme, and postprocess the measured values into purified expectation values. (5) Discussion and simulations of potential downfalls of virtual distillation on two different copies of ρ and correcting for coherent errors. (6) Numerical simulations of the experiment, including investigation of pre-initalisations of the circuit parameters and optimal layouts, and helped writing the manuscript.

Purification-based quantum error mitigation of pair-correlated electron simulations

T. E. O'Brien,[1, 2, *] G. Anselmetti,[3] F. Gkritsis,[3] V. E. Elfving,[4] S. Polla,[1, 2] W. J. Huggins,[1] O. Oumarou,[3] K. Kechedzhi,[1] D. Abanin,[1] R. Acharya,[1] I. Aleiner,[1] R. Allen,[1] T. I. Andersen,[1] K. Anderson,[1] M. Ansmann,[1] F. Arute,[1] K. Arya,[1] A. Asfaw,[1] J. Atalaya,[1] D. Bacon,[1] J. C. Bardin,[1, 5] A. Bengtsson,[1] S. Boixo,[1] G. Bortoli,[1] A. Bourassa,[1] J. Bovaird,[1] L. Brill,[1] M. Broughton,[1] B. Buckley,[1] D. A. Buell,[1] T. Burger,[1] B. Burkett,[1] N. Bushnell,[1] J. Campero,[1] Y. Chen,[1] Z. Chen,[1] B. Chiaro,[1] D. Chik,[1] J. Cogan,[1] R. Collins,[1] P. Conner,[1] W. Courtney,[1] A. L. Crook,[1] B. Curtin,[1] D. M. Debroy,[1] S. Demura,[1] I. Drozdov,[1, 6] A. Dunsworth,[1] C. Erickson,[1] L. Faoro,[1] E. Farhi,[1] R. Fatemi,[1] V. S. Ferreira,[1] L. Flores Burgos,[1] E. Forati,[1] A. G. Fowler,[1] B. Foxen,[1] W. Giang,[1] C. Gidney,[1] D. Gilboa,[1] M. Giustina,[1] R. Gosula,[1] A. Grajales Dau,[1] J. A. Gross,[1] S. Habegger,[1] M. C. Hamilton,[1, 7] M. Hansen,[1] M. P. Harrigan,[1] S. D. Harrington,[1] P. Heu,[1] J. Hilton,[1] M. R. Hoffmann,[1] S. Hong,[1] T. Huang,[1] A. Huff,[1] L. B. Ioffe,[1] S. V. Isakov,[1] J. Iveland,[1] E. Jeffrey,[1] Z. Jiang,[1] C. Jones,[1] P. Juhas,[1] D. Kafri,[1] J. Kelly,[1] T. Khattar,[1] M. Khezri,[1] M. Kieferová,[1, 8] S. Kim,[1] P. V. Klimov,[1] A. R. Klots,[1] R. Kothari,[1] A. N. Korotkov,[1, 9] F. Kostritsa,[1] J. M. Kreikebaum,[1] D. Landhuis,[1] P. Laptev,[1] K. Lau,[1] L. Laws,[1] J. Lee,[1, 10] K. Lee,[1] B. J. Lester,[1] A. T. Lill,[1] W. Liu,[1] W. P. Livingston,[1] A. Locharla,[1] E. Lucero,[1] F. D. Malone,[1] S. Mandra,[1] O. Martin,[1] S. Martin,[1] J. R. McClean,[1] T. McCourt,[1] M. McEwen,[1, 11] A. Megrant,[1] X. Mi,[1] A. Mieszala,[1] K. C. Miao,[1] M. Mohseni,[1] S. Montazeri,[1] A. Morvan,[1] R. Movassagh,[1] W. Mruczkiewicz,[1] O. Naaman,[1] M. Neeley,[1] C. Neill,[1] A. Nersisyan,[1] H. Neven,[1] M. Newman,[1] J. H. Ng,[1] A. Nguyen,[1] M. Nguyen,[1] M. Y. Niu,[1] S. Omonije,[1] A. Opremcak,[1] A. Petukhov,[1] R. Potter,[1] L. P. Pryadko,[1, 12] C. Quintana,[1] C. Rocque,[1] P. Roushan,[1] N. Saei,[1] D. Sank,[1] K. Sankaragomathi,[1] K. J. Satzinger,[1] H. F. Schurkus,[1] C. Schuster,[1] M. J. Shearn,[1] A. Shorter,[1] N. Shutty,[1] V. Shvarts,[1] J. Skruzny,[1] V. Smelyanskiy,[1] W. C. Smith,[1] R. Somma,[1] G. Sterling,[1] D. Strain,[1] M. Szalay,[1] D. Thor,[1] A. Torres,[1] G. Vidal,[1] B. Villalonga,[1] C. Vollgraff Heidweiller,[1] T. White,[1] B. W. K. Woo,[1] C. Xing,[1] Z. J. Yao,[1] P. Yeh,[1] J. Yoo,[1] G. Young,[1] A. Zalcman,[1] Y. Zhang,[1] N. Zhu,[1] N. Zobrist,[1] C. Gogolin,[3, †] R. Babbush,[1, ‡] and N. C. Rubin[1, §]

[1] *Google Research, Mountain View, CA 94043, United States*
[2] *Instituut-Lorentz, Universiteit Leiden, 2300 RA Leiden, The Netherlands*
[3] *Covestro Deutschland AG, 51365 Leverkusen, Germany*
[4] *Qu & Co B.V., 1070 AW Amsterdam, The Netherlands*
[5] *Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst MA, USA*
[6] *Department of Physics, University of Connecticut, Storrs, CT*
[7] *Department of Electrical and Computer Engineering, Auburn University, Auburn, AL*
[8] *QSI, Faculty of Engineering and Information Technology, University of Technology Sydney, NSW, Australia*
[9] *Department of Electrical and Computer Engineering, University of California, Riverside, CA USA*
[10] *Department of Chemistry, Columbia University, New York, NY, USA*
[11] *Department of Physics, University of California, Santa Barbara, CA, USA*
[12] *Department of Physics and Astronomy, University of California, Riverside, CA*

An important measure of the development of quantum computing platforms has been the simulation of increasingly complex physical systems. Prior to fault-tolerant quantum computing, robust error mitigation strategies are necessary to continue this growth. Here, we study physical simulation within the seniority-zero electron pairing subspace, which affords both a computational stepping stone to a fully correlated model, and an opportunity to validate recently introduced "purification-based" error-mitigation strategies. We compare the performance of error mitigation based on doubling quantum resources in time (echo verification) or in space (virtual distillation), on up to 20 qubits of a superconducting qubit quantum processor. We observe a reduction of error by one to two orders of magnitude below less sophisticated techniques (e.g. post-selection); the gain from error mitigation is seen to increase with the system size. Employing these error mitigation strategies enables the implementation of the largest variational algorithm for a correlated chemistry system to-date. Extrapolating performance from these results allows us to estimate minimum requirements for a beyond-classical simulation of electronic structure. We find that, despite the impressive gains from purification-based error mitigation, significant hardware improvements will be required for classically intractable variational chemistry simulations.

---

[*] teobrien@google.com
[†] christian.gogolin@covestro.com
[‡] babbush@google.com
[§] nickrubin@google.com

The prospect of accurately simulating ground states of quantum systems on quantum hardware has motivated substantial theory and hardware developments over the last decade. With fault-tolerant quantum computing in its infancy [1] and many years from promised applications [2–8] attention has focused on algorithms requiring only short-depth quantum circuits, such as the variational quantum eigensolver (VQE) [9–11]. Theoretical developments in ansatz design [10, 12–16] and measurement optimization [17–22] have enabled small to mid-scale VQE experiments [14, 23–30]. A key target of variational quantum algorithms has been the electronic structure problem in chemistry [10, 14, 23–27, 31]. Such simulations are challenging to implement on quantum hardware due to a long-range two-body fermionic Hamiltonian and stringent accuracy requirements. This makes it unclear whether a beyond-classical [32] simulation of chemistry can be achieved without fault tolerance. Determining the requirements for such a simulation is a critical open problem.

The electronic structure problem can be expressed in models of varying complexity and realism. Quantum simulations of chemistry within the Hartree-Fock (mean-field) approximation were implemented for system sizes up to 12 qubits in [27], and this retains the record for the largest VQE calculation of a chemical ground state on quantum hardware. As a next step, one can consider working in the seniority zero subspace of the entire Hilbert space, which assumes all electrons come in spin-up or spin-down pairs [15, 33–37]. This has the advantage of projecting a local fermionic problem onto a local qubit problem [15]. The S_0 ground state is not *a priori* classically efficiently simulable [15] (though good approximate methods are known to exist for many problems [38–40]). This makes it a good stepping stone beyond Hartree-Fock towards the full electronic structure problem.

Recent quantum experiments have relied on error mitigation techniques [41], which are not scalable like error correction [1, 42], but promise to substantially shrink experimental errors. Popular methods are based on postselection [43, 44], rescaling [45–48], purification [27, 49–51] and probabilistic cancellation [45, 52]. Various schemes and combinations of error mitigation techniques have been implemented in practice [24, 26–30, 47, 53]. However, many of these methods do not promise to remove bias to the level of accuracy needed for useful simulation of chemistry, or remain untested beyond few-qubit experiments. Shifting from non-interacting fermions to correlated electronic structure, one loses two error mitigation advantages that were crucial to the success of [27]: efficient density matrix purification via McWeeny iteration [54], and low-cost gradient estimation.

In this work, we mitigate errors accumulated during the preparation of electronic ground states in the seniority-zero space, comparing three different error mitigation techniques — postselection, echo verification, and virtual distillation — on up to 20 qubits of a superconducting quantum processor. Using either echo verification or a new combination of postselection and virtual distillation, we are able to reproduce the ground state energy and order parameter for an $N = 10$-qubit simulation of the Richardson-Gaudin (RG), or pairing model — the quintessential model of superconductivity — improving over the unmitigated estimates by $1 - 2$ orders of magnitude. This demonstrates an improvement over classical pair-coupled-cluster-doubles, and the non-interacting BCS theory, neither of which are qualitatively correct over the entire range of coupling values considered. Echo verification was further able to significantly improve over postselected VQE for 6- and 10-qubit simulations of the ring-opening of cyclobutene. While the stringent error requirements (< 0.05 Hartree) to differentiate between mean-field and the exact solution could only be achieved for the 6-qubit case, this still represents the largest VQE simulation of electronic structure for chemistry to date.

Finally, we considered the scaling of our simulation of the RG model, using data from simulations at $N = 4, 6, 8, 10$. We observe a clear difference in the asymptotic scaling of the mean absolute error in energy and order parameter when echo verification or virtual distillation are applied. From this data, we are able to estimate the minimum requirements for a beyond-classical VQE simulation of similar form: a 25× decrease in hardware error rates (from those observed in this work), a limit of $O(N)$-depth for future variational ansatzes, and the need to pre-optimize ansatzes classically without intermediate calls to a device. Even if this list of requirements is achieved, meeting the high level of accuracy required for the electronic structure problem will pose a serious challenge, as chemical accuracy is around 60× smaller than our mean accuracy for the 10-qubit cyclobutene problem.

I. METHODS

A. Simulating the seniority-zero subspace

The seniority of a Slater determinant is the number of unpaired electrons; thus, the seniority zero (S_0) sector of Hilbert space for an N-electron system in M orbitals is the space of $\binom{M}{N/2}$ determinants leaving no electrons unpaired given a particular pairing of the spin-orbitals. Seniority is not a global symmetry of the electronic structure Hamiltonian and it is basis dependent; it has been used as a way to classify determinant subspaces to generate better approximations for solving the Schrödinger equation [34–37] and as a starting point for modeling strong correlations from electron pair states [33].

Supported by the S_0 subspace there exists a set of operators satisfying the $su(2)$ algebra constructed from pairs of fermion ladder operators and the spatial orbital number operator [55]

$$P_p^\dagger = a^\dagger_{p\alpha} a^\dagger_{p\beta} \;\;,\;\; N_p = \sum_{p,\sigma} a^\dagger_{p\sigma} a_{p\sigma},$$

$$[P_p, P_q^\dagger] = (1 - N_p)\delta_{p,q} \;\;,\;\; [N_p, P_q] = -2P_q\delta_{p,q}, \qquad (1)$$

where p, q and α, β are orbital and spin indices respectively. These operators form a basis for Hamiltonians projected into the S_0 subspace. The equivalence to an $su(2)$ algebra means seniority zero models resemble Heisenberg spin$-1/2$ models which are easily expressed as Pauli operators.

In this work we focus on two Hamiltonians to validate purification-based error mitigation strategies. The first is the Richardson-Gaudin (RG), or pairing model

$$\hat{H} = \sum_{p=1}^{N} \epsilon_p N_p + g \sum_{p \neq q = 1}^{N} P_p^\dagger P_q, \qquad (2)$$

which is a model for a small superconducting grain when $g < 0$ [56–59], but with a g-dependent Debye frequency [57]. The second model is the electronic structure Hamiltonian (H_{elec}) projected into the S_0 subspace

$$H_{\mathrm{S}_0} = P_{\mathrm{S}_0} H_{\mathrm{elec}} P_{\mathrm{S}_0} = \sum_p (h_{p,p})\, N_p \qquad (3)$$

$$+ \frac{1}{4} \sum_{p \neq q} (2V_{pqpq} - V_{pqqp})\, N_p N_q + \sum_{pq} (V_{ppqq})\, P_p^\dagger P_q.$$

The all-to-all connected Heisenberg spin Hamiltonian is, in general, not known to be classically solvable, but good approximate methods exist. This is especially true for the RG model, which is often integrable [39], well-approximated by density-matrix renormalization group [38] and pair coupled cluster techniques in the repulsive regime, and solvable by quantum Monte Carlo in the attractive regime (where it has no sign problem). Pair coupled cluster theory is also known to work well for the electronic structure problem in the S_0 subspace [36, 40, 60, 61] while full configuration interaction quantum Monte Carlo shows a reduced sign problem [62]. As such, although we have strong evidence that the quantum circuits used in this text are not classically simulatable (App. B 1), we do not believe directly scaling S_0 simulations represents the easiest path to a quantum advantage in chemistry; this is instead a stepping stone between a mean field solution and the full electronic structure problem.

B. The unitary pair coupled cluster ansatz and energy estimation

In this work we use a Trotterized unitary pair-coupled-cluster doubles (UpCCD) ansatz [15] compiled into a set of qubits in a ladder geometry with nearest-neighbor coupling. The ladder ansatz (instead of a generic ring) allows us to efficiently measure terms in the Hamiltonian corresponding qubits that are not physically adjacent after encoding with a minimal number of SWAP operations. When mapped from fermions to qubits the UpCCD ansatz has the form

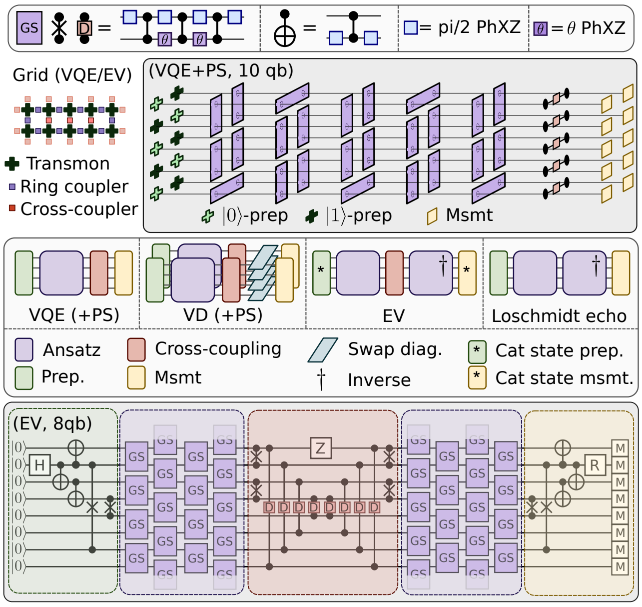


FIG. 1. The UpCCD ansatz and its compilation to a 2D superconducting transmon grid. (top) Decomposition of the gates used in this experiment to CZ and single-qubit gates. See supplemental material for details. (second from top, left) 2×5 grid with couplers in a square lattice geometry, showing couplers used during the ansatz (ring coupler, purple), and those used only during measurement (cross-coupler, red). (second from top, right) 2+1D circuit cartoon of a combined ansatz and measurement on a 2×5 transmon qubit array. (third from top) Cartoon of error mitigation techniques used in this experiment. Different circuit pieces are described in the legend. (bottom) an example 8-qubit echo verification circuit to measure the expectation value of $(X_1X_7+Y_1Y_7+Z_1+Z_7)/2$. Shaded gates at the top and bottom of the qubit array wrap around the 2×4 ring.

$$U(\theta) = \prod_{\ell}^{N_\ell} U_e(\theta^{e,\ell}) U_o(\theta^{o,\ell}) \qquad (4)$$

$$U_o(\theta^{o,\ell}) = \prod_{n=0}^{N/2-1} GS_{2n+1,(2n+2)\%N}(\theta^{o,l}_{2n+1,(2n+2)\%N}) \qquad (5)$$

$$U_e(\theta^{e,\ell}) = \prod_{n=0}^{N/2-1} GS_{2n,2n+1}(\theta^{e,l}_{2n,2n+1}) \qquad (6)$$

where each $GS_{ij}(\theta^{e/o,\ell}_{ij})$ is a Givens-SWAP gate corresponding to the product of a Givens rotation gate on a pair labeled by qubits i and j followed by a SWAP operation [15],

$$GS(\phi)=\underbrace{\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}}_{\text{SWAP}}\underbrace{\begin{bmatrix}1&0&0&0\\0&\cos(\phi)&-\sin(\phi)&0\\0&\sin(\phi)&\cos(\phi)&0\\0&0&0&1\end{bmatrix}}_{\text{Givens}}. \quad (7)$$

The GS gate corresponds to a coherent partial pair-excitation (by the angle ϕ), followed by a pair-SWAP. Given a number of layers N_ℓ in Eq. (4) and total number of qubits N there are a total of $N_\ell N/2$ free parameters in the ansatz. To minimize the amount of time qubits are idle we order the spatial orbitals such that the Fermi-vacuum is $|0101\ldots01\rangle$–e.g. the restricted Hartree-Fock state–corresponding to an interleaved list of occupied and virtual orbital labels in ascending energy order. The Hamiltonian qubit ordering is then chosen such that when all $\theta=0$, the Hartree-Fock state for each model is returned. The alternating SWAP gate arrangement allows us to couple each occupied pair with each unoccupied pair once in depth $N/2$ (see App. B 3). Thus, in this work we set $N_\ell = N/2$ for all systems. Each $GS(\theta)$ gate is compiled into a product of three controlled-Z (CZ) gates interleaved with tunable single-qubit microwave gates (Fig. 1 (top), see App. B 2).

To perform energy estimation on our two S_0 models, expectation values with respect to nearest-neighbor and non-nearest-neighbor qubits are required. The expectation value $\langle X_iX_j + Y_iY_j\rangle$ is estimated by performing a number preserving diagonalization [19, 27] mapping the expectation value to the difference of $\langle Z_i\rangle$ and $\langle Z_j\rangle$. The ladder geometry allows us to measure all non-nearest-neighbor pairs across the rungs of the ladder in a similar fashion at the additional cost of at most one SWAP operation. The full measurement protocol is detailed in Appendix B 3. All-to-all coupling is achieved in N circuits bringing the total number of different circuits to measure the Hamiltonian's expectation value to $N+1$. Strategies with fewer numbers of circuits exist, however they do not allow for post-selection on particle number.

C. Echo verification and virtual distillation

Echo verification (EV), introduced in [51] is an error mitigation technique that uses two copies of a quantum state $|\psi\rangle$ reflected in time (preparation ↔ unpreparation) to estimate $\langle\psi|O|\psi\rangle$ for a unitary O [53, 63]. EV can be implemented without control gates, given a known reference eigenstate $|\phi\rangle$ of O orthogonal to $|\psi\rangle$ (here $|\phi\rangle = |00\ldots\rangle$). To implement (control-free) EV, we act O on a prepared superposition of $|\psi\rangle$ and $|\phi\rangle$, generated by acting our UpCCD ansatz on the cat state $|00\ldots0\rangle + |0101\ldots01\rangle$. Then, we estimate the expectation value of $|\phi\rangle\langle\psi|$ [1] on the resulting state $|\Psi\rangle = O\frac{1}{\sqrt{2}}(|\psi\rangle + |\phi\rangle)$. The estimation is performed by inverting the preparation unitary. In the absence of noise, we have

$$\langle\Psi|\phi\rangle\langle\psi|\Psi\rangle = \tfrac{1}{2}\langle\psi|O|\psi\rangle e^{i\phi}, \quad (8)$$

where $O|\phi\rangle = e^{i\phi}|\phi\rangle$. The expectation value $\langle\psi|O|\psi\rangle$ can be recovered from Eq. (8) as the other terms are known. The largest effect of noise on the system is to dampen $\langle\Psi|\phi\rangle\langle\psi|\Psi\rangle \to F\langle\Psi|\phi\rangle\langle\psi|\Psi\rangle$, where F is the circuit fidelity [51]. We can estimate F independently by removing O from the circuit, which yields a Loschmidt echo of the preparation unitary [64]. This is achieved in practice by removing a virtual Z rotation (see Fig. 1, bottom), making the estimated Loschmidt fidelity an accurate estimate of F. Further EV implementation details can be found in App. B 4.

Virtual distillation (VD) [49, 50] is an error mitigation technique that uses collective measurements of k copies of a state ρ to estimate expectation values with respect to $\rho^k/\mathrm{Tr}[\rho^k]$. VD schemes are based on the observation that the cyclic shift operator $S^{(k)}$ is easily diagonalized, and therefore can be measured, which yields e.g. for $k=2$

$$\mathrm{Tr}[\rho\otimes\rho S^{(2)}] = \mathrm{Tr}[\rho^2], \mathrm{Tr}[\rho\otimes\rho S^{(2)}O_s] = \mathrm{Tr}[\rho^2 O], \quad (9)$$

with $O_s = \frac{1}{2}(I\otimes O + O\otimes I)$. $S^{(2)}$ can be simultaneously diagonalized with O_s when $O = Z_i$ by a $GS(\pi/4)$ rotation between pairs of identified qubits on the two registers. For two $N/2\times 2$ ladders on a square lattice geometry, this requires one round of SWAP gates to shift identified qubits next to each other. Operators $O \neq Z_i$ are measured by rotating to Z_i (see Sec. I B) and following the above procedure. The virtual distillation circuit is only 6 two-qubit gates deeper than post-selected VQE.

As the $GS(\pi/4)$ gate is number-conserving, VD can be combined with postselection: the global excitation number $\sum_j(Z_j\otimes I + I\otimes Z_j)$ is a good symmetry. This requires that the state prior to measurement also conserve number. This is true when estimating $\langle X_iX_j + Y_iY_j\rangle$, but not when estimating $\langle Z_iZ_j\rangle$: when mapping $Z_iZ_j \to Z_i$ one can only preserve the parity of the total number of excitations. In the main text of this work, we will present results showing VD with postselection only (PS-VD). We compare VD with and without postselection in App. F 1.

II. THE RICHARDSON-GAUDIN MODEL

We use our UpCCD ansatz to prepare approximate ground states of the RG model on 10 sites at half-filling across a range of coupling strengths using parameters optimized in noiseless simulations. We achieve half-filling

---

[1] The term $|\phi\rangle\langle\psi|$ is not Hermitian, but may be written as a sum of the Hermitian operators $|\phi\rangle\langle\psi| + |\psi\rangle\langle\phi|$ and $i|\phi\rangle\langle\psi| - i|\psi\rangle\langle\phi|$.

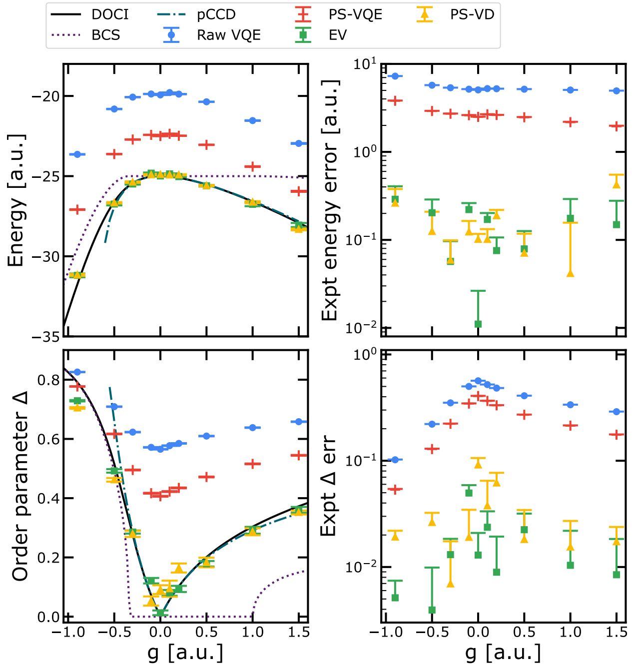


FIG. 2. Digital quantum simulation of ground states of the RG Hamiltonian for 10 spatial orbitals (Eq. (2) and Eq. (10)) on a superconducting quantum device. (top-left) Energy as a function of the coupling parameter g, for an unmitigated state preparation [blue circles], and state preparation mitigated by postselection [red crosses], echo verification [yellow triangles], and postselected virtual distillation [green squares]. This is compared to the exact DOCI result [black solid line], and BCS [purple dashed line] and pCCD [teal dashed-dotted line] classical approximations. The pCCD results do not converge below a critical value, resulting in their cut-off. (top-right) Log plot of experimental energy error (ignoring the model error from the UpCCD approximation). (bottom-left) Many-body order parameter for the RG Hamiltonian (see text), again compared to classical models. (bottom-right) Experimental error in estimating the superconducting order parameter vs the target state within the UpCCD approximation (again ignoring model error). Standard deviation error bars estimated by propagating variance (Raw VQE, PS-VQE) or bootstrapping (EV, PS-VD), see App. C 2 for details.

by adding a chemical potential to the single particle energies in Eq. (2);

$$\epsilon_p = p - \mu, \qquad \mu = \frac{1}{2}(N+1) \tag{10}$$

In Fig. 2 (top left), we estimate the prepared states' energy with and without error mitigation techniques (see caption), and compare it to exact diagonalization in the S_0 subspace, also known as double occupied configuration interaction (DOCI), and classical pair-coupled-cluster doubles (pCCD), and BCS solutions. We see that using EV or PS-VD we are able to reproduce the entire energy curve to high accuracy, which neither pCCD nor the non-interacting BCS theory can achieve. The experimental error in the result is the sum of the UpCCD model error and the experimental error. To disambiguate the effects of UpCCD model error, in Fig. 2 (top right) we plot the error between our experimental data and the UpCCD ground state energy. Postselection consistently mitigated around half the error present in the raw ansatz. By contrast, EV demonstrates an average 85-fold and maximum 460-fold error reduction. PS-VD achieves similar performance, with an average 60-fold and maximum 140-fold improvement. The residual error following EV or PS-VD drifts notably with fluctuations between points larger than error bars. We attribute this observation to device drift.

The RG Hamiltonian has a well-known phase transition in the attractive regime ($g \leq 0$) in the thermodynamic limit, which appears in the BCS state at finite N, but is not present in the true ground state due to finite size effects [56–58]. This presents an opportunity for a variational quantum simulation to determine qualitative features of a quantum Hamiltonian beyond non-interacting physics. The traditional order parameter for the BCS state, $\Delta_{BCS} = \frac{1}{N}\sum_j \langle a_{j\uparrow} a_{j\downarrow}\rangle$, is zero on the RG Hamiltonian ground state due to number conservation. However, one can confirm that $\Delta = \frac{1}{N}\sum_{j,\sigma}\sqrt{\langle n_{j\sigma}^2\rangle - \langle n_{j\sigma}\rangle^2}$ satisfies $\Delta = \Delta_{BCS}$ for the BCS ground state of the Hamiltonian, giving a many-body order parameter [57]. In Fig. 2 (bottom left), we plot experimental estimates of Δ across the range of g values considered. In the absence of error mitigation, though the order parameter dips around $g = 0$ the true cusp is not reproduced. Both EV and PS-VD clearly improve over the BCS approximation for $g > 0.5$, with EV particularly able to reproduce the cusp at $g = 0$. The performance of error mitigation is demonstrated by plotting the error in Δ against the noise free UpCCD energy in Fig. 2 (bottom right).

We see all experimental estimates have a slight peak in error at $g = 0$. This can be attributed to Δ being highly sensitive to error at this point ($\frac{\partial\Delta}{\partial\langle n_{j\sigma}\rangle} \to \infty$). Furthermore, the maximally-mixed state has $\Delta = 1$, so decoherence has a larger effect when targeting $\Delta << 1$. This contrasts with the error in the energy (Fig. 2 (top right)), which has a slight dip near $g = 0$. We attributed this to the increased contribution from $\langle X_iX_j + Y_iY_j\rangle$ to the energy when g is far from 0, as Δ is independent of these expectation values. The improvement from EV and PS-VD in estimating the order parameter is slightly less than that in estimating the energy, with a mean (max) 32-fold (56-fold) improvement from EV, and 18-fold (51-fold) improvement from PS-VD. We attribute this to the increased sensitivity of Δ to noise at $g = 0$, and the high performance of the raw results at $g << 0$ (where depolarizing noise has little effect as $\Delta \sim 1$).

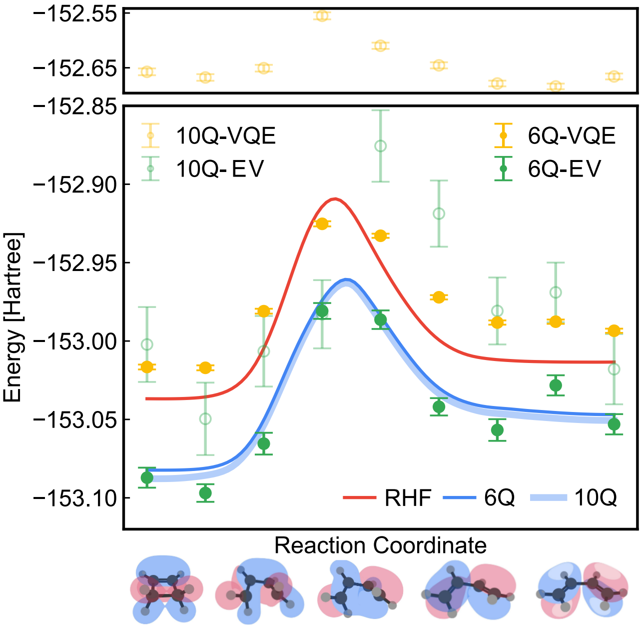


FIG. 3. The conrotatory Cyclobutene ring opening pathway simulated in the seniority zero subspace comparing post-selected VQE and echo verification (EV) on an optimized unitary pair-coupled-cluster ansatz. From left to right the reaction path corresponds to the ring opening reaction. For the ten orbital case the unitary pair-coupled-cluster ansatz (evaluated in simulation) has less than 1.8×10^{-4} energy difference from exact diagonalization in the seniority zero space. The blue curves correspond to the exact diagonalization of the seniority zero active space Hamiltonian spanning 10 orbitals (lighter-broad blue line) and 6 orbitals (darker-narrow blue line). The red curve is the restricted Hartree Fock (RHF) mean-field energy. Green points (darker green for 6 qubits and lighter green for 10 qubits) are the echo verified experimental data while yellow points (darker yellow for 6 qubits and lighter yellow for 10 qubits) are the post-selected VQE energies. The 10 qubit VQE data is plotted on a discontinuous and different scale to preserve the visual scale of the reaction energy along the reaction coordinate.

III. CYCLOBUTENE RING OPENING

We further validated scalable error mitigation protocols by simulating the conrotatory ring opening pathway for cyclobutene in an active space of six orbital and six electrons and ten orbitals and ten electrons corresponding to a six and ten qubit simulation of the Hamiltonian in Eq. (3). The mechanism of this ring opening is described by the Woodward-Hoffmann rules for pericyclic ring openings corresponding to the in-phase combination of the two carbon $2p$ orbitals when brought together to form the four-member carbon ring.

The geometries along the reaction path are determined from a nudged elastic band calculation using density functional theory (B3LYP) to evaluate forces. The final structures use a minimal basis set (STO-3G) to generate the active space Hamiltonians to project into the seniority zero sector. The Woodward-Hoffmann rules are a type of molecular orbital theory and thus we expect this reaction to be qualitatively described within mean-field theory. This is verified numerically for our seniority zero model where the largest CI coefficient has an average value of 0.974(9), for six-orbitals, and 0.973(9), for 10-orbitals, indicating a single-reference system. As such, our unitary pair-coupled-cluster doubles ansatz targets the dynamic correlation corrections to the mean-field.

The average post-selected-VQE absolute error is 0.058 ± 0.006 and 0.395 ± 0.023 Hartree for the six orbital and ten orbital systems, respectively. The average echo-verified absolute error is 0.011 ± 0.005 and 0.064 ± 0.035 Hartree for the six orbital and ten orbital system, respectively, showing a 5.51-fold and 6.12-fold improvement over post-selected-VQE average error. Comparing to the raw VQE data, we find a 55.1-fold and 38.4-fold mean error reduction for the six orbital system and 10-qubit system respectively. While there is notable improvement in energy across the reaction pathway for the 10 orbital system the magnitude of the errors is larger than the 0.037 Hartree energy difference between cyclobutene and 1,3-Butadiene. Furthermore, a visual inspection of Figure 3 indicates high parallelity errors in the 10 orbital system. Given the error bars on echo verification are smaller than the parallelity error (point scatter) we attribute the main source of error to device drift.

IV. OUTLOOK

We have observed the echo verification and virtual distillation error mitigation protocols suppressing errors by 1-2 orders of magnitude on a range of quantum simulation problems using up to 20 superconducting qubits. We now consider the requirements for scaling these experiments to the classical intractable regime.

In Fig. 4 (top) we plot the number of experiments (shots) used in this work to simulate the RG Hamiltonian at $g=-0.9$ (where pCCD does not describe the system well), and compare this to theoretical estimates targeting the same model to within a sampling noise of 0.1 a.u. using the experimental fidelities observed for 10 qubits (fidelities taken from Fig. 4 (bottom right)). The $50\times$ gap between theory and experiment for 10-qubit EV can be attributed mostly to extra circuits used to cancel out a background magnetic field (see App. B 4). The gap for our VD experiment is roughly $3\times$ by comparison. Assuming the ability to freely weight our shot distribution, we estimate that for a 50-qubit experiment (as a proxy lower bound for a beyond-classical quantum computation) using VD or EV, 10^8 or 10^9 shots would be required respectively. This is executable on current hardware in a wall-clock time (see App. D 1) of >1 hour or >10 hours respectively. Including the difference between experiment and theory at 10 qubits raises the cost of EV to 5×10^{10} shots, which would require multiple days to achieve. These numbers do not include the multiplica-

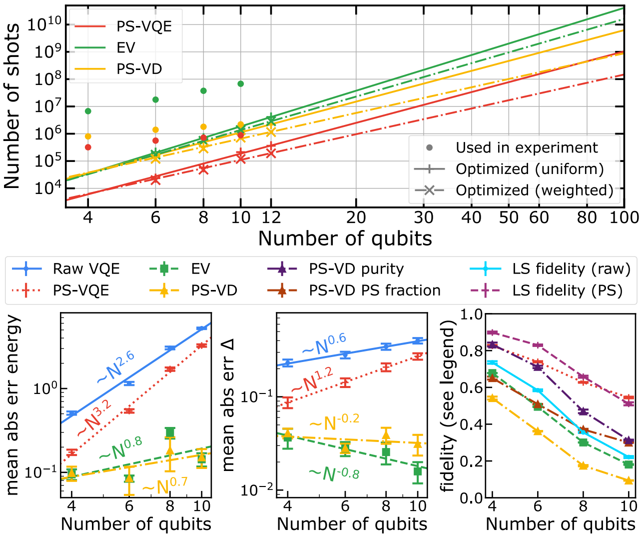


FIG. 4. Scaling the simulation of the RG model to larger qubit counts. (top) Number of shots required for convergence at $g = -0.9$. Dots give numbers chosen for the experiment, crosses and pluses give simulated estimations using two types of term grouping (see App. D 3) using observed fidelities of a 10 qubit experiment. (bottom-left) Experimental energy error (vs the UpCCD ground state), averaged over all points studied of the RG model. Error bars show sample standard deviation, and lines a power-law fit (exponent shown) as a guide to the eye. (bottom-middle) Experimental error in order parameter (vs the UpCCD ground state), averaged over all points studied of the RG model. Error bars and lines same as bottom-left. (bottom-right) Different fidelity metrics for post-selected VQE, EV, VD, and Loschmidt echo (see legend), averaged over all points studied of the RG model.

tive cost of variational optimization (see App. E). Furthermore, the requirements for accurate electronic structure simulations may be lower than the 0.1 a.u. requirement considered here. Methods to pre-optimize variational ansatzes classically, and applications of VQE to problems simpler than electronic structure, may thus be necessary for beyond-classical VQE experiments.

Device coherence presents an additional scaling challenge. To maintain circuit fidelity F over an $O(N)$-depth, fully parallel circuit as N scales from 5 to 50 requires all error rates to drop by and coherence times to increase by roughly a factor 25 (proportional to $O(N^2)$). As any reduction in F incurs an $O(\text{poly}(F^{-1}))$ sampling cost [49, 51], and as F scales exponentially in the error rate, and as $F \sim 10\%$ for PS-VD (Fig. 4 (bottom-right)), we see little room for negotiation on this 25× lower bound. To achieve a 25× decrease in error rate would require a 50-qubit device with XEB fidelities on all two-qubit gates $\leq 3 \times 10^{-4}$. However this analysis precludes ansatzes with depth $O(N^2)$ or higher or significantly larger constant factors (in our case, the circuit depth of the bare VQE is $3N/2$). For instance, successfully implementing a 50-qubit VQE with ansatz depth $3N^2/2$ with EV or VD would require error rates to drop $\sim 1000\times$.

On a more positive note, in Fig. 4 (bottom left), we plot the absolute error in the energy estimates, averaged across all points in our RG model experiment. The energy scales sublinearly after applying EV or VD (a clear asymptotic difference to raw or postselected VQE), which suggests that a 25× decrease in error rate required to keep sampling costs constant may yield significantly higher precision results. A similar gap between EV/VD and VQE/PS-VQE for estimating the order parameter can be observed in Fig. 4 (bottom middle); the discrepancy in absolute scaling can be attributed mostly to the energy scaling as $O(N^2)$, while Δ does not scale with N. This observation runs contrary to the observations in Fig. 3, where shifting from 6 to 10 qubits increased the mean error by a factor 10. Investigating the mean error in estimating Pauli operators (App. F 2) suggests that the true scaling lies somewhere in between these values. If the energy error scales linearly or better with the error rate per qubit (which is expected from simulations in Ref. [51]), and scales less than quadratically in N, our requirement to scale error rates as $O(N^{-2})$ to preserve the circuit fidelity F will yield a drop in absolute energy error as a function of N. Thus, pinning down this scaling of experimental error with system size and error rates is a key area for future work.

Author contributions

T.E.O. calibrated the device and ran the experiments. V.E.E., G.A., and C.G. designed the UpCCD ansatz. V.E.E., G.A., and T.E.O. designed the scheduling of the ansatz onto a 2 × N grid. F.G. and C.G. designed the conjugate model gradient descent algorithm and pre-optimized ansatz parameters for the ansatz. N.C.R. wrote the pair coupled cluster code, pre-optimized ansatz parameters, and performed the classical chemistry calculations for the cyclobutene model. T.E.O., W.J.H., S.P., K.K. and R.B. designed and optimized the error mitigation and EV measurement strategies. O.O. and C.G. developed the BQP completeness proof for the UpCCD ansatz. T.E.O., N.C.R., C.G., F.G., V.E.E. and R.B. wrote the paper. T.E.O., C.G., R.B. and N.C. led and co-ordinated the project. All authors contributed to revising the manuscript and writing the Supplementary Information. All authors contributed to the experimental and theoretical infrastructure to enable the experiment.

Acknowledgments

Some discussion and collaboration on this project occurred while using facilities at the Kavli Institute for Theoretical Physics, supported in part by the National Science Foundation under Grant No. NSF PHY-1748958.

Data availability

Raw and processed experimental data can be found at https://doi.org/10.5281/zenodo.7225821

---

[1] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, J. Basso, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, A. D. T. Barba, S. Demura, A. Dunsworth, D. Eppens, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, L. F. Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, A. G. Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. P. Harrigan, S. D. Harrington, O. Higgott, J. Hilton, M. Hoffmann, S. Hong, T. Huang, A. Huff, W. J. Huggins, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, K. Kechedzhi, J. Kelly, T. Khattar, M. Khezri, M. Kieferová, S. Kim, A. Kitaev, P. V. Klimov, A. R. Klots, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K.-M. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. Lill, W. Liu, A. Locharla, E. Lucero, F. D. Malone, J. Marshall, O. Martin, J. R. McClean, T. Mccourt, M. McEwen, A. Megrant, B. M. Costa, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, E. Mount, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, T. E. O'Brien, A. Opremcak, J. Platt, A. Petukhov, R. Potter, L. Pryadko, C. Quintana, P. Roushan, N. C. Rubin, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, V. Shvarts, J. Skruzny, V. Smelyanskiy, W. C. Smith, G. Sterling, D. Strain, Y. Su, M. Szalay, A. Torres, G. Vidal, B. Villalonga, C. V. Heidweiller, and T. White, Suppressing quantum errors by scaling a surface code logical qubit, ArXiv:2207.06431 (2022).

[2] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating Reaction Mechanisms on Quantum Computers, Proceedings of the National Academy of Sciences **114**, 7555 (2017).

[3] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, Quantum computing enhanced computational catalysis, Phys. Rev. Research **3**, 033055 (2021).

[4] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction, PRX Quantum **2**, 030305 (2021).

[5] J. J. Goings, A. White, J. Lee, C. S. Tautermann, M. Degroote, C. Gidney, T. Shiozaki, R. Babbush, and N. C. Rubin, Reliably assessing the electronic structure of cytochrome p450 on today's classical computers and tomorrow's quantum computers, ArXiv:2202.01244 (2022).

[6] C. Gidney and M. Ekerøa, How to factor 2048 bit rsa integers in 8 hours using 20 million noisy qubits, Quantum **5**, 433 (2021).

[7] E. T. Campbell, Early fault-tolerant simulations of the hubbard model, Quant. Sci. Technol. **7**, 015007 (2021).

[8] D. W. Berry, Y. Su, C. Gyurik, R. King, J. Basso, A. Del Toro Barba, A. Rajput, N. Wiebe, V. Dunjko, and R. Babbush, Quantifying quantum advantage in topological data analysis, ArXiv:2209.13581 (2022).

[9] 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 quantum processor, Nat. Comm. **5**, 4213 (2014).

[10] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The Theory of Variational Hybrid Quantum-Classical Algorithms, New Journal of Physics **18**, 23023 (2016).

[11] S. McArdle, S. Endo, A. Aspuru-Guzik, S. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[12] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015).

[13] 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**, 3007 (2019).

[14] 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 (2017).

[15] V. E. Elfving, M. Millaruelo, J. A. Gámez, and C. Gogolin, Simulating quantum chemistry in the seniority-zero space on qubit-based quantum computers, Phys. Rev. A **103**, 032605 (2021).

[16] 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).

[17] W. J. Huggins, J. McClean, N. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, Efficient and Noise Resilient Measurements for Quantum Chemistry on Near-Term Quantum Computers, npj Quantum Information volume **7** (2021).

[18] J. Cotler and F. Wilczek, Quantum overlapping tomography, Physical Review Letters **124**, 100401 (2020).

[19] 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).

[20] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, The Journal of Chemical Physics **152**, 124114 (2020).

[21] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measure-

ment of pauli operators in the presence of finite sampling error, Quantum **5**, 385 (2021).

[22] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[23] 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, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, Scalable Quantum Simulation of Molecular Energies, Physical Review X **6**, 31007 (2016).

[24] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature **567**, 491 (2019).

[25] 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, Physical Review X **8**, 031022 (2018).

[26] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. Rol, C. Bultink, X. Fu, C. Price, V. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. O'Brien, and L. DiCarlo, Error mitigation by symmetry verification on a variational quantum eigensolver, Physical Review A **100**, 010302 (2019).

[27] F. Arute, K. Arya, R. Babbush, D. Bacon, J. Bardin, R. Barends, S. Boixo, M. Broughton, B. Buckley, D. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, S. Habegger, M. Harrigan, A. Ho, S. Hong, T. Huang, W. Huggins, L. Ioffe, S. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, O. Martin, J. Martinis, J. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, H. Neven, M. Niu, T. O'Brien, E. Ostby, A. Petukhov, H. Putterman, C. Quintana, P. Roushan, N. Rubin, D. Sank, K. Satzinger, V. Smelyanskiy, D. Strain, K. Sung, M. Szalay, T. Takeshita, A. Vainsencher, T. White, N. Wiebe, Z. Jamie Yao, P. Yeh, and A. Zalcman, Hartree-fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020).

[28] S. Stanisic, J. L. Bosse, F. M. Gambetta, R. A. Santos, W. Mruczkiewicz, T. E. O'Brien, E. Ostby, and A. Montanaro, Observing ground-state properties of the fermihubbard model using a scalable algorithm on a quantum computer, ArXiv:2112.02025 (2021).

[29] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala, Scalable error mitigation for noisy quantum circuits produces competitive expectation values, ArXiv:2108.09197 (2021).

[30] E. van den Berg, Z. K. Minev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse pauli-lindblad models on noisy quantum processors, ArXiv:2201.09866 (2022).

[31] M. Motta, G. O. Jones, J. E. Rice, T. P. Gujarati, R. Sakuma, I. Liepuoniute, J. M. Garcia, and Y.-y. Ohnishi, Quantum chemistry simulation of ground- and excited-state properties of the sulfonium cation on a superconducting quantum processor, ArXiv:2208.02414 (2022).

[32] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[33] P. R. Surján, Á. Szabados, P. Jeszenszki, and T. Zoboki, Strongly orthogonal geminals: size-extensive and variational reference states, Journal of Mathematical Chemistry **50**, 534 (2012).

[34] F. Kossoski, Y. Damour, and P.-F. Loos, Hierarchy configuration interaction: Combining seniority number and excitation degree, The journal of physical chemistry letters **13**, 4342 (2022).

[35] K. Gunst, D. Van Neck, P. A. Limacher, and S. De Baerdemacker, The seniority quantum number in tensor network states, SciPost Chemistry **1**, 001 (2021).

[36] K. Boguslawski, P. Tecmer, P. W. Ayers, P. Bultinck, S. De Baerdemacker, and D. Van Neck, Efficient description of strongly correlated electrons with mean-field cost, Phys. Rev. B **89**, 201106 (2014).

[37] P. A. Limacher, P. W. Ayers, P. A. Johnson, S. De Baerdemacker, D. Van Neck, and P. Bultinck, A new mean-field method suitable for strongly correlated electrons: Computationally facile antisymmetric products of nonorthogonal geminals, Journal of chemical theory and computation **9**, 1394 (2013).

[38] J. Dukelsky, J. M. Roman, and G. Sierra, Comment on "polynomial-time simulation of pairing models on a quantum computer", Phys. Rev. Lett. **90**, 249803 (2003).

[39] J. Dukelsky, Integrable richardson-gaudin models in mesoscopic physics, J. Phys. Conf. Ser. **338**, 012023 (2012).

[40] N. Vu and A. E. D. III, Size-extensive seniority-zero energy functionals derived from configuration interaction with double excitations, J. Chem. Phys. **152**, 244103 (2020).

[41] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, ArXiv:2210.00921 (2022).

[42] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).

[43] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, Physical Review Letters **122**, 180501 (2019).

[44] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. O'Brien, Low-cost error mitigation by symmetry verification, Physical Review A **98**, 062339 (2018).

[45] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Physical Review Letters **119**, 180509 (2017).

[46] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[47] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, A. Bengtsson, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, Y.-A. Chen, B. Chiaro, R. Collins, S. J. Cotton, W. Courtney, S. Demura, A. Derk, A. Dunsworth, D. Eppens, T. Eckl, C. Erickson, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, J. A. Gross, S. Habegger, M. P. Harrigan, A. Ho, S. Hong, T. Huang, W. Huggins, L. B. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. N. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, M. Marthaler, O. Martin, J. M. Martinis, A. Marusczyk, S. McArdle, J. R. McClean, T. McCourt, M. McEwen, A. Megrant, C. Mejuto-Zaera, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, H. Neven, M. Newman, M. Y. Niu, T. E. O'Brien, E. Ostby, B. Pató, A. Petukhov, H. Putterman, C. Quintana, J.-M. Reiner, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, D. Strain, K. J. Sung, P. Schmitteckert, M. Szalay, N. M. Tubman, A. Vainsencher, T. White, N. Vogt, Z. J. Yao, P. Yeh, A. Zalcman, and S. Zanker, Observation of separated dynamics of charge and spin in the fermi-hubbard model, ArXiv:2010.07965 (2020).

[48] A. Montanaro and S. Stanisic, Error mitigation by training with fermionic linear optics, arXiv:2102.02120 (2021).

[49] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, Phys. Rev. X **11**, 041036 (2021).

[50] B. Koczor, Exponential error suppression for near-term quantum devices (2021) p. 031057.

[51] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, Error mitigation via verified phase estimation, PRX Quantum **2**, 020317 (2021).

[52] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Physical Review X **8**, 031027 (2018).

[53] M. Huo and Y. Li, Dual-state purification for practical error mitigation, Phys. Rev. A **105**, 022427 (2022).

[54] R. McWeeny, Some recent advances in density matrix theory, Rev. Mod. Phys. **35**, 668 (1963).

[55] P. Ring and P. Schuck, *The nuclear many-body problem* (Springer Science & Business Media, 2004).

[56] J. von Delft, A. Zaikin, D. Golubev, and W. Tichy, Parity-affected superconductivity in ultrasmall metallic grains, Phys. Rev. Lett. **77** (1996).

[57] F. Braun and J. von Delft, Superconductivity in ultrasmall metallic grains, Phys. Rev. B **59**, 9527 (1999).

[58] J. Dukelsky and G. Sierra, The crossover from the bulk to the few-electron limit in ultrasmall metallic grains, Phys. Rev. B **61**, 12302 (1999).

[59] A. Khamoshi, F. A. Evangelista, and G. E. Scuseria, Correlating agp on a quantum computer, Quantum Science and Technology **6**, 014004 (2020).

[60] T. M. Henderson, I. W. Bulik, T. Stein, and G. E. Scuseria, Seniority-based coupled cluster theory, The Journal of chemical physics **141**, 244104 (2014).

[61] T. Stein, T. M. Henderson, and G. E. Scuseria, Seniority zero pair coupled cluster doubles theory, The Journal of chemical physics **140**, 214113 (2014).

[62] J. J. Shepherd, T. M. Henderson, and G. E. Scuseria, Using full configuration interaction quantum monte carlo in a seniority zero space to investigate the correlation energy equivalence of pair coupled cluster doubles and doubly occupied configuration interaction, The Journal of Chemical Physics **144**, 094112 (2016).

[63] S. Polla, G.-L. R. Anselmetti, and T. E. O'Brien, Optimizing the information extracted by a single qubit measurement, ArXiv:2207.0947 (2022).

[64] X. Mi, P. Roushan, C. Quintana, S. Mandrà, J. Marshall, C. Neill, F. Arute, K. Arya, J. Atalaya, R. Babbush, J. C. Bardin, R. Barends, J. Basso, A. Bengtsson, S. Boixo, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. R. Derk, A. Dunsworth, D. Eppens, C. Erickson, E. Farhi, A. G. Fowler, B. Foxen, C. Gidney, M. Giustina, J. A. Gross, M. P. Harrigan, S. D. Harrington, J. Hilton, A. Ho, S. Hong, T. Huang, W. J. Huggins, L. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, J. Kelly, S. Kim, A. Kitaev, P. V. Klimov, A. N. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, E. Lucero, O. Martin, J. R. McClean, T. McCourt, M. McEwen, A. Megrant, K. C. Miao, M. Mohseni, S. Montazeri, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, M. Newman, M. Y. Niu, T. E. O'Brien, A. Opremcak, E. Ostby, B. Pato, A. Petukhov, N. Redd, N. C. Rubin, D. Sank, K. J. Satzinger, V. Shvarts, D. Strain, M. Szalay, M. D. Trevithick, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, I. Aleiner, K. Kechedzhi, V. Smelyanskiy, and Y. Chen, Information scrambling in quantum circuits, Science , eabg5029 (2021).

[65] C. Neill, T. McCourt, X. Mi, Z. Jiang, M. Y. Niu, W. Mruczkiewicz, I. Aleiner, F. Arute, K. Arya, J. Atalaya, R. Babbush, J. C. Bardin, R. Barends, A. Bengtsson, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, J. Campero, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. R. Derk, A. Dunsworth, D. Eppens, C. Erickson, E. Farhi, A. G. Fowler, B. Foxen, C. Gidney, M. Giustina, J. A. Gross, M. P. Harrigan, S. D. Harrington, J. Hilton, A. Ho, S. Hong, T. Huang, W. J. Huggins, S. V. Isakov, M. Jacob-Mitos, E. Jeffrey, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. N. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, E. Lucero, O. Martin, J. R. McClean, M. McEwen, A. Megrant, K. C. Miao, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, M. Newman, T. E. O'Brien, A. Opremcak, E. Ostby, B. Pató, A. Petukhov, C. Quintana, N. Redd, N. C. Rubin, D. Sank, K. J. Satzinger, V. Shvarts, D. Strain, M. Szalay, M. D. Trevithick, B. Villalonga, T. C. White, Z. Yao, P. Yeh, A. Zalcman, H. Neven, S. Boixo, L. B.

Ioffe, P. Roushan, Y. Chen, and V. Smelyanskiy, Accurately computing the electronic properties of a quantum ring, Nature **594**, 508 (2021).

[66] Quantum AI team and collaborators, Recirq (2020).

[67] N. C. Rubin, R. Babbush, and J. McClean, Application of fermionic marginal constraints to hybrid quantum algorithms, New J. Phys. **20**, 053020 (2018).

[68] 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).

[69] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards **49**, 409 (1952).

[70] J. W. Daniel, Convergence of the conjugate gradient method with computationally convenient modifications, Numerische Mathematik **10**, 125 (1967).

[71] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, The Computer Journal **7**, 149 (1964), https://academic.oup.com/comjnl/article-pdf/7/2/149/959725/070149.pdf.

Appendix A: Calibration of the processor

All experiments were implemented on a subgrid of a 25-qubit superconducting processor with the Sycamore architecture. For all methods other than virtual distillation, a $2 \times N/2$ qubit grid was calibrated to within 0.008 XEB fidelity [32] and 0.008 speckle purity [32]. For virtual distillation, a $4 \times N/2$ qubit grid was calibrated to within 0.01 XEB fidelity and 0.01 speckle purity.

We were further required to calibrate the single-qubit Z-phases accumulated during a CZ gate. This is a well-documented issue [47, 65], but is complicated in our case by the addition of microwave gates. These are observed to bleed into the CZ gate, which made standard Floquet calibration techniques inaccurate. To solve this issue, we calibrate CZ gates *in-situ*. The Givens-SWAP gate was altered by, after each CZ between qubits i and j, inserting virtual rotations $\exp(iZ_i\beta_i^{(j)}), \exp(iZ_j\beta_j^{(i)})$ on qubit i and j respectively. The phases $\beta_i^{(j)}$ were calibrated by running two experiments in series. Firstly, a single $GS(0) = \text{SWAP}$ gate was implemented between qubits i and j (with virtual gates inserted); the qubits were prepared in the state $|0+\rangle$ measured in the ZX or ZY basis, or prepared in the state $|+0\rangle$ and measured in the XZ or YZ basis. Sweeping $\beta_i^{(j)}$ and $\beta_j^{(i)}$ gave four datasets that could be fitted to extract optimal phase offsets. The resulting gate was then benchmarked by estimating $\langle XI\rangle$ and $\langle YI\rangle$ on the state $[GS(0)]^{2k}|0+\rangle$ and $\langle IX\rangle$ and $\langle IY\rangle$ on $[GS(0)]^{2k}|+0\rangle$, and fitting this to an oscillatory decay curve. Under this benchmark, the initial calibration typically reduced the accumulated phase per CZ to less than 30 milliradians. This benchmark was further used to calibrate, by sweeping $\beta_i + \beta_j$ on pairs i, j that are being acted on by the same GS gate to remove the remaining oscillations. We find in practice that a cubic fit to 11 datasets is a robust way to perform a final estimate of $\beta_i + \beta_j$, with the residual phase less than 5 milliradians when calibration was successful. If the estimated fidelity of the resulting GS gate underperformed ($> 1.5\%$ error per CZ gate), qubit or coupler frequencies were reoptimized before recalibrating. Calibration was performed in parallel on sets of CZ gates that were run in parallel during an experiment, to mimic the local environment and compensate for 2-qubit gate crosstalk.

Appendix B: Further details of the UpCCD ansatz

1. BQP-completeness of nearest neighbor Givens-swap circuits

Here we substantiate the claim that the UpCCD circuits realized on hardware in this work are in general not efficiently classically simulable. We do so by constructing a universal quantum gate set on a reduced Hilbert space (dual-rail encoding) with an $O(1)$ depth overhead. This construction shows that any nearest neighbor depth-$O(N)$ circuit on a line of qubits can be mapped to a depth-$O(N)$ UpCCD ansatz (and circuits with arbitrary connectivity to a depth-$O(N^2)$ UpCCD ansatz), when allowing for the omission of gates (as the identity is not a GS gate). For this to hold it is pivotal that the GS gate family includes the SWAP gate and is thus not a matchgate.

To demonstrate a universal gate set we use a dual-rail encoding of one logical qubit into two physical qubits (onto which the GS gates will act). We use tilde to denote logical states and operations and set

$$|\tilde{0}\rangle := |01\rangle \tag{B1}$$

$$|\tilde{1}\rangle := |10\rangle. \tag{B2}$$

It is then straightforward to verify by direct computation that a GS gate acting on two physical qubits belonging to the same logical qubit can be used to realize the following logical Hadamard, Pauli, and Pauli rotation gates:

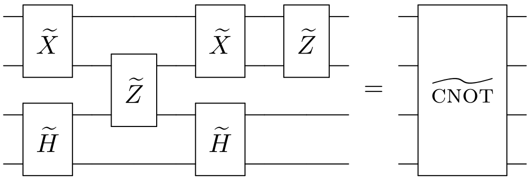
$$GS(\frac{\pi}{4}) = \widetilde{H} \tag{B3}$$

$$GS(0) = \widetilde{X} \tag{B4}$$

$$GS(\frac{\pi}{2}) = \widetilde{Z} \tag{B5}$$

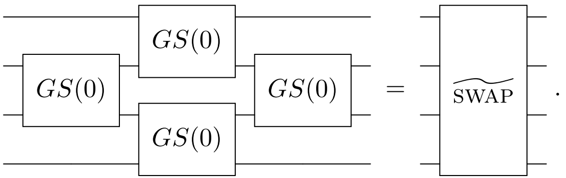
$$GS(0) \cdot GS(\theta) = \widetilde{RY}(\theta) \tag{B6}$$

Logical two qubit entangling gates can be realized by acting with GS gates on qubits belonging to two different logical qubits. The $\widetilde{\text{CNOT}}$ gate can for instance be made by means of



(B7)

and since $GS(0) = \text{SWAP}$ a $\widetilde{\text{SWAP}}$ is obtainable by means of the planar circuit



(B8)

Since CNOT, SWAP, and the above single qubit gates are universal for quantum computation, any circuit from a family recognizing a language in BQP can be represented as a planar GS gate circuit on twice as many qubits and with an at most polynomially larger depth.

2. Gate decompositions

To implement the $GS(\theta)$ gate on superconducting hardware requires us to decompose it into native gates. In our case this is arbitrary single-qubit rotations and two-qubit number-conserving excitation gates [47]. To minimize calibration overhead, we limit ourselves to a fixed two-qubit gate; in all experiments performed this was a controlled-Z gate. We can write an arbitrary single-qubit rotation in phased-XZ form [32]

$$R(\alpha_x, \alpha_a, \alpha_z) = e^{i(\alpha_z+\alpha_a)Z} e^{i\alpha_x X} e^{-i\alpha_a Z} \tag{B9}$$

A $GS(\theta)$ gate can be executed at arbitrary θ on qubits i and j using a combination of 3 CZ gates and single-qubit rotations

$$\begin{aligned} GS_{i,j}(\theta) =& R_i(\tfrac{\pi}{4}, \tfrac{\pi}{4}, 0) \cdot CZ_{i,j} \\ & \cdot R_i(\tfrac{\pi}{4}, \tfrac{-\pi}{4}, 0) \cdot R_j(\tfrac{\pi}{4} + \tfrac{\theta}{2}, \tfrac{-\pi}{4}, \tfrac{\pi}{2}) \cdot CZ_{i,j} \\ & \cdot R_i(\tfrac{\pi}{4}, \tfrac{\pi}{4}, 0) \cdot R_j(\tfrac{\pi}{4} + \tfrac{\theta}{2}, \tfrac{-\pi}{4}, \tfrac{\pi}{2}) \cdot CZ_{i,j} \\ & \cdot R_i(\tfrac{\pi}{4}, \tfrac{-\pi}{4}, 0). \end{aligned} \tag{B10}$$

Alternatively, if one defines a $\sqrt{\text{SWAP}}$ gate to be

$$\sqrt{\text{SWAP}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{e^{-i\frac{\pi}{4}}}{\sqrt{2}} & \frac{e^{i\frac{\pi}{4}}}{\sqrt{2}} & 0 \\ 0 & \frac{e^{i\frac{\pi}{4}}}{\sqrt{2}} & \frac{e^{-i\frac{\pi}{4}}}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{B11}$$

we can decompose a $GS(\theta)$ gate into two $\sqrt{\text{SWAP}}$ gates and single-qubit Z rotations

$$GS_{i,j}(\theta) = \sqrt{\text{SWAP}}_{i,j} e^{i\frac{\theta}{2}Z_i} e^{-i\frac{\theta}{2}Z_j} \sqrt{\text{SWAP}}_{i,j}. \tag{B12}$$

This is similar to the decomposition of a Givens rotation gate into two $\sqrt{I\text{SWAP}}$ and z-rotation gates [27]: the $\sqrt{\text{SWAP}}$ and $\sqrt{I\text{SWAP}}$ gates differ by a $e^{i\frac{\pi}{4}ZZ}$ rotation, which doubles to yield the ZZ term which separates the Givens and GS rotation (up to a redefinition of angle and single-qubit Z rotations).

It is further possible to decompose a $GS(\theta)$ gate into three $\sqrt{I\text{SWAP}}$ gates and arbitrary single-qubit rotations, though the calculation is more involved. We start from a decomposition of a SWAP ($= GS(0)$) gate into three $\sqrt{I\text{SWAP}}$ gates

$$\begin{aligned} \text{SWAP} &= \left(e^{-i\frac{\pi}{4}(YI+IY)}\sqrt{\text{iSWAP}}e^{i\frac{\pi}{4}(YI+IY)}\right)\sqrt{\text{iSWAP}}\left(e^{i\frac{\pi}{4}(XI+IX)}\sqrt{\text{iSWAP}}e^{-i\frac{\pi}{4}(XI+IX)}\right) && \text{(B13)} \\ &= \left(\exp\left[i\tfrac{\pi}{8}(YY+ZZ)\right]\right)\exp\left[i\tfrac{\pi}{8}(XX+YY)\right]\left(\exp\left[i\tfrac{\pi}{8}(XX+ZZ)\right]\right). && \text{(B14)} \end{aligned}$$

If we perform a basis rotation on the bracketed terms by $\exp(i\alpha Z)$ on the appropriate qubit, we can generate a gate of the form

$$G(\alpha) = \exp\left[i\tfrac{\pi}{8}(\cos(\alpha)YY + \sin(\alpha)XY)\right]\exp\left[i\tfrac{\pi}{8}(XX+YY+2ZZ)\right]\exp\left[i\tfrac{\pi}{8}(\cos(\alpha)XX - \sin(\alpha)XY\right]. \tag{B15}$$

This function can be expanded to give

$$G(\alpha) = g_1(\alpha)II + g_2(\alpha)(XX+YY) + g_3(\alpha)(IZ-ZI) + g_4(\alpha)ZZ, \tag{B16}$$

with $g_i(\alpha)$ the following complex functions

$$
\begin{aligned}
g_1(\alpha) &= \frac{1}{\sqrt{2}}\Big\{\cos^2(\tfrac{\pi}{8})(\cos^2(\tfrac{\pi}{8})+i\sin^2(\tfrac{\pi}{8}))+i\cos(\alpha)\cos(\tfrac{\pi}{8})\sin(\tfrac{\pi}{8})\frac{(1+i)}{\sqrt{2}} \\
&\quad +\cos^2(\alpha)\sin^2(\tfrac{\pi}{8})(i\cos^2(\tfrac{\pi}{8})+\sin^2(\tfrac{\pi}{8}))+\sin^2(\alpha)\sin^2(\tfrac{\pi}{8})(\cos^2(\tfrac{\pi}{8})+i\sin^2(\tfrac{\pi}{8}))\Big\} \qquad \text{(B17)} \\
g_2(\alpha) &= \frac{1+i}{4\sqrt{2}}(1+\cos(\alpha)) \qquad \text{(B18)} \\
g_3(\alpha) &= \frac{1}{\sqrt{2}}\Big\{\sin(\alpha)\sin(\tfrac{\pi}{8})\cos(\tfrac{\pi}{8})\frac{(1+i)}{\sqrt{2}}+i\sin(\alpha)\cos(\alpha)\sin^2(\tfrac{\pi}{8})(\cos^2(\tfrac{\pi}{8})+i\sin^2(\tfrac{\pi}{8})) \\
&\quad -i\cos(\alpha)\sin(\alpha)\sin^2(\tfrac{\pi}{8})(i\cos^2(\tfrac{\pi}{8})+\sin^2(\tfrac{\pi}{8}))\Big\} \qquad \text{(B19)} \\
g_4(\alpha) &= \frac{1}{\sqrt{2}}\Big\{\cos^2(\tfrac{\pi}{8})(i\cos^2(\tfrac{\pi}{8})+\sin^2(\tfrac{\pi}{8}))-i\cos(\alpha)\cos(\tfrac{\pi}{8})\sin(\tfrac{\pi}{8})\frac{(1+i)}{\sqrt{2}} \\
&\quad +\cos^2(\alpha)\sin^2(\tfrac{\pi}{8})(\cos^2(\tfrac{\pi}{8})+i\sin^2(\tfrac{\pi}{8}))+\sin^2(\alpha)\sin^2(\tfrac{\pi}{8})(i\cos^2(\tfrac{\pi}{8})+\sin^2(\tfrac{\pi}{8}))\Big\}\,. \qquad \text{(B20)}
\end{aligned}
$$

This is of the correct form for our desired gate up to single-qubit Z rotations as long as the phase on $\langle 00|G(\alpha)|00\rangle$, $\langle 11|G(\alpha)|11\rangle$, $\langle 01|G(\alpha)|10\rangle$ and $\langle 10|G(\alpha)|01\rangle$ are equal. One can confirm that all have a phase of $e^{i\frac{\pi}{4}}$ for any angle of α. There remains a residual phase on the two on-diagonal elements of $G(\alpha)$,

$$
\begin{aligned}
\langle 01|G(\alpha)|01\rangle &= g_1(\alpha)-g_4(\alpha)+2g_3(\alpha) \\
&= A(\alpha)e^{i\phi(\alpha)}e^{i\pi/4} \qquad \text{(B21)} \\
\langle 10|G(\alpha)|10\rangle &= g_1(\alpha)-g_4(\alpha)-2g_3(\alpha) \\
&= -A(\alpha)e^{-i\phi(\alpha)}e^{i\pi/4}, \qquad \text{(B22)}
\end{aligned}
$$

which can be removed by shifting

$$G(\alpha)\rightarrow e^{i\phi/4(IZ-ZI)}G(\alpha)e^{i\phi/4(IZ-ZI)}. \qquad \text{(B23)}$$

The precise value of ϕ here is

$$
\phi = -\arctan\Bigg\{ \frac{\sqrt{2}\left(\cos^2(\tfrac{\pi}{8})-\sin^2(\tfrac{\pi}{8})\cos(2\alpha)-\tfrac{1}{\sqrt{2}}\cos(\alpha)\right)}{\sin(\alpha)+\sin(2\alpha)\sin^2(\tfrac{\pi}{8})}\Bigg\}. \qquad \text{(B24)}
$$

We notice that our formula for $g_2(\alpha)$ only takes positive values. To get the full range of $GS(\theta)$, one can finally send $G(\alpha)\rightarrow e^{i\pi/2ZI}e^{i\pi/2IZ}G(\alpha)e^{i\pi/2ZI}e^{i\pi/2IZ}$ (again at no extra cost), and solve for $\sin(\theta)=2g_2(\alpha)$.

3. Scheduling details

A key advance in this work was the development of a mapping of our UpCCD ansatz to a 2D grid with local connectivity, such that a) the entire ansatz could be implemented in depth $N/2$ GS gates, and b) all $X_iX_j+Y_iY_j$ terms could be estimated using only N unique mappings. In this section, we explain this mapping in more detail and prove both a) and b) true.

Let us first expand the discussion of the implementation of the UpCCD ansatz. The standard UpCCD ansatz takes the form

$$U(\theta)=\exp\Big\{\sum_{p\in\text{unocc},q\in\text{occ}}\theta_{pq}a^{\dagger}_{p\alpha}a^{\dagger}_{p\beta}a_{q\alpha}a_{q\beta}-\text{h.c.}\Big\}, \qquad \text{(B25)}$$

which when mapped to qubits in the S_0 approximation, becomes

$$U(\theta)=\exp\Big\{\sum_{p\in\text{unocc},q\in\text{occ}}\theta_{pq}X_pY_q-\text{h.c.}\Big\}. \qquad \text{(B26)}$$

This in turn can be Trotterized to a product of coherent pair excitations $\exp(\theta_{pq}X_pY_q-\text{h.c.})$. As mentioned in the main text, the effect of a single GS gate (Eq. (7)) in the fermionic picture is to implement a single coherent pair excitation between the spatial orbitals assigned to qubits i and j, and then to SWAP the orbitals. This means a given orbital is not assigned to a fixed qubit throughout the experiment. For our implementation of the UpCCD ansatz (see e.g. Fig. 1, bottom) GS gates are applied in layers: between qubits $i,j=2n,2n+1$ during an even-numbered layer, and between qubits $i,j=2n+1,(2n+2)\%N$ during an odd-numbered layer (for $n=0,\ldots,N/2-1$). We claim that, at half-filling, any initial assignment of occupied orbitals to odd-numbered qubits and unoccupied orbitals to even-numbered qubits will cause $N/2$ such layers to implement a Trotterized form of Eq. (B26). To see this, note that during an even-numbered layer, orbitals sitting on even-numbered (odd-numbered) qubits shift to the right (left) around the ring of qubits, and vice-versa during an odd-numbered layer. This in turn implies that the empty orbitals, that are initially assigned to an even-numbered qubit, will only ever move to the right around this ring as the ansatz proceeds, as they will be assigned to an odd-numbered qubit on odd-numbered rounds. Likewise, the filled orbitals will only ever move to the left around this ring. As an occupied orbital must cross (and thus interact with) every unoccupied orbital before it encounters the same one twice, this implies that the first $N/2$ layers of our UpCCD ansatz will yield precisely the $(N/2)^2$ coherent pair excitations between each unoccupied and each occupied orbital, as required. (This is demonstrated for 10 qubits in Fig. 5[top].)

Let us now consider the measurement of non-local $X_iX_j + Y_iY_j$ terms as performed in this work. As mentioned in the main text, these are diagonalized on a pair of orbitals i, j via a $GS(\pi/4)$ rotation, which necessitates the orbitals be on neighbouring qubits. The $GS(\pi/4)$ rotation maps the operators $Z_i, Z_j \rightarrow D^+_{ij}, D^-_{ij}$, where we define

$$D^{\pm}_{ij} = \frac{1}{2}\Big[Z_i + Z_j \pm \big(X_iX_j + Y_iY_j\big)\Big] \qquad \text{(B27)}$$

As we implement our ansatz on a $2 \times N/2$ grid, qubit i is not only connected to qubit $(i \pm 1)\%N$; the cross-links in the grid connect qubit i and qubit $N - i - 1$. Moreover, note that our ansatz remains unchanged if we perform the cyclic permutation $i \rightarrow (i + k)\%N$; that is, we shift our orbital assignments, and all ansatz gates and parameters, around the ring of qubits. (Note that this technically SWAPs the definition of even and odd layers when k is odd: after this permutation the first layer of GS gates will be between qubits $i, j = 2n + 1, (2n + 2)\%N$.) Following this permutation, cross-links will connect the orbital that was on qubit i to that which was on qubit $[N - (i + 2k) - 1]\%N$. One can confirm that by running over $k = 0, \ldots, N/2 - 1$ we find $N/2$ cyclic permutations, such that each occupied orbital is coupled to each unoccupied orbital by a cross-link for exactly one permutation. With just the cyclic permutation operation and no additional SWAP gates, this gives $N/2$ unique circuits that allow for measurement of all $D^{\pm}_{ij}$ where i is occupied and j unoccupied. To obtain the circuits that couple occupied orbitals to occupied orbitals (and unoccupied to unoccupied), we require an additional layer of SWAP gates. (This is unavoidable given our initial ansatz ordering: all occupied qubits are connected by an even number of couplings, so a further direct coupling would yield an odd-order cycle, which does not exist on a square lattice.) After each permutation k above, we perform SWAPs between qubits $l + (k\%2), l + (k\%2) + 1$ for $0 \leq l < N/4$. One can confirm that this yields $N/2$ circuits such that a coupling between any pair of occupied orbitals can be achieved in one such circuit. (In this second set of circuits, some qubits are not coupled, and were not used to estimate expectation values in this experiment.) This can be confirmed by a visual inspection of Fig. 5[bottom]. Code that implements this scheduling has also been uploaded to ReCirq [66].

4. Scheduling of EV circuits

In this section we outline the additional experimental details required to implement EV in this work. We

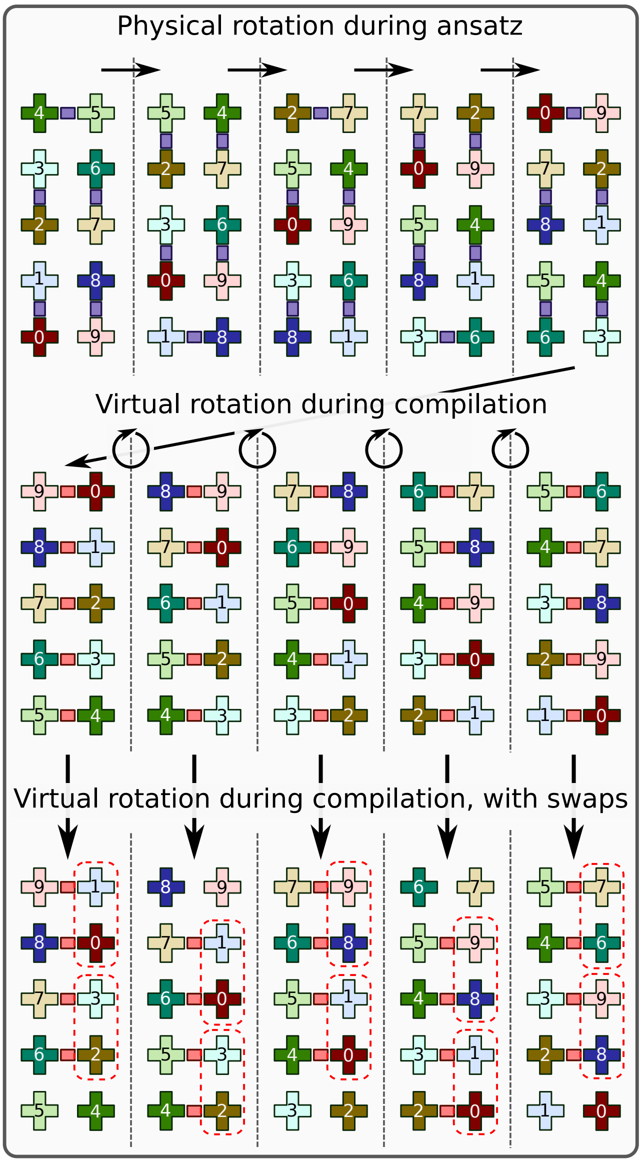


FIG. 5. Detail of the scheduling of pair excitation interactions during the UpCCD ansatz, and measurement scheduling. (top) The physical rotations during the execution of SWAP layers of the UpCCD ansatz, showing that all occupied (even index) and all unoccupied (odd index) orbitals are coupled (purple squares) at some point during the ansatz. (middle) After executing the UpCCD ansatz we have not used the cross-couplers (red squares), allowing us to virtually rotate the entire grid during compilation for the purposes of measurement. This virtual rotation allows all occupied and unoccupied orbitals to be coupled at some step, without any increase in circuit depth. (bottom) The virtual rotation cannot however, bring two occupied or two unoccupied orbitals to nearest-neighbour qubits so that they may be coupled. To achieve this, we require an extra round of SWAPs (red dashed boxes). One can confirm that across the middle and bottom layers, all pairs of qubits are coupled in at least one configuration.

implemented control-free EV for $O = Z_i$, or Z_iZ_j, or D^+_{ij}(Eq. (B27)), using a vacuum reference state $|\phi\rangle = |00\ldots\rangle$. We prepare the superposition $\frac{1}{\sqrt{2}}(|\psi\rangle + |\phi\rangle)$ by acting the UpCCD ansatz circuit (Eq. (4)) on the cat state $\frac{1}{\sqrt{2}}(|0000\ldots\rangle + |0101\ldots\rangle)$ (see Fig. 1[middle, 'Echo verification'] and Fig. 1[bottom]). Then, all operators O were implemented by compiling them to a virtual Z rotation on a single qubit. Finally, to measure $\langle\Psi|\phi\rangle\langle\psi|\Psi\rangle$, we inverted the UpCCD circuit and cat state preparation. This maps the desired matrix element $|\phi\rangle\langle\psi| \rightarrow |0\rangle\langle 1| \otimes |00\ldots\rangle\langle 00\ldots|$, allowing its measurement via single-qubit rotation on a single 'measurement' qubit, and readout of all qubits in the computational basis. (Reading out all qubits is essential, as we record an estimate of 0 for the measurement of $\langle\Psi|\phi\rangle\langle\psi|\Psi\rangle$ unless all qubits other than the measurement qubit read out 0.)

Our implementation of O as a virtual Z rotation allows us to replace $O \rightarrow O^\alpha = \cos(\alpha) + i\sin(\alpha)O$ to remove our susceptibility to a uniform background magnetic field $e^{ih\sum_j Z_j}$. Such a field transforms Eq. (8) to $\frac{1}{2}\langle\psi|O^\alpha|\psi\rangle e^{i(\phi+hN/2)}$; fitting this to three points of α allows us to simultaneously estimate h, the fidelity F, and $\langle O\rangle$. (This is preferable to independent calibration as h fluctuates with a $1/f$ spectrum.)

Some further experimental optimization was made for the EV circuit that was not available for the VQE or VD circuits. Many of the gates in the final EV circuit [Fig. 1, bottom] cancel to the identity, as the second half of the circuit is the inversion of the first half. We identify and prune these to increase the overall circuit fidelity, and insert echo pulses into the resulting empty space. We further compile an echo pulse for the entire second half of the EV circuit [this can be done as $XX \cdot GS(\theta) \cdot XX = GS(\theta)$]. To unbias readout, we measure the single measurement qubit in the $\pm X$ and $\pm Y$ bases. This, combined with the additional circuits to remove a background magnetic field, raises the total number of circuits to $12N^2$ $(6N^2 + 6N)$ to estimate the expectation value of our chemistry (RG) Hamiltonian. Shots were distributed across these circuits following the term weight in the Hamiltonian with some additional restrictions imposed by classical readout hardware (see App. D 3).

Appendix C: Data processing

1. Optimal linear combinations of non-independent expectation values

Once we have used our variational ansatz to prepare an approximation $\rho(\boldsymbol{\theta}) \sim |\psi(\boldsymbol{\theta})\rangle\langle\psi(\boldsymbol{\theta})|$ to the ground state of our target problem, it remains to measure the quantum device to estimate the energy (or other properties of the state). As our devices are heavily coherence limited, rather than attempting to perform this estimation in a single shot, we write our Hamitonian as a sum of simpler terms

$$H = \sum_i c_i Q_i, \tag{C1}$$

and estimate the expectation value of each such term

$$E = \mathrm{Trace}[H\rho] = \sum_i c_i \mathrm{Trace}[Q_i\rho]. \tag{C2}$$

(Note that we are not placing any restrictions on Q_i at this point.) The method in which we estimate $\mathrm{Trace}[Q_i\rho]$ will depend on which error mitigation methods are being implemented. However, all schemes will return a set of estimates of $\mathrm{Trace}[Q_i\rho]$ with a covariance matrix

$$\Sigma_{i,j} = \mathrm{Covar}\Big[\mathrm{Trace}[Q_i\rho], \mathrm{Trace}[Q_j\rho]\Big]. \tag{C3}$$

In this experiment, it turns out that our choice of $\{Q_i\}$ will not be linearly independent. The reason for this is post-selection: we desire our choice of $\{Q_i\}$ to allow us to measure $S_z = \sum_i Z_i$ for each experiment. To achieve this, we measure the operators Z_j, Z_jZ_k, and $D^\pm_{jk}$ (Eq. (B27)), but $D^+_{jk} + D^-_{jk} = Z_j + Z_k$. This leaves us with a degree of freedom in our choice of c_i that we may optimize upon, once a dataset is taken.

In order to choose c_i, we perform a constrained Lagrangian minimization. Our target cost function is the variance on the estimate in Eq. (C2)

$$\mathrm{Var}(E) = \sum_{i,j} c_i \Sigma_{i,j} c_j, \tag{C4}$$

subject to Eq. (C1) as a constraint. Let us fix some linearly independent basis of operators $\{P_j\}$ (e.g. Pauli operators), and we can write $H = \sum_j h_j P_j$, and $Q_i = \sum_j q_{i,j} P_j$. (As P_j is a basis, we have no freedom in our choice of h_j or $q_{i,j}$.) Our constraints then take the form

$$\sum_i c_i q_{i,j} = h_j. \tag{C5}$$

This can be written as a Lagrange multiplier, yielding a Lagrangian

$$\mathcal{L} = \sum_{i,j} c_i \Sigma_{i,j} c_j - \sum_j \Big(\sum_i c_i q_{i,j} - h_j\Big)\lambda_j. \tag{C6}$$

Differentiating with respect to c_i and setting equal to 0 yields (using the fact that $\Sigma_{i,j}$ is a positive matrix)

$$2\sum_j \Sigma_{i,j} c_j - \sum_j q_{i,j}\lambda_j = 0 \rightarrow c_j. \tag{C7}$$

Recognising this as a vector equation, as the matrix Σ is invertible we have

$$\mathbf{c} = \frac{1}{2}\Sigma^{-1} q \boldsymbol{\lambda}. \tag{C8}$$

Here, $\boldsymbol{\lambda}, \mathbf{c}$ are vectors containing the λ_j and c_j components, and q and Σ are matrices containing the $q_{i,j}$ and $\Sigma_{i,j}$ components respectively. Substituting this into our Lagrangian yields

$$\mathcal{L} = -\frac{1}{4}\boldsymbol{\lambda}^T q^T \Sigma^{-1} q \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \mathbf{h}. \tag{C9}$$

Then, differentiating with respect to λ (and using the fact that $q^T \Sigma^{-1} q$ is Hermitian), we have

$$q^T \Sigma^{-1} q \boldsymbol{\lambda} = 2\mathbf{h} \rightarrow \boldsymbol{\lambda} = 2(q^T \Sigma^{-1} q)^* \mathbf{h} \tag{C10}$$

$$\rightarrow \mathbf{c} = \Sigma^{-1} q (q^T \Sigma^{-1} q)^* \mathbf{h}, \tag{C11}$$

where here, $*$ denotes the Moore-Penrose pseudoinverse of a matrix.

In practice, though we find that this produces low-variance estimates, it is unstable to uncertainty in our estimate of Σ. As such, we set $\Sigma = I$ for the purposes of determining $\mathbf{c}$ (which corresponds to assuming that all uncertainties are equal and all covariances are 0). This yields

$$\mathbf{c} = q(q^T q)^* \mathbf{h}. \tag{C12}$$

2. Error propagation and bootstrapping

The exact form of the error for raw and post-selected VQE is well-known. The covariance between estimates of the expectation value of two reflection operators P_i and P_j, estimated simultaneously from M repeated preparations and measurements on a target state, is given by

$$\mathrm{Covar}[\langle P_i \rangle \langle P_j \rangle] = \frac{\langle P_i P_j \rangle - \langle P_i \rangle \langle P_j \rangle}{M}. \tag{C13}$$

The resulting number can be substituted into the propagation of variance formula described in App. C 1 above to get an estimate of the variance in energy, whilst errors in order parameters can be obtained by propagation of variance through ($\Delta = \frac{1}{N}\sum_{j,\sigma}\sqrt{\langle n_{j\sigma}^2 \rangle - \langle n_{j\sigma} \rangle^2}$)

$$\frac{\partial \Delta}{\partial \langle n_{j\sigma} \rangle} = \frac{1}{2N} \frac{1 - 2\langle n_{j\sigma} \rangle}{\sqrt{\langle n_{j\sigma} \rangle - \langle n_{j\sigma} \rangle^2}} \tag{C14}$$

$$\mathrm{Var}[\Delta] = \sum_{j\sigma} \frac{1}{2N} \frac{(1 - 2\langle n_{j\sigma} \rangle)^2}{\langle n_{j\sigma} \rangle - \langle n_{j\sigma} \rangle^2} \mathrm{Var}[\langle n_{j\sigma} \rangle]. \tag{C15}$$

We note that this diverges when $\langle n_{j,\sigma} \rangle \rightarrow 0, 1$, which is what happens at $g = 0$ when $\Delta \rightarrow 0$. We suggest that this explains the peak in the observed experimental error in the order parameter around $g = 0$ somewhat.

The experimental covariances (Eq. (C13) for raw and postselected VQE when $i \neq j$ were corrupted during data taking; these terms were set to 0 when generating error bars for Fig. 3 and Fig. 2. As said error bars are negligible due to the large number of samples used in this experiment, more complex recovery procedures will not noticeably change the figures.

For echo verification and virtual distillation, the formulas for variance are more complicated (the form derived in Ref. [63] is not appropriate here due to our fitting to remove a background magnetic field). Instead, error bars were determined by bootstrapping the raw data (resampling with replacement and taking the sample standard deviation) over 100 and 25 samples respectively. This was made complicated due to data loss during the the EV experiment. Arrays were stored of the verified expectation value of X and Y on the measurement qubit; this is insufficient to recover the shot distribution, as the EV measurement can return three values; $+1$, -1, and 0 [51]. Counts M_+, M_-, and M_0 of these values were approximated from these expectation valus using the estimated EV fidelity F_{EV} and the fact that in the absence of error

$$\frac{M_0}{M} = 1 - |\langle U \rangle|^2, \tag{C16}$$

where U is the operator to be estimated via EV, and $M = M_+ + M + - + M_0$. Assuming the fraction $1 - F_{\mathrm{EV}}$ fails verification, we have $M_0 = M(1 - F_{\mathrm{EV}}|\langle U \rangle|^2)$, and M_+ and M_- may be then distributed so that $(M_+ - M_-)/M$ is the observed expectation value on the measurement qubit. (This can result in $M_+ < 0$ or $M_- < 0$, in which case we reduced $M_0 \rightarrow M_0 - 2\min(M_+, M_-)$, $M_+, M_- \rightarrow M_+ + \min(M_+, M_-), M_- + \min(M_+, M_-)$.)

Appendix D: Quantum run time estimation

In this section, we estimate the cost of running our quantum experiments in terms of number of experiments (shots) and the wall-clock time. This is done in terms of established theoretical cost estimates [12, 67], which allows us to compare between this and what was implemented in the actual experiment in Fig. 4 (top). This is further necessary for the tuning of hyperparameters of the variational optimizers we will present in App. E.

1. Wall-clock model

The cost of running a set of experiments on real-world hardware in terms of the actual time spent or *wall-clock time* is not a linear function of the number of samples used. We can estimate the wall-clock time as a function of three parameters. First, the number a of calls to the device from the computed executing the Cirq code. Second, the number b of distinct circuits the device needs to execute. Third, the total number of shots c spend over all calls and distinct circuits. We found empirically that for the device used the time for a call to the device from Cirq is about $1s$. In each call, a batch of distinct circuits can be sent to the device for execution but re-programming the device to execute a different circuit takes $0.042s$. When estimating energy expectation values or order parameters, all circuits that need to be executed are known in advance and thus can be sent in a single batch. However, during variational optimization, depending on the optimizer used, at least one call to the device per epoch is needed, because the quantum circuits to be executed next depend on the step taken by the optimizer, which in turn depends on the measurement results of the first batch. This needs to be taken into account when comparing the performance of different optimizers. The time per shot was found to be $1 * 10^{-5}$, essentially independently of the circuit depth (probably because it is dominated by readout, reset, and time needed by the control electronics). The number of shots c of a circuit can currently only be set for a batch as a whole, which in practice limits our ability to distribute shots in a way that would minimize the variance of the resulting estimator. The total wall-clock time this takes the form

$$t_{\text{wall}} = a * 1s + b * 0.042s + c * 5 * 10^{-5}s \tag{D1}$$

2. Hamiltonian decomposition schemes

In this section we define the different options chosen to decompose a Hamiltonian into terms Q_i for measurement. We are free to group the measurement of all Q_i terms together, as long as they commute and an appropriate diagonalization circuit is found. For our numerics, we study the following measurement strategies:

- *Termwise* — we take individual Pauli operators $Q_i \in \{Z_j, X_iX_j, Y_iY_j, Z_iZ_j\}$, and measure each Q_i using an independent measurement circuit.
- $XX + YY$ — we first measure all $Q_i \in \{Z_j, Z_iZ_j\}$ in a single-shot measurement. Next, we measure $Q_i \in \{X_jX_k + Y_jY_k\}$, grouping disjoint pairs j, k together into N sets following the scheduling outlined in Sec. B 3. This mostly matches the scheme used for the estimation of expectation values Raw VQE, PS-VQE and PS-VD in Fig. 2, and the scheme used for the PS-VQE estimates in Fig. 3.
- $XX + YY + IZ + ZI$ — here, we draw $Q_i \in \{Z_i, Z_iZ_j, D^+_{ij}\}$ ($D^\pm_{ij}$ is defined in Eq. (B27)), and measure each term separately. This matches the measurement scheme used for EV. As the operators chosen are Hermitian and unitary, the EV variance is equivalent to the standard tomography variance [63]. By choosing only D^+_{ij} and not adding D^-_{ij} to our set of measurements, the set $\{Q_i\}$ becomes linearly independent, and so the relative coefficients c_i (Eq.(C2)) are fixed.

Some notable differences occur between the numerical estimates made here and those implemented on hardware. (Ultimately the number of shots in the experiment was chosen to be low enough that the experimental error mostly dominated, rather than being optimized based on preliminary calculations.) In the experiment, additional estimates of $Z_j + Z_k$ were extracted alongside each measurement of $X_jX_k + Y_jY_k$ and combined using the techniques outlined in App. C 1. Each circuit was repeated with and without an additional π pulse on all qubits to unbias readout noise. In the experiment, the shot distribution was chosen to be uniform for VQE (40,000 per circuit) and VD (100,000 per circuit). (One can observe in Fig. 4 that an excess of shots was taken in both cases.) For EV, shots were distributed according to the relative weight of Q_i in the Hamiltonian. However, this was made more complicated by a technical requirement that all shots executed in a single batch must have the same number of repetitions. To accommodate this, the number of shots used to estimate a single Q_i was rounded up or down to be an integer multiple of a fixed base (40000). Furthermore, as mentioned in App. B 4, when performing EV to estimate each Q_i, 12 unique circuits were run to cancel out a background magnetic field and depolarize readout noise.

3. Shot distribution

Once the decomposition of the Hamiltonian is decided, the variance of the resulting energy estimator further depends on how the available shot budget is distributed over the Q_i (or the jointly measurable groups of Q_i). In principle, estimates of the (co-)variances from a small number of shots, from previous measurements at close by VQE parameter values according to (C13), or in adaptive schemes can be used to distribute shots in an asymptotically optimal way to reduce the variance. In practice, one is limited by the overhead of calling the device and limitations in setting the shots for individual circuits inside a batch (see Section D 1).

In our estimates we assume the ability to allocate shots per distinct circuit (and not only on a per-batch basis) but only consider the two non-adaptive shot distribution schemes that need no input from the quantum computer:

- *Uniform Distribution* — distribute the total shot budget per expectation value uniformly over the term groups, i.e., spend the same number of shots on each group of Q_i that are measured jointly according to the decomposition scheme.
- *According to term group weights* — the total shot budget is distributed proportional to the coefficients $|c_i|$ in Eq. C2. In the case that multiple Q_i are measured, shots are distributed according to $[\sum_i |c_i|^2]^{1/2}$. This is optimal in case the variances of all Q_i are equal and covariances vanish [17]. (Note that as the three measurement strategies measure only linearly-independent operators, the $|c_i|$ can be fixed prior to measurement.)

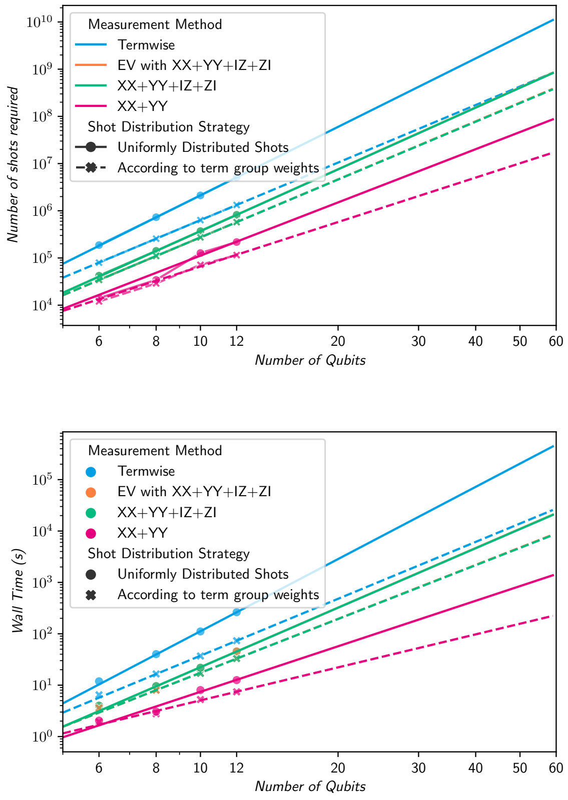


FIG. 6. Number of shots (top) and wall-clock time (bottom) required to estimate the ground-state energy of the RG Hamiltonian at $g = -0.9$ to within 0.1a.u.. Estimates were obtained by standard Lagrangian optimization [67] using term distributions defined in App. D 3 and measurement strategies defined in App. D 2. Wall-clock time model is described in App. D 1. The orange and green data points for the $XX+YY+IZ+ZI$ scheme with and without EV coincide.

4. Wall-clock time extrapolation to large system sizes

Using the protocol described in the previous sections, we are able to estimate the cost of measuring the RG model at $g = -0.9$ to within 0.1 a.u. as we enlarge the system size, while optimizing our shot distribution (App. D 3) for the wall-clock time (App. D 1). In Fig. 6, we plot the cost in terms of the number of shots (top) and in wall-clock time for superconducting hardware (bottom). Estimates are performed for $N = 6, 8, 10, 12$; given good fit to a line on a log-log plot and that a powerlaw scaling is expected, we are able to extrapolate this to larger system sizes. This data was combined with fidelity estimates for a 10-qubit system (Fig. 4[bottom-right]) to yield the curves in Fig. 4[top].

Appendix E: Variational optimization of a 6-qubit experiment

1. The Conjugate Model Gradient Descent optimizer

The large number of evaluations needed for ansatz parameter optimization on quantum hardware is a major impediment towards keeping the cost of variational algorithms (in terms of wall-clock time) manageable. To mitigate the overhead incurred by sending jobs to and receiving results from a device (see App. D 1), it is beneficial if the optimizer can request a batch of cost-function evaluations at once before making a step. In [68], surrogate model based optimizers were found to have good performance under this constraint. Here, a quadratic model function was fitted to present and past expectation value estimates in the vicinity of the current ansatz parameter vector. Then, after making a step, a batch of circuits corresponding to points in the vicinity of the new parameter vector were evaluated and the stepping procedure repeated. In this appendix we develop a natural extension of this procedure, by combining it with the conjugate descent algorithm.

The Conjugate Model Gradient Descent optimizer is a surrogate model-based optimization algorithm, with the additional improvement that the gradient which is extrapolated from fitting a quadratic model to the cost function, is used in the framework of conjugate gradient descent to make a step in the parameter space. The Conjugate Gradient Descent method was developed by Hestenes and Stiefel in [69]. For the special case of quadratic cost functions over an n dimensional space, conjugate gradient methods can be proven to converge in n iterations [70]. In practice the conjugate gradient method is found to work well for cost functions that are locally approximately quadratic. This is an assumption one anyway needs to make when using quadratic model based optimizers and thus makes it natural to combine conjugate gradient descent and quadratic surrogate model based optimizers with quadratic model functions. In conjugate gradient descent the steps are taken in the direction of the so-called conjugate gradient $s_n = g_n + \beta_n s_{n-1}$, where g_n is the gradient (in our case the one of the quadratic model fitted to the samples around the current position) and β_n is a scaling factor. The formula for the n-th iteration scaling factor is given in [71] as: $\beta_n^{FR} = \frac{g_n^T g_n}{g_{n-1}^T g_{n-1}}$. With $s_0 = g_0$ fixed for the first step of the algorithm.

In Alg. 1, we present our Conjugate Model Gradient Descent algorithm. We assume in this algorithm access to a (noisy) oracle to the target function f to be optimized. In practice, this is given by a call to a quantum device with a target error, that must be made small enough to enable convergence.

Algorithm 1 Conjugate Model Gradient Descent

```
Input: Initial point $x_0$, learning rate $\gamma$, sample radius $\delta$, maximum iterations $n$, number of evaluations per iteration $k$, rate decay exponent $\alpha$, stability constant A, sample radius decay exponent $\xi$, tolerance $\varepsilon$, oracle for function $f$.
Initialize lists L, L'
Initialize a list G
Initialize a list H
Let $x \leftarrow x_0$
for m in 0...n do
    Let $\delta' \leftarrow \delta/(m+1)^{\xi}$
    Sample $k$ points uniformly at random from the $\delta'$-neighborhood of $x$ to generate a set $S$
    for each $x'$ in $S \cup \{x\}$ do
        Add $(x', f(x'))$ to $L$
    end for
    Clear list $L'$
    for each tuple $(x', y')$ in $L$ do
        if $|x' - x| < \delta$ then
            Add $(x', y')$ in $L'$
        end if
    end for
    Fit $f(x) = x^T Ax + bx + c \sim y$ to the points $(x, y)$ in $L'$ using least squares linear regression.
    Let $g_m$ be the gradient of $f$ at $x$ (i.e. $g_m = b$).
    if $|g_m| < \varepsilon$ then
        return $x$
    end if
    if $m = 0$ then
        Let $h_0 \leftarrow g_0$
    else
        $\beta_m \leftarrow g_m^T g_m / g_{m-1}^T g_{m-1}$
        $h_m \leftarrow g_m + \beta_m h_{m-1}$
    end if
    $\gamma' = \gamma/(m+1+A)^{\alpha}$
    Add $g_m$ to the list G
    Add $cg_m$ to the list H
    Let $x \leftarrow x - \gamma' \cdot h_m$
    Let $m \leftarrow m + 1$
end for
return $x$
```

2. Hyperparameter tuning

For all experiments shown in the main text, parameters were obtained by a noiseless simulation using L-BFGS-B, starting from $\boldsymbol{\theta} = 0$. We find that in the absence of experimental noise, gradient-based optimizers such as L-BFGS-B converge well to an optimal solution. However, this is not the case in the presence of experimental or sampling noise, which can make many estimators unstable. This has lead previous efforts towards stochastic-based [28] or model-based [27] optimizers, including the Model Gradient Descent optimizer that we have based our conjugated Model Gradient Descent algorithm upon in the previous section.

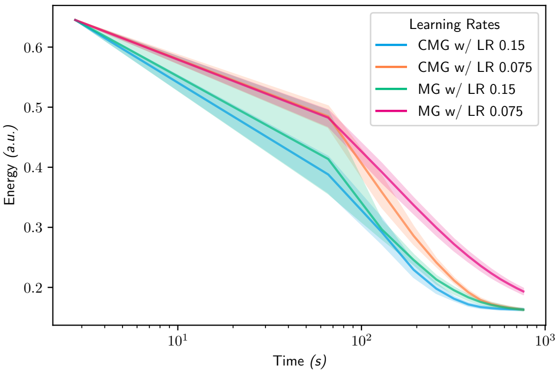


FIG. 7. In this figure, the distance between the energy that was calculated in the noiseless regime by the L-BFGS-B optimizer is plotted over the wall-time required to reach such resolution. The system displayed is a RG Hamiltonian for 6 qubits, with a coupling constant of $g = -0.9$, consisting of 10 runs and the lines being the average of those runs for the Conjugate Model Gradient Descent (CMG) and the Model Gradient Descent (MG) found in [68].

The hyperparameters used in the Conjugate Model Gradient Descent algorithm were either chosen through an intuitive approach due to their physical meaning, such as the sample radius due to the parameters being perturbed by a random variable from [-0.25,0.25] and the maximum number of iterations to stay within reasonable wall-time, or via grid search for the rest of the hyperparameters, ensuring they are good enough to work for the variety of cases examined while avoiding overfitting. Explicitly, the hyperparameters used are included in Table I:

Hyperparameter	Values Used
Sample radius δ	1.0
Learning rate γ	0.15
Stability constant A	0
Sample number k	$0.409(N+1)(N+2)$
Sample radius decay exponent ξ	0
Rate decay exponent α	0.2
Maximum evaluations n	12

TABLE I. Hyperparameters for the Conjugate Model Gradient Descent optimizer during the experiment, where N is the number of parameters in the optimization. These were also used in the later comparisons with Model Gradient Descent, to gauge performance in equal footing.

The hyperparameters we chose work equally well for Model Gradient Descent, as well our Conjugate variant. We found that the conjugate method can sometimes speed up convergence or help prevent getting stuck in local minima. We illustrate this with two examples: In Fig 7 how Conjugate Model Gradient Descent has the ability to speed up in case the initial learning rate was chosen to be too small. In Fig. 8 we show an example of hyper parameters for which Conjugate Model Gradient Descent is able to converge to a lower minimum than plain Model Gradient Descent.

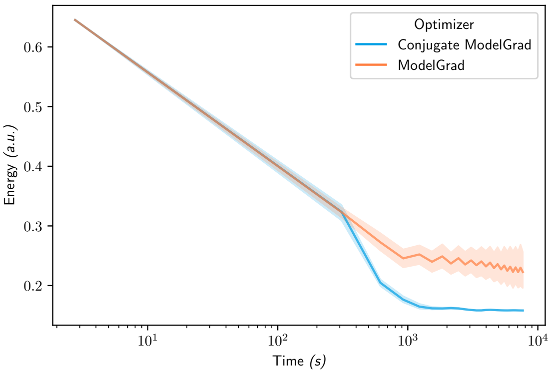


FIG. 8. Comparison of the two variants of the Model Gradient Descent optimizer, where in this case the number of samples used to construct the quadratic model one uses is higher, namely $k = 4(N+1)(N+2)$, allowing for the performance difference of the two optimizers to become more pronounced, with Conjugate Model Gradient Descent managing to consistently converge at better parameters, while Model Gradient Descent gets stuck in a local minimum, averaged over 10 runs for each optimizer, for 25 maximum iterations.

3. Experimental results

We test the hyperparameter-tuned Conjugate Model Gradient Descent algorithm on the problem of optimizing the UpCCD ansatz for the ground state of the RG Hamiltonian at a coupling $g = -0.9$. In order to demonstrate convergence, we perturb our ansatz parameters from values optimized on a noiseless classical simulation (using the L-BFGS-B optimizer as described above), by a random variable drawn from $[-0.25, 0.25]$. From this point, Conjugate Model Gradient Descent converged in 9 iterations with each iteration requiring 46 calls to the cost function (414 calls total). We observe that the optimizer successfully finds an energy below the initial point ($0.07 a.u.$), which demonstrates the well-known VQE ability to mitigate coherent noise [9, 10]. This improvement is reflected in the results mitigated with echo verification as well, which demonstrated a $0.04 a.u.$ reduction in error. However, this is a relatively small reduction compared to the overall error observed. We note that this is a significantly higher error than that observed in Fig. 12, which we attribute to poor calibration on the day. If the absolute gain in energy was replicated in Fig. 12, this would account for $\sim$ 40% of the overall error. However, the relative gain in error in this case was only $\sim$ 10%. Ultimately, as the cost of optimization was already significant enough for this 6-qubit example, and as the number of calls for Conjugate Model Gradient Descent scales as $O(\#\text{parameters}) \sim O(N^2)$, we did not see it practical to continue this line of research in this work. With current qubit counts and shot repetition rates the optimization of variational algorithms of meaningful size remains a major obstacle.

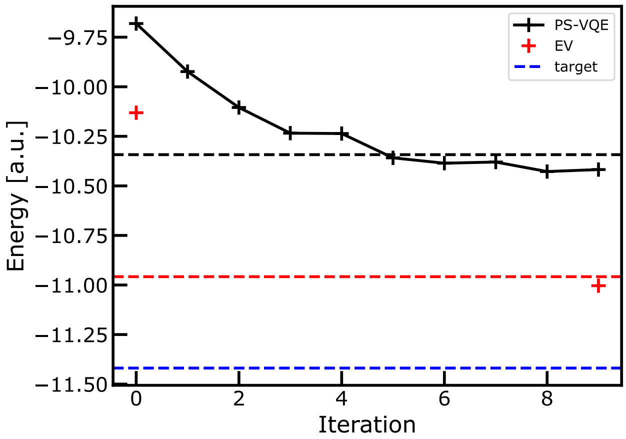


FIG. 9. Optimization trace of a 6 qubit RG simulation at $g = -0.9$, targeting the blue dashed line (ground state energy in the seniority-zero space). After an energy estimation using postselection (black dashed line) and echo verification (red dashed line) at optimized parameters from a noiseless simulation, all parameters were perturbed by a random variable drawn from $[-0.25, 0.25]$, and reoptimized over 9 iterations of Conjugate Model Gradient Descent, using the post-selected experimental data as a cost function. After perturbation and after convergence, a single call was made with the converged parameters to an echo verified energy estimation (red cross).

Appendix F: Additional results and analysis

1. Virtual distillation with and without postselection

In this appendix, we show the effect that postselection has on virtual distillation. In Fig. 10, we repeat the plot of Fig. 2, but with data from VD without postselection on top. (Other lines are removed to make the plot clearer.) We see that by itself, VD [light blue curve] outperforms postselection in terms of energy estimation across most points of the energy curve, but it is not variationally bound. This is because without postselection, VD returns unphysical estimates $|\langle Z_i\rangle| \geq 1$ and $|\langle D^{\pm}_{ij}\rangle| \geq 1$. (This is a consequence of the two copies of the state not necessarily being subject to the same noise.) As a consequence of this, our estimate of the order parameter is complex and non-physical, and therefore not plotted. This issue can be rectified somewhat by bounding the estimates of $\langle Z_i\rangle$ and $\langle D^{\pm}_{ij}\rangle$ to lie in $[-1, 1]$. Performing this (Fig. 2, purple curve) allows for an estimate of the order parameter to be made, however it is clearly qualitatively incorrect. Moreover, although the energies are shifted up, some are still not variationally bounded, and those that are, often overshoot the energy, making the absolute error worse. (The variational bound could be rectified by enforcing positivity conditions on the set of

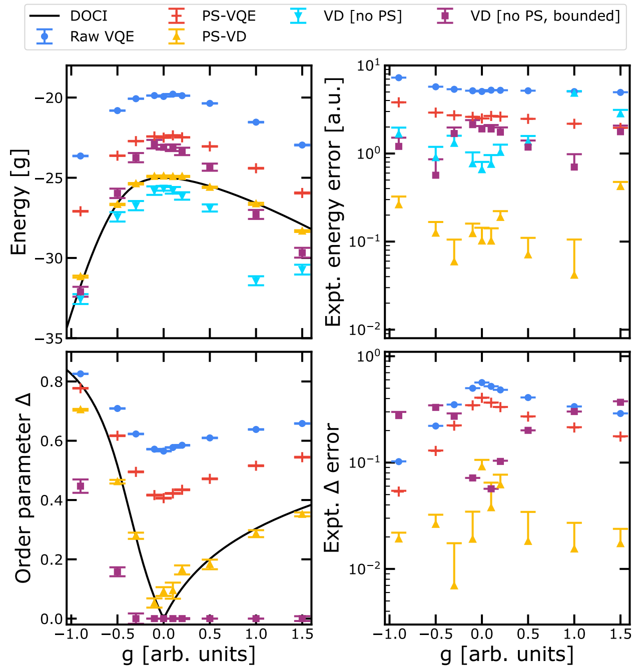


FIG. 10. Comparison of virtual distillation with postselection (yellow), without postselection (light blue), and without postselection but while bounding expectation values of Z_j and $D^{\pm}_{ij}$ in $[-1, 1]$ (purple), for a 10-qubit simulation of ground states of the RG Hamiltonian across a range of g values. Raw VQE (blue), postselected VQE (red), and post-selected VD [PS-VD] data taken from Fig. 2. VQE error bars obtained by error propagation (1 standard deviation, see App. C 2), VD error bars obtained by bootstrapping (1 standard deviation).

generated estimates [67].) In summary, the combination of postselection and virtual distillation is seen here to have a greater effect than the sum of its parts for energy error.

We attribute the gain from postselection in virtual distillation to two sources. Firstly, postselection removes final readout noise, which VD does not naturally correct against. (As the estimation of $\mathrm{Tr}[\rho^2]$ involves a highly correlated measurement, this cannot be easily corrected by most readout correction schemes designed to estimate local Pauli expectation values.) Secondly, though we are in principle only post-selecting on the sum of the number of excitations in $\rho^{(1)}$ and the number of excitations in $\rho^{(2)}$ being equal to N, when combined with virtual distillation this projects out all noise such as T1 decay and suppresses any particle-non-conserving noise to second order. In short, this is because breaking number conservation separately on $\rho^{(1)}$ and $\rho^{(2)}$ yields states which do not overlap (as they must have different particle number), and these states cancel out when taking the product $\rho^{(1)}\rho^{(2)}$. Let us study this in more detail. Consider a post-selected estimate of $\mathrm{Tr}[\rho^2 O]$ using VD for $O = Z_j$ (as described in the main text). After postselecting on $\sum_j (1 \otimes Z_j + Z_j \otimes 1) = N$, we can write our global state as

$$\rho_2 = \sum_{p,q,r,s} c_{p,q,r,s} |p\rangle\langle q| \otimes |r\rangle\langle s|, \tag{F1}$$

where p, q, r, s index the basis states of both systems. (Following projection, ρ_2 will no longer be a product state.) Let us write $n_p = \langle p| \sum_j Z_j |p\rangle$ etc as the number of excitations in these basis states (i.e. the Hamming weight of the index), and our projection requires $c_{p,q,r,s} = 0$ unless $n_p + n_r = n_q + n_s = N$. (This assumes WLOG we are at half-filling, and ideally $n_p = n_r = n_q = n_s = N/2$.) Then, as O_s preserves particle number for $O = Z_j$, the only contribution to

$$\begin{aligned} &\mathrm{Tr}\big[\rho_2 S^{(2)} (Z_j \otimes I + I \otimes Z_j)\big] \\ &\quad = \sum_{p,q,r,s} c_{p,q,r,s} \delta_{s,p} \delta_{r,q} \Big(\langle s|Z_j|s\rangle + \langle r|Z_j|r\rangle \Big), \end{aligned} \tag{F2}$$

comes from those terms $c_{p,q,p,q}$. (The same is true for our estimate of $\mathrm{Trace}[\rho_2 S^{(2)}]$.) This implies that a non-zero contribution to $\mathrm{Trace}[\rho_2 S^{(2)}]$ comes only from matrix elements $|p\rangle\langle q| \otimes |p\rangle\langle q|$ where $n_p = \frac{N}{2} + \delta$, $n_q = \frac{N}{2} - \delta$. When $\delta = 0$, our matrix element $|p\rangle\langle q| \otimes |p\rangle\langle q|$ is in the number-conserving sector. When $\delta \neq 0$, our matrix element corresponds to a product of coherent superpositions between the $N/2 + 1$ and $N/2 - 1$ sector on both qubits. This implies that noise channels such as the T1 channel will be completely mitigated as these off-diagonal elements do not exist. Coherent noise may cause some of these off-diagonal elements to appear (consider e.g. a single-qubit X rotation on the state $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$.), however this is required to happen on both states to contribute, which suppresses it to second order. (This is in contrast to coherent noise that preserves number, to which we can be first-order sensitive.) The above analysis has assumed that the postselection is perfect; readout noise would complicate the above.

2. Distribution of errors in Pauli expectation value estimation

In Fig. 11, we plot a histogram of the error in estimating the expectation values of Pauli operators, taking data from Fig. 2, Fig. 12, and Fig. 3. We observe that the mean error in both cases for cyclobutene is slightly worse than for the RG model, but only by a small percentage. (This justifies our claim in Sec. IV that a 0.1 a.u. error in the RG Hamiltonian is approximately $1-2$ orders of magnitude larger than chemical accuracy for a 10-qubit system.) As the difference between systems here is minimal (changing only the value of some virtual Z rotations), we can only attribute this difference to the performance of the device whilst taking these datasets. Going from 6 to 10 qubits increases the mean error in all experiments by a factor $1.5-2\times$. As the Hamiltonian for both the RG model and cyclobutene has a number of terms scaling as

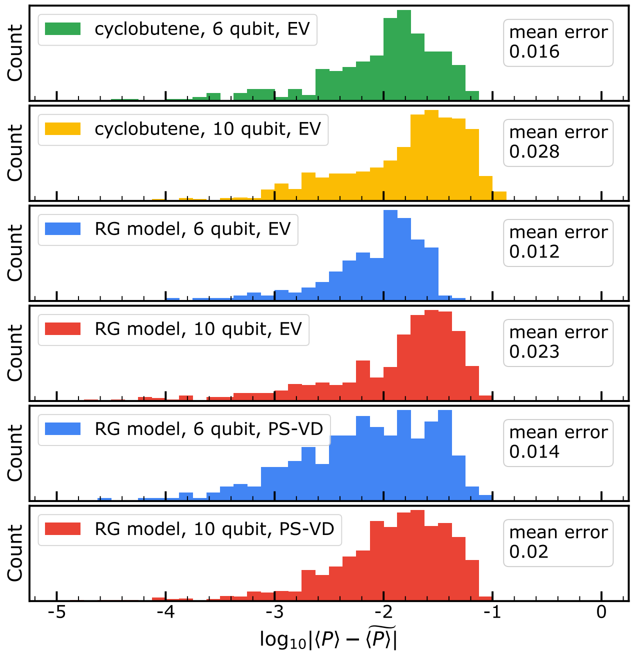


FIG. 11. Histogram of the expectation value error in Pauli operators $P = Z_j$, $P = Z_iZ_j$ or $P = D^{\pm}_{ij}$ (Eq. (B27)), across all data taken for the experiment mentioned. Mean error across the dataset is given.

$O(N^2)$; assuming that the errors follow a roughly Gaussian distribution would predict the energy error scales $O(N)\times$ the error per each individual operator. This then predicts a gain in energy of $2.5-3.3\times$, which lies in between the observation of Fig. 3 and Fig.4. This suggests that the RG model energy estimation at large N may have had some benevolent cancellation of noise, and likewise for the cyclobutene energy estimation at small N.

3. Smaller studies of the RG Hamiltonian

In Fig. 12, we present experimental simulations of the ground state of the RG Hamiltonian for 4, 6, and 8 qubits. The method to produce these figures is identical to that used in the production of Fig. 2, save for the number of qubits and shots used. Aggregated data from these figures was used to generate the scaling plots of Fig. 4.

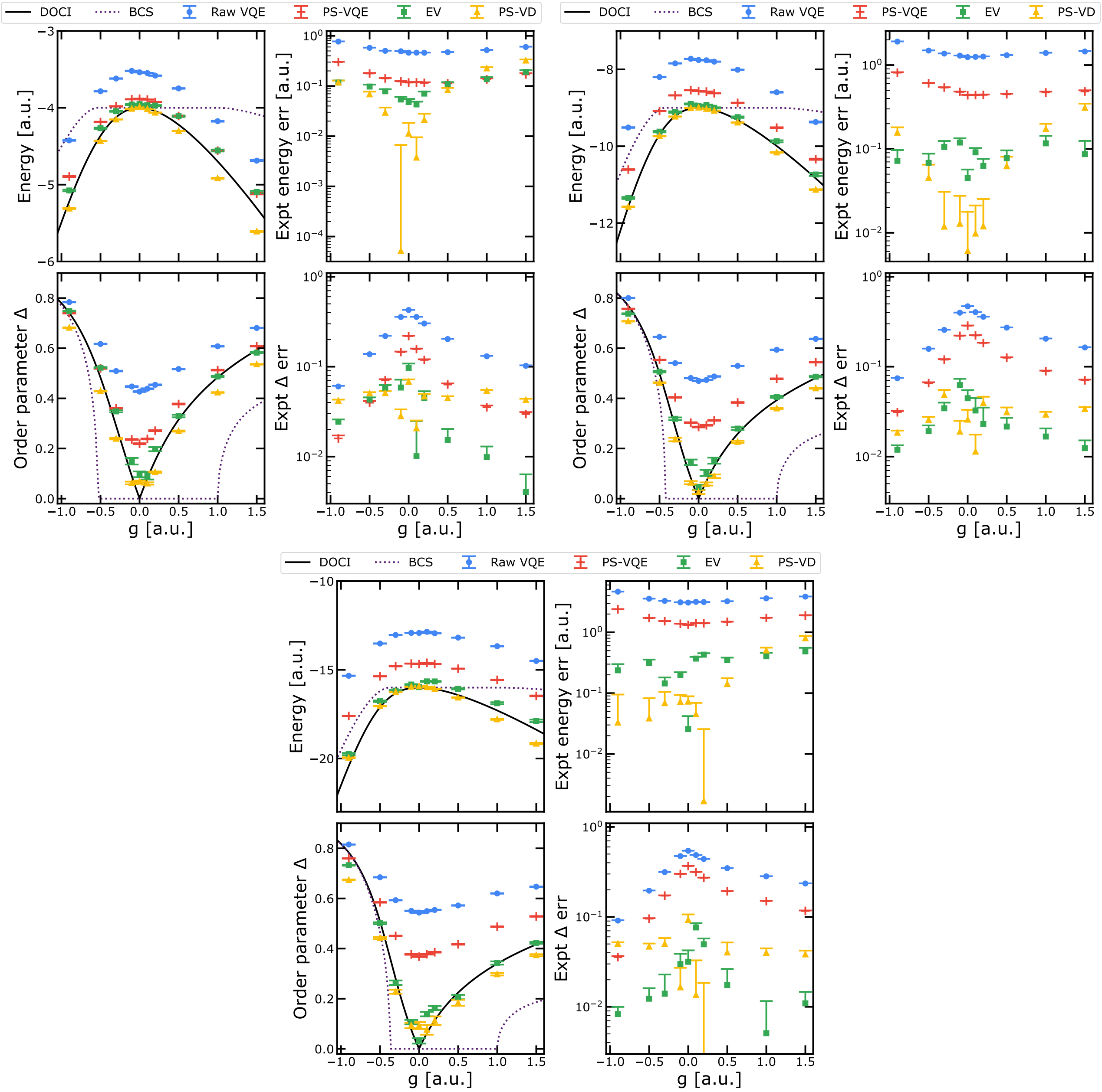


FIG. 12. Identical experiment to Fig. 2, but for 4 (top-left), 6(top-right), and 8 (bottom) qubits instead of 10. Aggregate data is used in Fig. 4. (top-left) Energy plot for the RG system as a function of the coupling parameter g, for an unmitigated state preparation [blue circles], and state preparation mitigated by postselection [red crosses], echo verification [yellow triangles], and postselected virtual distillation [green squares]. This is compared to the exact DOCI result [black solid line], and BCS [purple dashed line]. (top-right) Log plot of experimental energy error (ignoring the model error from the UpCCD approximation). (bottom-left) Superconducting order parameter for the RG Hamiltonian ($\Delta = \frac{1}{N}\sum_{j,\sigma}\sqrt{\langle n_{j\sigma}^2\rangle - \langle n_{j\sigma}\rangle^2}$), again compared to classical models. (bottom-right) Experimental error in estimating the superconducting order parameter vs the target state within the UpCCD approximation (again ignoring model error). 1 std. dev. error bars estimated by either propagating variance (Raw VQE, PS-VQE) or bootstrapping (EV, PS-VD).

5 Local, expressive, quantum - number - preserving VQE ansätze for fermionic systems

Bibbliographic information

Anselmetti, G.-L. R., Wierichs, D., Gogolin, C., Parrish, R. M. (2021, November). Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems. New Journal of Physics, 23 (11), 113010. Retrieved 2023-04-06, from https://dx.doi.org/10.1088/1367-2630/ac2cb3 (Publisher: IOP Publishing) doi: 10.1088/1367-2630/ac2cb

Summary

In this work we are concerned with ansätze in VQE tailored towards fermionic simulations for quantum chemistry. We propose VQE circuit fabrics with advantageous properties for the simulation of strongly correlated ground and excited states of molecules and materials under the Jordan–Wigner mapping that can be implemented linearly locally and preserve all relevant quantum numbers: number of spin up (α) and down (β) electrons and the total spin squared $\hat{S}^2$. We derive decompositions of all the operations used in our gate fabric and parameter-shift rules to calculate derivatives for optimization. Large scale simulations running on high-performance-computing hardware simulating VQEs on up to 20 qubits, up to 500 layers of our gate fabric and up to a week of node-computation time investigate convergence behaviour for known strongly correlated systems. We discuss initialisation strategies to avoid barren plateaus during convergence and to improve general trainability of these systems and investigate the shape of the wave function-approximations these circuits generate in relation to their depth. A qualitative comparison with other entanglement circuits is part of the work.

Contributions

The author contributed the main part of the work - (1) establishing the quantum-number preserving gate set, (2) numerical experiments on convergence on up to 20-qubit VQEs with up to 4500 parameters to chemical accuracy on high-performance-computing hardware, (3) finding initialisation strategies in the quantum-number preserving ansatz to avoid barren plateaus and (4) investigation of the behaviour of the approximated wave functions present in the VQE after optimisation. (5) Further contributions are writing substantial parts of the manuscript, e.g. the numerical demonstration and contributing to introduction, comparison and discussion.

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems

To cite this article: Gian-Luca R Anselmetti *et al* 2021 *New J. Phys.* **23** 113010

IOP Publishing New J. Phys. **23** (2021) 113010 https://doi.org/10.1088/1367-2630/ac2cb3

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

RECEIVED
17 June 2021

REVISED
20 September 2021

ACCEPTED FOR PUBLICATION
4 October 2021

PUBLISHED
12 November 2021

PAPER

Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems

Gian-Luca R Anselmetti[1], **David Wierichs**[2], **Christian Gogolin**[1] and **Robert M Parrish**[3,*]

[1] Covestro Deutschland AG, Leverkusen 51373, Germany
[2] Institute for Theoretical Physics, University of Cologne, Germany
[3] QC Ware Corporation, Palo Alto, CA 94301, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: gian-luca.anselmetti@covestro.com, wierichs@thp.uni-koeln.de, christian.gogolin@covestro.com and rob.parrish@qcware.com

Keywords: variational quantum eigensolver, particle symmetries, entangler circuits

Abstract

We propose VQE circuit fabrics with advantageous properties for the simulation of strongly correlated ground and excited states of molecules and materials under the Jordan–Wigner mapping that can be implemented linearly locally and preserve all relevant quantum numbers: the number of spin up (α) and down (β) electrons and the total spin squared. We demonstrate that our entangler circuits are expressive already at low depth and parameter count, appear to become universal, and may be trainable without having to cross regions of vanishing gradient, when the number of parameters becomes sufficiently large and when these parameters are suitably initialized. One particularly appealing construction achieves this with just orbital rotations and pair exchange gates. We derive optimal four-term parameter shift rules for and provide explicit decompositions of our quantum number preserving gates and perform numerical demonstrations on highly correlated molecules on up to 20 qubits.

1. Introduction

Hybrid quantum classical variational algorithms, including those of the variational quantum eigensolver (VQE) type [1, 2], are among the leading candidates for quantum algorithms that may yield quantum advantage in areas such as computational chemistry or machine learning already in the era of noisy intermediate scale quantum (NISQ) computing [3]. A foundational issue in VQE [1, 2], and in many of its extensions and alternatives [4–11], is finding a 'good' definition of the entangler circuit. Here the qualifier 'good' has many facets, possibly including: (1) providing an efficient approximate representation of the target quantum states in the limit of an intermediate (ideally polynomial-scaling) depth (2) consisting of a low number of distinct physically realizable gate elements (3) exhibiting a simple pattern of how these gate elements are applied (4) exhibiting sparse spatial locality that is further compatible with device connectivity (5) exhibiting simple analytical gradient recipes and robust numerical convergence behavior during optimization of the VQE energy, e.g. by mitigating the effects of barren plateaus [12] (6) respecting exactly the natural particle and spin quantum number symmetries of the target quantum states, e.g. as notably explored in [13] (7) providing an exact representation of the target quantum states in the limit of sufficient (usually exponential-scaling) depth.

Especially within the context of VQE for spin-1/2 fermions governed by real, spin-free Hamiltonian operators (e.g. electrons in molecules and materials, the prime application of VQE), a variety of compelling VQE entangler circuit recipes have been discussed in the literature. Prominent examples include UCCSD [1, 14, 15], *k*-UpCCGSD [16, 17], Jastrow-factor VQE [18], the symmetry-preserving ansätze [13], the hardware efficient ansätze [19, 20], ADAPT-VQE [21], pUCCD [22], and additional methods discussed below [23–28]. Each of these generally satisfies a subset of the 'good' facets listed above, though no extant

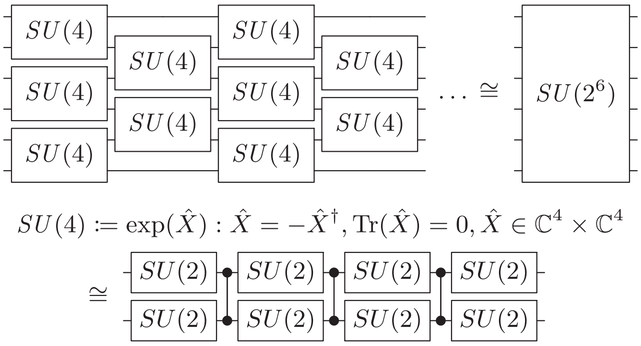


Figure 1. Sketch for $N = 6$ of a gate fabric universal for $\mathcal{SU}(2^N)$ providing inspiration for the fermionic QNP gate fabrics developed here. The gate fabric is a two-local-nearest-neighbor tessellation of alternating 15-parameter, two-qubit $SU(4)$ gates. The $SU(2)$ gate in the $SU(4)$ gate decomposition on the bottom line is the three-parameter universal gate for the one-qubit Bloch sphere. The indicated 24-parameter decomposition of $SU(4)$ is overcomplete for the 15-parameter $\mathcal{SU}(4)$ group.

ansatz that we are aware of obtains all of them, with the notable exception of the generalized swap network form of k-UpCCGSD of [17].

In this work we develop a VQE entangler circuit recipe for fermions in the Jordan–Wigner representation and show, or at least provide evidence, that it obtains all facets, with facet (5) partially left to future numerical studies. Perhaps the most notable property of our fabrics is the exact preservation of all relevant quantum numbers the individual gate elements of the fabric, which is why we refer to them as quantum number preserving (QNP). This property may be critical for employment of VQE in larger systems, where contaminations from or even variational collapse onto states with different particle or spin quantum numbers can severely degrade the quality of the VQE wavefunction.

Note that after we posted the first version of our manuscript, we became aware of the generalized swap network reformulation of k-UpCCGSD of [17]. This paper refactors k-UpCCGSD to use nearest-neighbor connectivity, yielding a circuit fabric that could be written in terms of four-qubit gates containing diagonal pair exchange and orbital rotation elements in a very similar manner as our $\hat{Q}$-type QNP gate fabric discussed below. There are some tactical differences in the qubit ordering and the generalized swap network paper does not emphasize the role of quantum number symmetry as much as the present manuscript. Moreover, the origin of the $\hat{Q}$-type QNP gate fabric as a simplification of our more-complete $\hat{F}$-type QNP gate fabric of appendix D provides a markedly different approach to developing this gate fabric. In any case, we encourage any readers interested in the present manuscript to also explore [17].

2. Gate fabrics

Our VQE entangler circuit recipe draws inspiration from the well known fact that the qubit Hilbert space (without any fermionic symmetries) can be spanned by a tessellation of two-qubit gates universal for $\mathcal{SU}(4)$ in alternating layers (see figure 1). This tessellation can formally be repeated to infinite depth. However, one finds that after some finite, N-dependent critical depth of order $\mathcal{O}(2^{2N})$, additional gate layers do not increase the expressiveness of the circuit, as formal completeness (denoted 'universality') in $\mathcal{SU}(2^N)$ is achieved. In practice usually shorter circuit depths are of interest. For instance, one may consider the case where the tessellation is restricted to be polynomial scaling in N, in which case universality cannot be exactly achieved. However, a good approximation of specialized (e.g. physically relevant) parts of some subgroup may still be achievable in a way that is tractable to compute even on an NISQ computer but intractable to compute with a classical device.

Note that it is difficult to find a single landmark reference explicitly proving that the fabric in figure 1 is universal for $\mathcal{SU}(2^N)$, but every research group we have discussed the matter with acknowledges that this gate fabric is widely known in the field. Moreover, it is simple to prove that this gate fabric is universal for $\mathcal{SU}(4)$ if one starts from the well-known fact that circuits composed of arbitrary one-qubit operations and CX or CZ gates between arbitrary pairs of qubits are universal for $\mathcal{SU}(2^N)$ [29]. One can picture an arbitrary-depth version of the gate fabric in figure 1 where only one $\hat{S}U(2)$ or CZ gate at desired qubit indices is active in each layer (the latter possible by repeating layers of CZ gates with trivial $\hat{S}U(2)$ gates interleaving). Then the only requirement is to extend the local nearest neighbor connections of CZ gates to arbitrary pairs of qubit indices. This is easily accomplished by adding additional layers of the fabric whose $\hat{S}U(2)$ gates interleaving). Then the only requirement is to extend the local nearest gates are initialized to implement SWAP gates to expand the linear nearest-neighbor CZ gates to arbitrary connectivity. And thus

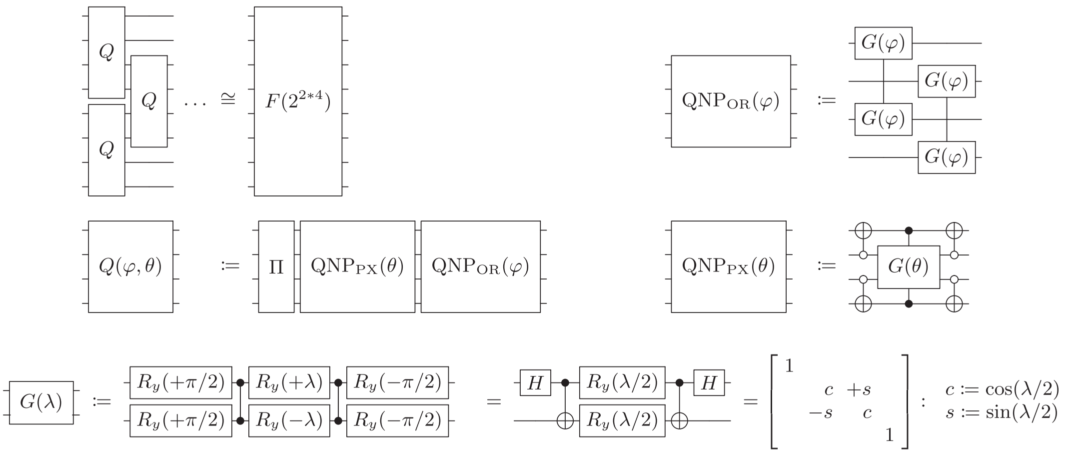


Figure 2. Proposed gate fabric for $\mathcal{F}(2^{2M})$ (sketched for $M = 4$). The spin orbitals in Jordan–Wigner representation are in 'interleaved' ordering with even (odd) qubit indices denoting α (β) spin orbitals. The Jordan–Wigner strings are taken to be in 'α-then-β' order. The gate fabric is a four-local-nearest-neighbor-tessellation of alternating even and odd spatial-orbital-pair two-parameter, four-qubit $\hat{Q}$ gates. Each $\hat{Q}$ gate has two independent parameters and contains a one-parameter, four-qubit spatial orbital rotation gate $\mathrm{QNP}_{\mathrm{OR}}(\varphi)$ and a one-parameter, four-qubit diagonal pair exchange gate $\mathrm{QNP}_{\mathrm{PX}}(\theta)$. The order of $\mathrm{QNP}_{\mathrm{OR}}$ and $\mathrm{QNP}_{\mathrm{PX}}$ (note $[\mathrm{QNP}_{\mathrm{OR}}, \mathrm{QNP}_{\mathrm{PX}}] \neq 0$) does not seem to substantially change expressiveness at intermediate depths. $\mathrm{QNP}_{\mathrm{OR}}(\varphi)$ implements the spatial orbital Givens rotation $|\phi_0\rangle = c|\phi_0\rangle + s|\phi_1\rangle$ and $|\phi_1\rangle = -s|\phi_0\rangle + c|\phi_1\rangle$, with the same orbital rotation applied in the α and β spin orbitals. $\mathrm{QNP}_{\mathrm{PX}}(\theta)$ implements the diagonal pair Givens rotation, $|0011\rangle = c|0011\rangle + s|1100\rangle$ and $|1100\rangle = -s|0011\rangle + c|1100\rangle$. The real one-parameter, one-qubit rotation gate is $\hat{R}_y(\lambda) := \mathrm{e}^{-\mathrm{i}\lambda\hat{Y}/2}$. In $\hat{Q}$, we include the optional constant $\hat{\Pi}$ gate, for which natural choices include the four-qubit identity gate, i.e. $\hat{\Pi} = \hat{I}$, or the fixed spin-adapted orbital rotation gate $\hat{\Pi} = \mathrm{QNP}_{\mathrm{OR}}(\pi)$. In the latter case, the gate fabric with all parameters $\{\theta = 0\}$ and $\{\varphi = 0\}$ promotes exchange of orbitals. We find that the choice of $\hat{\Pi} \in \{\hat{I}, \mathrm{QNP}_{\mathrm{OR}}(\pi)\}$ does not appear to affect the expressiveness of the quantum circuit, but the latter choice has turned out to be advantageous during gradient-based parameter optimization. This can additionally be mixed with a non-trivial initialization of the $\mathrm{QNP}_{\mathrm{OR}}$ gates, to, e.g. angles of $\varphi = \pi/2$ as we illustrate below. Regardless of the choice of $\hat{\Pi}$, this gate fabric exactly preserves the real nature of the subgroup, exactly commutes with the $\hat{N}_\alpha$, $\hat{N}_\beta$, and $\hat{S}^2$ symmetry operators, and numerically appears to provide universality at sufficient parameter count.

one obtains a circuit with one-qubit $\mathcal{SU}(2)$ gates with free parameters at arbitrary desired qubit positions interleaved with CZ gates between arbitrary pairs of qubit positions, which is well known to be universal. This construction suffices to prove universality, but is obviously extremely wasteful. In practice, we find that unconstraining all of the $\mathcal{SU}(2)$ parameters markedly improves the expressivity of the gate fabric at a given depth, and that action or operator universality is numerically obtained when the number of free $\mathcal{SU}(2)$ parameters is similar (strictly greater than or equal to) the number of free parameters in the many-body unitary action or operator.

Particularly striking in figure 1 is the locality (alternating nearest neighbor connectivity) and simplicity (single gate element) of the circuit, properties of what we call a 'gate fabric.' More precisely, throughout this manuscript, we define a gate fabric for a subgroup of $\mathcal{SU}(2^N)$ to be a tessellation of gates over N-qubits with the following properties:

(a) Simplicity: composed of a single type of k-qubit, l-parameter gate element (with a known decomposition into elementary gates), where k and l are independent of N.

(b) Linear locality: when the qubits are thought of as arranged on a vertical line the gate elements are arranged in layers and connect up to k contiguous qubits.

(c) Universality: achieving universality within the target subgroup of $\mathcal{SU}(2^N)$ within a finite number of layers depending on N.

(d) Symmetry: commuting with all symmetry operators used to define the subgroup of $\mathcal{SU}(2^N)$, i.e. $[\hat{U}, \hat{N}] = 0$, where $\hat{U}$ is the circuit unitary for any set of parameters and $\hat{N}$ is the symmetry operator.

Depending on the subgroup of $\mathcal{SU}(2^N)$ of interest it can be more or less difficult to find fabrics akin to the one shown in figure 1. In appendix A we discuss the trivial restriction to $\mathcal{SO}(2^N)$ and the less-trivial restriction to subspaces of definite Hamming weight.

3. Gate fabrics for fermions under the Jordan–Wigner mapping

The focus of this work is the construction of gate fabrics for the subgroup $\mathcal{F}(2^{2M}) \in \mathcal{SU}(2^N)$ constrained to spin-restricted fermionic symmetry under the Jordan–Wigner representation. To make this more precise, we define M real orthogonal spatial orbitals $\{|\phi_p\rangle\}_{p=0}^{M}$. For each spatial orbital, we define corresponding α

(β) spin orbitals $|\psi_p\rangle := |\phi_p\rangle|\alpha\rangle$ ($|\psi_{\bar{p}}\rangle := |\phi_p\rangle|\beta\rangle$) for a total of $N := 2M$ spin orbitals in a spin-restricted formalism. We associate the occupation numbers of these spin orbitals with the occupation numbers of N qubits. We number the qubits in 'interleaved' ordering $\ldots|1_\beta\rangle|1_\alpha\rangle|0_\beta\rangle|0_\alpha\rangle$. The fermionic creation/annihilation operators are defined in terms of the qubit creation/annihilation operators via the Jordan–Wigner mapping in 'α-then-β' ordering, $p^\pm := (\hat{X}_p \mp \mathrm{i}\hat{Y}_p)/2\bigotimes_{q=0}^{p-1}\hat{Z}_q$ and $\bar{p}^\pm := (\hat{X}_{\bar{p}} \mp \mathrm{i}\hat{Y}_{\bar{p}})/2\bigotimes_{q=0}^{p-1}\hat{Z}_{\bar{q}}\bigotimes_{q=0}^{M-1}\hat{Z}_q$. We note that for the majority of applications in the space of spin-1/2 fermions, the governing Hamiltonians are real (e.g. for non-relativistic electronic structure theory), and so we restrict from complex to real unitary operators, i.e. $\mathcal{SU}(2^N) \to \mathcal{SO}(2^N)$. The spin-restricted fermionic subgroup is then defined as the subgroup of $\hat{U} \in \mathcal{SO}(2^N)$ that respect the commutation relations $[\hat{U}, \hat{N}_\alpha] = 0$, $[\hat{U}, \hat{N}_\beta] = 0$, and $[\hat{U}, \hat{S}^2] = 0$. Here the α (β) number operator is $\hat{N}_\alpha := \sum_p p^\dagger p = \sum_p(\hat{I} - \hat{Z}_p)/2$ $[\hat{N}_\beta := \sum_p \bar{p}^\dagger \bar{p} = \sum_p(\hat{I} - \hat{Z}_{\bar{p}})/2]$ [30]. The spin-squared operator is $\hat{S}^2 := \sum_{pq} pq^\dagger \bar{p}^\dagger \bar{q} + (\hat{N}_\alpha - \hat{N}_\beta)/2 + (\hat{N}_\alpha - \hat{N}_\beta)^2/4$, and does not admit a local description in terms of Pauli operators in the Jordan–Wigner basis (we provide further details in appendices B and C). We denote this real subgroup, preserving $\hat{N}_\alpha$, $\hat{N}_\beta$, and $\hat{S}^2$, as $\mathcal{F}(2^{2M})$.

Naively one might expect that there should not be any local gate fabric exactly preserving all three fermionic quantum numbers, since the $\hat{S}^2$ operator is non-local. The crux of this work is thus the simple quantum-number-preserving gate fabric of figure 2. This gate fabric is composed of two-parameter four-qubit gate elements $\hat{Q}$, each composed of a one-parameter four-qubit spin-adapted spatial orbital rotation gate $\mathrm{QNP_{OR}}(\varphi)$ and a one-parameter four-qubit diagonal pair exchange gate $\mathrm{QNP_{PX}}(\theta)$. We describe further related quantum-number-preserving gate fabrics for $\mathcal{F}(2^{2M})$ in appendix D—these were the progenitors of the simpler gate fabrics shown in the main text, and may have advantageous properties in specific realizations of VQE entangler circuits.

Facets (2)–(4) of gate fabrics are manifestly fulfilled for all these proposals and facet (6) holds by construction as all gates individually preserve all quantum numbers. For facets (1) and (7) we provide numerical evidence below and in appendix G. It is worth noting that these tests numerically indicate that our gate fabrics are universal in the vast bulk of quantum number irreps, i.e. they may be used for cases where $S \neq 0$ and/or where $N_\alpha \neq N_\beta$ (including both even and odd spin cases). Note that there are a few edge case quantum number irreps where the $\hat{Q}$-type gate fabrics are not universal—see next paragraph for discussion. We believe that the methods from [31, 32] or [23] can be used to rigorously show universality in the (bulk of the) quantum number sectors of $\mathcal{F}(2^{2M})$ in all cases (as well as that our circuits are polynomial depth ϵ-approximate unitary t-designs and form ϵ-nets). Working out the details of a rigorous proof, which we believe has to be done spin sector by spin sector in some cases, is however beyond the scope of this work.

There are a few pathological edge case irreps for which the $\hat{Q}$-type QNP gates are not universal. These cases correspond to quantum number irreps with N_Δ open-shell high-spin orbitals and then all other orbitals wholly occupied or all other orbitals wholly unoccupied. In such cases, the $\hat{Q}_{\mathrm{PX}}$ gates have trivial action, as there are no pairs of orbitals with one orbital unoccupied and one orbital doubly occupied. The remaining $\hat{Q}_{\mathrm{OR}}$ gates (which form a match gate circuit) then have insufficient support to move the high-spin particles into all $\binom{M}{|N_\Delta|}$ possible configurations. Such irreps are rarely of interest in chemical physics. Moreover, universal gate fabrics for these edge case irreps do exist in the form of the extended quantum-number-preserving gate fabrics for $\mathcal{F}(2^{2M})$ in appendix D. For complete details on these edge case irreps, the reader is referred to a more-thorough analysis in appendix G.3.

For $\mathrm{QNP_{OR}}$, $\mathrm{QNP_{PX}}$ (and all other parametrized QNP gates introduced in appendix D) we provide explicit decompositions into elementary gates in appendix E. We further provide generalized parameter shift rules [33–36] for theses gates in appendix F.2, enabling a computation of the gradients with respect to their circuit parameters with a maximum of four distinct circuits and without increasing circuit depth, gate count, or qubit number. We compare this gradient recipe to the generalization presented in [37], extend the variance minimization technique from [38] to the QNP gate gradients and note that our new rule can be applied to a large variety of other gates.

4. Numerical demonstrations

To numerically investigate the properties of the gate fabric from figure 2 and to collect evidence that it satisfies all facets of a 'good' entangler circuit, we consider two prototypical examples of highly correlated molecular ground states: the first is p-benzyne, which exhibits a biradical open-shell singlet ground state, with two unpaired electrons indicated by significant deviations from Hartree–Fock (HF) natural orbital occupation numbers, and four other moderate deviations from HF natural orbital occupation numbers. We use the geometry from [39], build the orbitals at RHF/cc-pVDZ, and construct a (6e, 6o) active space Hamiltonian with the orbitals ranging from HOMO − 2 to LUMO + 2. This corresponds to a case of

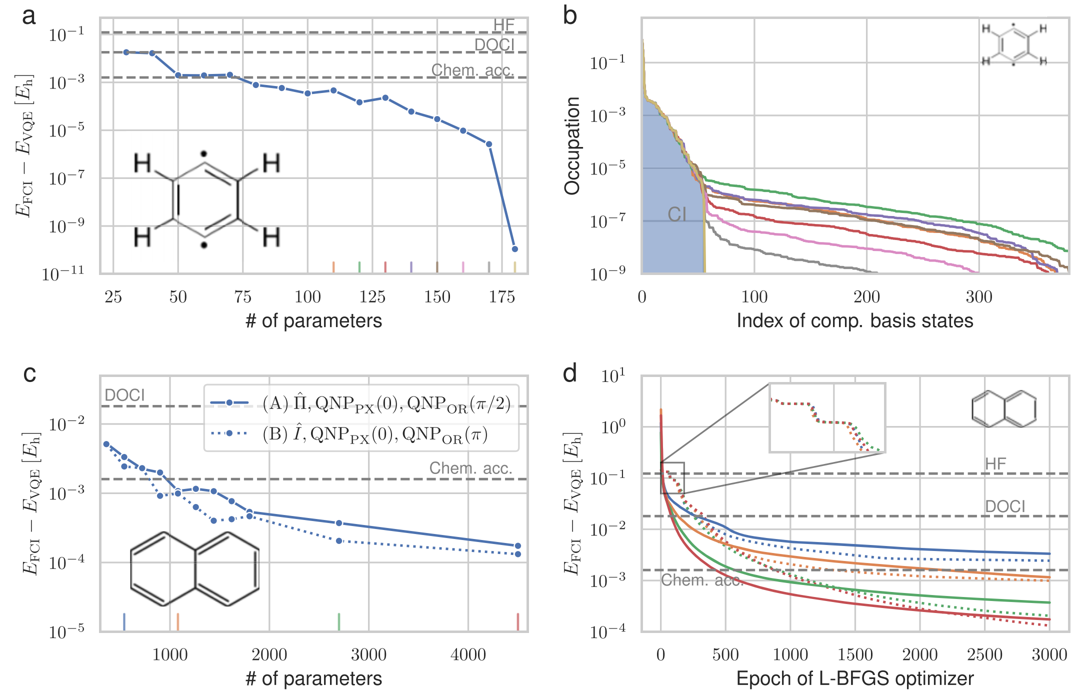


Figure 3. Results of the discussed VQE fabric for representative molecular test cases: (a) convergence of the VQE energy relative to the exact ground state energy E_{FCI} of the 12-spin-orbital active space of *p*-benzyne as a function of the number of parameters in the fabric, (b) occupation $|\langle I|\Psi\rangle|^2$ of each computational basis state $\langle I|$ in the optimized VQE state $|\Psi\rangle$ at the color-indicated parameter counts in (a). Blue area indicates the computational basis states of the full configuration interaction (FCI) ground state in the active space. Each set of computational basis states is sorted in descending order, we show a figure with consistent ordering between sets in see appendix G, (c) convergence of the VQE energy to the exact ground state energy E_{FCI} in the 20-spin-orbital active space of naphthalene as a function of the number of parameters an for the two different initialization schemes, (d) convergence under the L-BFGS optimizer for the color coded parameter counts indicated in (c), dotted (solid) convergence lines correspond to a data point from the dotted (solid) curve in (c), inset highlights the plateaus encountered with initialization method B during the first 180 epochs.

$M = 6$ spatial orbitals (i.e. $N = 12$ qubits), and we focus on the ground state irreducible representation ($N_\alpha = 3, N_\beta = 3, S = 0$). For a larger test case, we consider naphthalene, which while not intrinsically biradical, has multiple natural orbitals with significant deviations from HF natural orbital occupation numbers. We build the orbitals at RHF/STO-3G, and then construct a (10e, 10o) active space Hamiltonian consisting of the π and π^* orbitals. This corresponds to a case of $M = 10$ spatial orbitals (i.e. $N = 20$ qubits), and we focus on the ground state irreducible representation ($N_\alpha = 5, N_\beta = 5, S = 0$). In both cases, we consider VQE gate fabrics of the form of figure 2.

Our final VQE circuit starts with the preparation of an uncorrelated product state by applying local Pauli $\hat{X}$ gates to appropriate qubits of an all-zero state depending on the number of alpha and beta electrons. The qubits are chosen such that for all parameters equal to zero in the following fabric the state is transformed to the state with the energetically lowest orbitals occupied. We then consider two parameter initialization strategies: (A) the fabric is initialized with all $\theta = 0$ and all $\varphi = \pi/2$ and $\hat{\Pi} = \mathrm{QNP}_{\mathrm{OR}}(\pi)$ (solid lines in 3). (B) The fabric is initialized with all $\theta = 0$ and all $\varphi = \pi$ and $\hat{\Pi} = \hat{I}$ (dotted lines in 3).

In both cases we optimize the VQE ground state energy with respect to the VQE gate fabric parameters via L-BFGS. As the purpose of this study is to explore the expressive power of this gate fabric, we consider neither shot noise nor decoherence. This restriction permits the use of analytical expressions for the Hamiltonian expectation values and VQE parameter gradients thereof, greatly accelerating the classical statevector simulation of the VQE.

Figure 3 shows the salient results of this study. For the case of *p*-benzyne, figure 3(a) shows the VQE ground state energy vs FCI with respect to the depth of the gate fabric. The first notable point is that the fabric is able to provide higher accuracy than either HF (a fabric of $\mathrm{QNP}_{\mathrm{OR}}$ gates—redundant here due to the use of HF orbitals in the active space) or doubly-occupied configuration interaction (DOCI) (a fabric of $\mathrm{QNP}_{\mathrm{PX}}$ gates—equivalent to the pUCCD ansatz from [22]). Focusing on the early convergence behavior on the left side of the plot, even with only a few layers of the VQE gate fabric, e.g. $\sim$50–80 parameters, absolute accuracy of 1 kcal mol^{-1} is achieved, which is commonly referred to as chemical accuracy. As the gate fabric depth is increased, roughly geometric (exponential) convergence of the absolute energy is achieved, modulo some minor aberrations due to difficulties in tightly converging the L-BFGS-based numerical optimizations of the VQE gate fabric parameters. Focusing on the later convergence behavior on the right side of the plot, as the number of parameters in the VQE gate fabric approaches the number of parameters in the FCI problem (note that in this irrep there are 175 configuration state functions (CSFs), see appendix C.3), the error convergence turns sharply downward. At 180 parameters we are able to achieve very tight convergence to errors of $\sim 10^{-10} E_{\mathrm{h}}$ relative to FCI, numerically indicating the onset of universality. Figure 3(b) shows the sorted power spectra of the computational basis state (determinant) amplitudes of the various VQE gate fabrics and the FCI state. The exact zeros in the FCI state amplitudes are an artifact of the $D_{2\mathrm{h}}$ spatial point group symmetry of this molecule, which our VQE gate fabric was not optimized to capture. Even for low VQE gate fabric circuit depths, we see that all determinants are populated by nonzero amplitudes, with a compromise apparently being made to allow for some nonzero error in all amplitudes to provide for the best variational energy. As more layers are added to the gate fabrics, the precision of the amplitude spectra increases, as indicated by, e.g. significant attenuation of the symmetry-driven zero block of the amplitude spectrum. The tail of amplitudes exactly zero in FCI is exactly extinguished in the VQE state only when numerical universality is achieved at a 180-parameter VQE gate fabric. This behavior is reminiscent of the nonzero but structured tensor factorized representation of the determinant amplitudes in coupled cluster theories, where here the tensor structure is provided by the local quantum gate fabric.

Moving to the larger test case of naphthalene, figure 3(c) tells a similar story as the corresponding plot for *p*-benzyne. Here we see similar and roughly geometric convergence of energy error vs VQE gate fabric depth and parameter count, albeit with a smaller prefactor. As with *p*-benzyne, the VQE gate fabric rather quickly outstrips both the HF and DOCI ansätze, which its primitive gates are constituted from, and achieves chemical accuracy of ~ 1 kcal mol^{-1} in absolute energy at just $\sim$800–1000 parameters (there are 19 404 CSFs in this irrep, so universality is not reached for any of the depth explored here). Figure 3(d) considers convergence of the energy with respect to the L-BFGS epoch for a number of different VQE gate fabric depths. A first key finding is that making the gate fabric deeper decreases the epoch count needed to converge to chemical accuracy. A second key insight is that, while initialization strategy (B) has shallower circuits and ultimately achieves lower energy error at very high epoch count, plateaus are visible during the optimization with the L-BFGS optimizer (figure 3). Strategy (A), which exchanges orbitals by means of the non trivial choice $\hat{\Pi} = \mathrm{QNP}_{\mathrm{OR}}(\pi)$ and also initializes all $\mathrm{QNP}_{\mathrm{OR}}$ gates with angles of $\pi/2$, appears to circumvent the plateaus entirely and for deeper circuits speeds up (the power-law like) convergence to below chemical accuracy.

The fabric presented here has favorable properties for implementation on NISQ hardware: the 12 qubit ansatz at 110 parameters is without (with) $\hat{\Pi}$ gates decomposable into elementary gates (two-qubit controlled Pauli and one-qubit gates) with resulting depth of 507 (617). The 20 qubit ansatz at 1080 parameters without (with) $\hat{\Pi}$ gates has depth 2761 (3361) in such decomposition. To put this into perspective, a single trotter step of a naive UCCSD circuit has gate depth $\approx$6600 (12 qubits), respectively $\approx$ 57 600 (20 qubits). Another considerable advantage is that only $N-2$ unique four-qubit gates have to be calibrated on hardware as the structure is repetitive after the first two layers.

Simulations were done with the support of PennyLane [36] and OpenFermion [40].

5. Comparison with other entangler circuits

Having numerically demonstrated the features of the VQE gate fabric, it is worth considering the relationship of this gate fabric to other proposed VQE entangler circuits. There has been substantial prior work along these lines in the past few years.

For one instance, the hardware efficient ansatz [19, 20] is manifestly a local gate fabric, using essentially $\mathcal{SU}(4)$ entangler elements or subsets thereof from the native gate set of the underlying quantum circuit architecture. However, this gate fabric does not respect the particle or spin quantum number symmetries, and therefore is likely to encounter substantial difficulties in locating low-lying states within a target quantum number irrep, particularly in larger active spaces.

In another direction, there are myriad proposed entangler circuit constructions which are either already explicitly or in principle could be adapted to real amplitudes and strict commutation with the number and/or spin-squared symmetry operators, but which are either nonlocal circuits or composed of heterogeneous gate layers. For instance, UCCSD [1, 14, 15], (here referring to the Trotterized version thereof) and its sparse and/or low-rank cousins *k*-UpCCGSD [9, 16], ADAPT-VQE [21], and Jastrow-factor VQE [18] all may have the power to achieve universality at sufficient depth, e.g. as proved in a recent analysis of distangled UCC [23] and have been either partially or completely symmetrized already. However, as written, all of these ansätze require nonlocal gate elements that, e.g. mediate excitations among non-adjacent spin orbitals in UCCSD, and thus are not gate fabrics. Moreover, many of these constructions involve heterogeneous gate layers. For a canonical instance, Jastrow-factor VQE [18] involves alternating circuit layers of orbital rotations and substitutions (with the last of these being nonlocal and complex-valued in the usual formulation). Of all methods discussed in the prior literature, k-UpCCGSD is likely closest to our proposed method, with products of single and diagonal double substitution operators comprising the method. k-UpCCGSD as described in [16] involves nonlocal pair substitutions and therefore does not yield a local gate fabric. Note however that the generalized swap network reformulation of k-UpCCGSD described in and around equation (12) and figure 7 of [17] (noticed after the first version of this manuscript was posted) appears to realized k-UpCCGSD by means of a local circuit composed of four-qubit gates that is almost a fabric.

Yet another interesting direction to consider is previously proposed true gate fabrics that preserve quantum number symmetry, but do not achieve universality. Orbital rotation fabrics [41–43], (i.e. HF) are clearly one example here, but so too is doubly occupied configuration interaction (DOCI), for which a gate fabric was developed with the pUCCD ansatz [22]. Both of these ansätze have the interesting property that they can be mapped into gate fabrics requiring only M qubits, but both fail to reach FCI universality as the parameter depth is increased.

Another interesting gate fabric construction is the 'gate-efficient ansatz' presented in [24], which presents as a gate fabric that preserves total particle number $\hat{N}_{\alpha} + N_{\beta}$, but does not appear to respect high-spin particle number $\hat{N}_{\alpha} - \hat{N}_{\beta}$ or $\hat{S}^2$ symmetry. Yet another interesting entangler is the 'qubit coupled cluster' approach presented in [25], which essentially implements a partial spin adaption of UCCSD to preserve $\hat{N}_{\alpha}$ and $\hat{N}_{\beta}$ symmetry within the single and double excitation operations, but neither preserves $\hat{S}^2$ symmetry nor attains the structure of a local gate fabric. Another related approach presented in [26] constructs fermion-adapted excitation operators which preserve particle number symmetry but not spin symmetry, and additionally are aimed at optimizing the number of CNOT gates in non-gate-fabric UCCSD methods. An entangler that has the potential to preserve all quantum number symmetries with additional spin-adaption work is the quantum approximate optimization algorithm (QAOA)-inspired Pauli-term approach, presented in [27], but this approach yields highly nonlocal circuits which do not resemble gate fabrics. Another approach which has some intersection with the present work is the correlating antisymmetric geminal power (AGP) approach explored in [28], which first implements classically-tractable APG to provide state preparation in quantum circuits and then augments AGP with an anti-Hermitian pair hopping entangler which resembles our pair exchange gate. However, the correlating AGP is not written in the form of a local gate fabric.

Another interesting direction to explore that we propose here is alternative local gate fabrics that fully preserve quantum number symmetry, but which exhibit different gate constructions than the $\mathrm{QNP}_{\mathrm{PX}}$ and $\mathrm{QNP}_{\mathrm{OR}}$ gates used in figure 2. Examples of such gate fabrics using generic four-qubit five-parameter FCI gates and decompositions of these gates into $\mathrm{QNP}_{\mathrm{PX}}$, $\mathrm{QNP}_{\mathrm{OR}}$, one-hole/particle substitution, and pair-break up/down gates are described in appendix D.

One additional interesting direction is the 'symmetry preserving state preparation circuits' of [13]. This work primarily focuses on total number symmetry, but does introduce a four-qubit gate that preserves particle and spin quantum numbers via a hyperspherical parametrization.

Note that an alternative approach to the exact symmetry preservation explored in this manuscript is symmetry projection [44, 45], which often requires ancilla qubits and extra measurements due to the necessarily non-unitary nature of the projection operation.

6. Discussion

It is important to note that the main focus of this paper is the existence and performance of the QNP gate fabrics in the limit of an ideal quantum computer with infinite shot resolution. One particular topic that we defer for future consideration is the performance of methods using the QNP gate fabrics in the presence of realistic shot and/or decoherence noise channels. Here, one key concern to be studied is if one encounters the potential for variational collapse to a state with incorrect symmetry through decoherence noise channels (note that we believe that variational collapse through shot noise channels is overwhelmingly unlikely due to the non-systematic nature of this noise channel). We hypothesize that this is unlikely to be a practical issue for three reasons: (1) the forces driving a parameter search from the correct number manifold to the incorrect number manifold will be at most proportional to the decoherence noise strength, which must be fairly small for practical computation to be carried out. (2) The decoherence channels are not directly parametrized by the variational parameters of the QNP circuit, i.e. it is unlikely for the variational parameters of the decoherence-including circuit to be able to provide enough support for complete transport of population from one number manifold to another. (3) It is likely that the contaminants from incorrect quantum number blocks can be removed by efficient quantum error mitigation strategies, as was recently demonstrated for the N_T (and trivially extendible to N_α/N_β) on a hardware implementation of HF [46]. Note that it is not yet clear how (3) could be implemented for the spin-squared operator. A promising result along the lines of this discussion was recently found in simulations and hardware deployment of the N_α/N_β-preserving 'ASWAP' ansatz (essentially orbital rotations, which are universal for certain quantum number irreps of the H_2/STO-3G test system) [47]. Here empirical evidence was found of noise reduction through the use of a symmetry-preserving ansatz, and no variational collapses were encountered despite the targeting of remarkably high-energy states such as the $N_\alpha = 2$, $N_\beta = 2$ double-anion state of H_2/STO-3G.

On the topic of initialization strategies: a different strategy for mitigating barren plateaus was recently presented in the literature [48], and works by creating blocks of identity operators in the ansatz to limit the effective depth and reduce the issue of vanishing gradients. Our strategy instead draws from the idea that our fabric fully exchanges populations at initialised parameters of π. If we initialize halfway at $\pi/2$, all of the possible arms or light cones are populated in the beginning and have a contribution to the final gradient, which seems to help in the optimisations we show here. In the future, it might be possible to merge the ideas from these two strategies into an improved approach.

It is also worth pointing out that there is an apparent asymmetry in the implementation complexity of the $\hat{Q}_\mathrm{OR}$ gate (4× CNOTs) and the $\hat{Q}_\mathrm{PX}$ gate (13× CNOTs). This suggests that it might be advantageous to apply a $\hat{Q}_\mathrm{OR}$ gate on both sides of each $\hat{Q}_\mathrm{PX}$ gate to obtain a circuit with more parameters and fewer total CNOT gates. It is also worth considering that an overhead of 4–13 CNOT gates per effective VQE parameter is decidedly higher than, e.g. the overhead of the hardware efficient ansatz. Therefore, it is highly likely that the hardware efficient ansatz and/or an intermediate ansatz that only preserves $\hat{S}_z$ symmetry might be more accurate on the extremely small experiments allowable on today's quantum hardware. i.e. in the limit of high decoherence noise, the longer circuits of the QNP ansatz may accrue more error than that from the loss of symmetry with a simple ansatz. A related finding was observed on IBM hardware experiments for LiH/STO-3G using the symmetry-preserving state preparation circuits, in which it was found that decoherence errors overwhelmed the proper $\hat{S}^2$-preserving circuits, while circuits the only preserved N_α and N_β fared better [13]. Our opinion is that the QNP gate fabrics developed herein will likely become much more important in intermediate timescales as decoherence error rates diminish and tractable problem size increases.

It is also worth considering how the gate fabric concept could be extended to other types of physical symmetries than the distinguishable, Hamming-weight constrained, or spin-$1/2$ fermionic cases encountered here, e.g. for higher-spin fermions, bosons, nucleons, or elementary particles. For instance, one could consider the case of systems composed of higher-spin fermions. Here each single particle orbital would require effective representation with more than one qubit, as is often done in qudit-to-qubit mappings. Next, the relevant composition operators would have to be mapped to qubits in analogy with the Jordan–Wigner mapping. Finally, localized gate fabrics that preserve the symmetry operators but allow for universal exploration of the Hilbert space would have to be developed. Such gate fabrics would almost surely require larger gate elements than the four-qubit gates encountered here. Similar extensions are likely possible also for bosons and for systems with additional quantum numbers such as isospin, strangeness, charm, etc. The conceptual pathway to obtaining such gate fabrics would likely be similar to the present manuscript, but it remains to be seen if such extended gate fabrics are possible.

7. Summary and outlook

In this work, we set out to construct doppelgängers of the well known gate fabric (i.e. a potentially infinitely repeatable, simple and geometrically local pattern of gate elements that span the parent group at sufficient depth) for the unrestricted qubit Hilbert space $\mathcal{SU}(2^N)$ consisting of simple two-qubit gate elements $SU(4)$. Our major result is the construction of a gate fabric for the important special case of spin-1/2 fermionic systems in the Jordan–Wigner representation $\mathcal{F}(2^{2M})$ consisting of simple four-qubit gate elements $\hat{Q}$. Each two-parameter $\hat{Q}$ gate comprises a one-parameter spatial orbital rotation gate $\mathrm{QNP}_\mathrm{OR}(\varphi)$ and a one-parameter diagonal pair exchange gate $\mathrm{QNP}_\mathrm{PX}(\theta)$. A fabric made of either of these gate elements alone does not achieve FCI universality with sufficient parameter depth, but our VQE gate fabric, being an amalgamation of the two appears to be able to do so. Moreover, at intermediate depths, the VQE gate fabric appears to be pragmatically expressive as evidenced by tests of the ground state energy convergence in strongly correlated molecular systems. It is worth emphasizing that these properties seem to hold in the vast bulk of quantum number irreps, i.e. that these fabric circuits can be applied for cases where $S \neq 0$ (including even or odd spin cases) and/or where $N_\alpha \neq N_\beta$ (see appendix G for details on specific high-spin edge cases that are not universal with the $\hat{Q}$-type QNP gate fabrics of the main text, but that can be addressed with elements of the $\hat{F}$-type QNP gate fabrics of appendix D). Many important questions remain regarding our QNP gate fabrics. These include: (1) how does the numerical optimization of parameters for such gate fabrics behave in the presence of shot and/or decoherence noise? (2) How can numerical optimization algorithms be adapted to exploit the knowledge that the VQE entangler circuit is a gate fabric? (3) Is the fixed $\hat{\Pi}$ gate construction or an extension thereof an effective way to mitigate barren plateaus during numerical optimization? (4) How does the VQE gate fabric perform for relative properties, for properties at different nuclear geometries, and for properties in different quantum number irreps? (5) Is the construction of the VQE gate fabric in terms of $\hat{Q}$ gates optimal, or do more elaborate constructions, e.g. using the $\hat{F}$ gates of appendix D provide additional benefits? (6) What is the scaling behavior of the error in absolute and/or relative properties as a function of parameter depth for representative interesting molecular systems? (7) Can the gate fabric be adapted to additionally exploit external symmetries such as spatial point group symmetries, e.g. as explored in [49]? Taken together, the results of this work might provide an interesting guide for the required symmetries and limiting simplicities when constructing more elaborate VQE entanglers for fermionic systems.

Acknowledgments

RMP is grateful to Dr Edward Hohenstein for many discussions on the structure of the $\hat{S}^2$ operator. The authors further thank Fotios Gkritsis for discussions and acknowledge useful comments from Will Simmons, Seyon Sivarajah and David Ramo that have led to a reduction of depth of the decompositions of our QNP gates. QC Ware Corp. acknowledges generous research funding from Covestro Deutschland AG for this project. Covestro acknowledges funding from the German Ministry for Education and Research (BMBF) under the funding program quantum technologies as part of project HFAK (13N15630). DW acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC2004/1 390534769.

Conflict of interest

The VQE gate fabrics described in this work are elements of two US provisional applications for patents both filed jointly by QC Ware Corp. and Covestro Deutschland AG. RMP owns stock/options in QC Ware Corp.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. Symmetry-constrained subgroups of $\mathcal{SU}(2^N)$

This section discusses some of the technical hurdles encountered in developing universal gate fabrics for certain subgroups of $\mathcal{SU}(2^N)$, using well-known literature results for restriction to real operators $\mathcal{SO}(2^N)$ and further restriction to Hamming-weight-preserving operators $\mathcal{H}(2^N)$.

The imposition of specific symmetries which constrain the subgroup of $\mathcal{SU}(2^N)$ may or may not present considerable difficulties in constructing gate fabrics of the type defined above. For an example that does not introduce significant difficulty, consider the case where we restrict the Hilbert space operators to have real value, i.e. a restriction to $\mathcal{SO}(2^N)$. In this case, one may simply substitute $\mathcal{SU}(4) \rightarrow \mathcal{SO}(4)$ in the gate fabric of figure 1 to construct the desired gate fabric sketched in figure 4 for $\mathcal{SO}(2^N)$.

For an example that does introduce significant difficulty, consider the case where we restrict $\mathcal{SO}(2^N)$ Hilbert space operators to preserve Hamming weight, i.e. to respect the commutation constraint $[\hat{U}, \hat{P}] = 0$ where $\hat{P} := \sum_p (\hat{I} - \hat{Z}_p)/2$ is the Hamming weight or 'particle counting' operator. We denote this subgroup as $\mathcal{H}(2^N)$. Here, we might be tempted to continue restricting the two-qubit $\mathcal{SO}(4)$ gates to preserve Hamming weight, mandating that we substitute $\mathcal{SO}(4) \rightarrow \mathcal{H}(4)$, where $\mathcal{H}(4)$ implements a Givens rotation between configurations $|01\rangle$ and $|10\rangle$ while acting as the identity in $|00\rangle$ and $|11\rangle$. This is sketched in figure 5. However, a tessellation of two-qubit Givens gates is not a gate fabric for $\mathcal{H}(2^N)$, as it does not provide universality for this subgroup. In fact, it can be shown that a tessellation of Givens gates amounts to a one-particle rotation of the qubit creation and annihilation operators $\hat{p}^{\pm} := \sum_q V_{qp} \hat{q}^{\pm}$ for $V_{qp} \in \mathcal{SO}(N)$

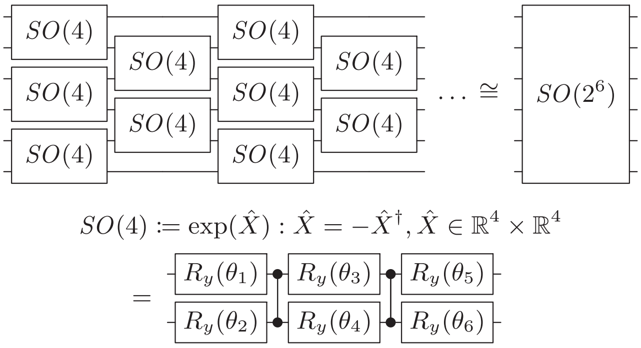


Figure 4. Gate fabric universal for $\mathcal{SO}(2^N)$ (sketched for $N = 6$). The gate fabric is a two-local-nearest-neighbor tessellation of alternating even and odd qubit-pair six-parameter, four-qubit $SO(4)$ gates.

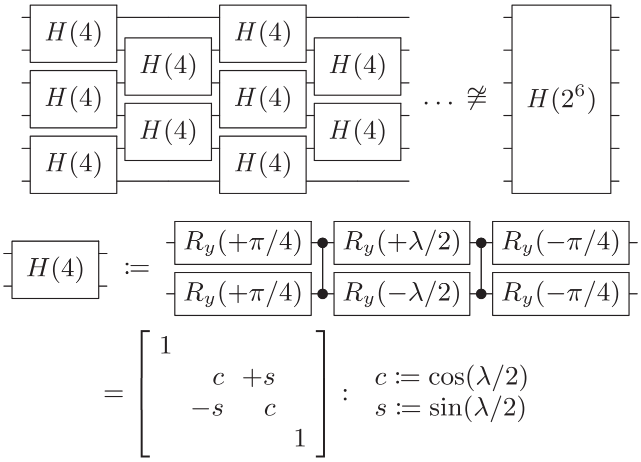


Figure 5. Gate fabric attempt *not* universal for the Hamming-weight-preserving subgroup $\mathcal{H}(2^N)$ (sketched for $N = 6$). The gate fabric is a two-local-nearest-neighbor tessellation of alternating even and odd qubit-pair one-parameter, two-qubit Hamming-weight-preserving $\hat{H}(4)$ gates. The gate fabric exactly commutes with the Hamming weight operator $\hat{P} \equiv \sum_p (\hat{I} - \hat{Z}_p)/2$, but the gate fabric does not span $\mathcal{H}(2^N)$ for any depth.

and $\hat{q}^\pm := (\hat{X}_q \mp \mathrm{i}\hat{Y}_q)/2$, and thus after exactly N layers and $N(N-1)/2$ gates the part of Hilbert space reachable with the fabric does no longer increase anymore, and in fact the fabric is classically simulable in polynomial time via techniques such as the match gate formalism or direct implementation with classical photons and beamsplitters. Note that $\mathcal{H}(2^N)$ has irreducible representations of dimension up to $\binom{N}{\lfloor N/2 \rfloor}$, so failure to reach universality can be shown by simply parameter counting. Speaking more practically, this proposed gate fabric has very limited expressive power for most irreps of $\mathcal{H}(2^N)$, and does not provide a good approximation to most desired actions within this space.

In fact, no gate fabric for $\mathcal{H}(2^N)$ is possible with two-qubit gate elements. One possible construction of a gate fabric for $\mathcal{H}(2^N)$ with three-qubit gate elements is sketched in figure 6. Note that this might not be a minimal representation—we will see examples of fermionic systems shortly where a much simpler gate element than the fully explicitly universal k-minimal qubit gate provides a gate fabric.

Appendix B. Additional details on the Jordan–Wigner mapping

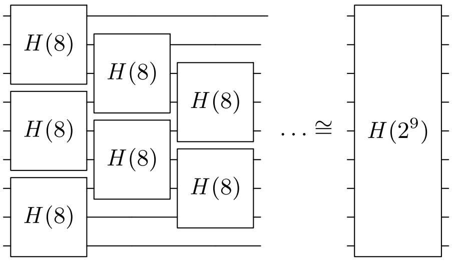
This section is included to enumerate the expansion of same-spin number operators and the $\hat{S}^2$ operator into Pauli operators in the Jordan–Wigner mapping defined in the main text.

B.1. Same-spin occupation and substitution operators

B.1.1. Same-spin occupation number operators

The same-spin occupation number operators are

$$p^\pm p^\mp = (\hat{I}_p \mp \hat{Z}_p)/2, \tag{B1}$$



$$H(8) := \begin{bmatrix} 1 & & & & & & & \\ & u^1_{00} & u^1_{01} & & u^1_{02} & & & \\ & u^1_{10} & u^1_{11} & & u^1_{12} & & & \\ & & & u^2_{00} & & u^2_{01} & u^2_{02} & \\ & u^1_{20} & u^1_{21} & & u^1_{22} & & & \\ & & & u^2_{10} & & u^2_{11} & u^2_{12} & \\ & & & u^2_{20} & & u^2_{21} & u^2_{22} & \\ & & & & & & & 1 \end{bmatrix}$$

$$\hat{u}^d := \exp(\hat{x}^d) \in \mathcal{SO}(3) : \ \hat{x}^{d\dagger} = -\hat{x}^d : \hat{x}^d \in \mathbb{R}^3 \times \mathbb{R}^3, \ d \in [1, 2]$$

Figure 6. Gate fabric universal for the Hamming-weight-preserving subgroup $\mathcal{H}(2^N)$ (sketched for $N = 9$). The gate fabric is a three-local-nearest-neighbor tessellation of cascading qubit-triple six-parameter, three-qubit Hamming-weight-preserving $\hat{H}(8)$ gates. Each $\hat{H}(8)$ gate is composed of a three-parameter $\mathcal{SO}(3)$ rotation in the d-Hamming-weight subspace, where $d \in [1, 2]$ for a total of 6 parameters. The gate fabric exactly commutes with the Hamming weight operator $\hat{P} \equiv \sum_p (\hat{I} - \hat{Z}_p)/2$ and spans $\mathcal{H}(2^N)$ at sufficient depth.

whereby p^+p^- is a 'particle occupation number operator' (counts 1s). p^-p^+ is a 'hole occupation number operator' (counts 0s). Note that the Jordan–Wigner strings cancel for these operators.

B.1.2. Same-spin substitution operators

The same-spin one-particle substitution operator is

$$p^{\pm}q^{\mp}\big|_{p<q} = \pm\hat{P}^{\pm}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{P}^{\mp}_q. \tag{B2}$$

Here $Z^{\leftrightarrow}_{p+1,q-1} := \bigotimes_{r=p+1}^{r=q-1} \hat{Z}_r$. For completeness

$$p^{\pm}q^{\mp}\big|_{p>q} = \pm\hat{P}^{\mp}_q \bigotimes \hat{Z}^{\leftrightarrow}_{q+1,p-1} \bigotimes \hat{P}^{\pm}_p. \tag{B3}$$

Technically, p^+q^- is the 'one-particle substitution operator' and p^-q^+ is the 'one-hole substitution operator.'

With some algebra, one can show that,

$$p^{\pm}q^{\mp} + q^{\pm}p^{\mp}\big|_{p<q} \tag{B4}$$

$$= \pm\hat{X}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{X}_q/2 \pm \hat{Y}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{Y}_q/2$$

(the formula for $p > q$ is the same except for the indices on $Z^{\leftrightarrow}_{p+1,q-1}$.)

Here, the Jordan–Wigner strings only cancel partially. However, the α-then-β Jordan–Wigner ordering does provide the advantage that the remaining $Z^{\leftrightarrow}_{p,q}$ strings are supported only on the intermediate α (β) spin orbital indices for α (β) substitution operators.

B.2. Quantum number operators

B.2.1. Alpha number operator

The α number operator is

$$\hat{N}_{\alpha} := \sum_p p^{\dagger}p = (M/2)\hat{I} - \sum_p \hat{Z}_p/2. \tag{B5}$$

The eigenvalues of the α number operator are $N_{\alpha} \in [0, 1, \ldots M]$ with degeneracy $\binom{M}{N_{\alpha}} 2^M$. The determinants are eigenfunctions of the α number operator, with eigenvalues given by the α population count,

$$\hat{N}_{\alpha}|\vec{I}\rangle = \text{popcount}(\vec{I}_{\alpha})|\vec{I}\rangle = N_{\alpha}|\vec{I}\rangle \tag{B6}$$

and popcount$(\vec{I}_{\alpha})$ counts the number of ones in $\vec{I}_{\alpha}$.

B.2.2. Beta number operator

The β number operator is,

$$\hat{N}_\beta := \sum_p \bar{p}^\dagger \bar{p} = (M/2)\hat{I} - \sum_{\bar{p}} \hat{Z}_{\bar{p}}/2. \tag{B7}$$

The eigenvalues of the β number operator are $N_\beta \in [0, 1, \ldots M]$ with degeneracy $\binom{M}{N_\beta} 2^M$. The dets are eigenfunctions of the β number operator, with eigenvalues given by the β population count,

$$\hat{N}_\beta |\vec{I}\rangle = \mathrm{popcount}(\vec{I}_\beta)|\vec{I}\rangle = N_\beta |\vec{I}\rangle. \tag{B8}$$

B.2.3. Total spin squared operator

The total spin squared operator is,

$$\hat{S}^2 = \hat{S}_- \hat{S}_+ + \hat{S}_z + \hat{S}_z^2 \tag{B9}$$

with the spin lowering operator,

$$\hat{S}_- := \sum_p \bar{p}^\dagger p \tag{B10}$$

the spin raising operator,

$$\hat{S}_+ := \sum_p p^\dagger \bar{p} \tag{B11}$$

and the z-spin,

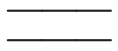
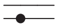
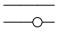
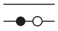
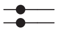
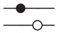
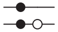
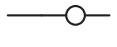
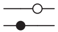
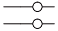
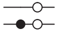
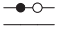
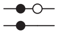
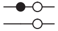
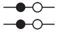
$$\hat{S}_z := \frac{1}{2}\left[\hat{N}_\alpha - \hat{N}_\beta\right]. \tag{B12}$$

After some algebra, under the chosen Jordan Wigner mapping, this resolves to

$$\begin{aligned}
\hat{S}^2 = \frac{3M}{8}\hat{I} &- \frac{3}{8}\sum_p \hat{Z}_p \bigotimes \hat{Z}_{\bar{p}} + \sum_{p<q}\frac{1}{8}\hat{Z}_p \bigotimes \hat{Z}_q \\
&+ \sum_{p<q}\frac{1}{8}\hat{Z}_{\bar{p}} \bigotimes \hat{Z}_{\bar{q}} - \sum_{p<q}\frac{1}{8}\hat{Z}_p \bigotimes \hat{Z}_{\bar{q}} - \sum_{p<q}\frac{1}{8}\hat{Z}_{\bar{p}} \bigotimes \hat{Z}_q \\
&- \frac{1}{8}\sum_{p<q}\hat{X}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{X}_q \bigotimes \hat{X}_{\bar{p}} \bigotimes \hat{Z}^{\leftrightarrow}_{\bar{p}+1,\bar{q}-1} \bigotimes \hat{X}_{\bar{q}} \\
&- \frac{1}{8}\sum_{p<q}\hat{X}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{X}_q \bigotimes \hat{Y}_{\bar{p}} \bigotimes \hat{Z}^{\leftrightarrow}_{\bar{p}+1,\bar{q}-1} \bigotimes \hat{Y}_{\bar{q}} \\
&- \frac{1}{8}\sum_{p<q}\hat{X}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{Y}_q \bigotimes \hat{X}_{\bar{p}} \bigotimes \hat{Z}^{\leftrightarrow}_{\bar{p}+1,\bar{q}-1} \bigotimes \hat{Y}_{\bar{q}} \\
&+ \frac{1}{8}\sum_{p<q}\hat{X}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{Y}_q \bigotimes \hat{Y}_{\bar{p}} \bigotimes \hat{Z}^{\leftrightarrow}_{\bar{p}+1,\bar{q}-1} \bigotimes \hat{X}_{\bar{q}} \\
&+ \frac{1}{8}\sum_{p<q}\hat{Y}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{X}_q \bigotimes \hat{X}_{\bar{p}} \bigotimes \hat{Z}^{\leftrightarrow}_{\bar{p}+1,\bar{q}-1} \bigotimes \hat{Y}_{\bar{q}} \\
&- \frac{1}{8}\sum_{p<q}\hat{Y}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{X}_q \bigotimes \hat{Y}_{\bar{p}} \bigotimes \hat{Z}^{\leftrightarrow}_{\bar{p}+1,\bar{q}-1} \bigotimes \hat{X}_{\bar{q}} \\
&- \frac{1}{8}\sum_{p<q}\hat{Y}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{Y}_q \bigotimes \hat{X}_{\bar{p}} \bigotimes \hat{Z}^{\leftrightarrow}_{\bar{p}+1,\bar{q}-1} \bigotimes \hat{X}_{\bar{q}} \\
&- \frac{1}{8}\sum_{p<q}\hat{Y}_p \bigotimes \hat{Z}^{\leftrightarrow}_{p+1,q-1} \bigotimes \hat{Y}_q \bigotimes \hat{Y}_{\bar{p}} \bigotimes \hat{Z}^{\leftrightarrow}_{\bar{p}+1,\bar{q}-1} \bigotimes \hat{Y}_{\bar{q}}.
\end{aligned} \tag{B13}$$

The eigenvalues of the $\hat{S}^2$ operator can be written as $S/2(S/2+1)$ with $S \in [0, 1, 2, \ldots]$ (singlet, doublet, triplet, etc).

Table 1. Enumeration of characteristics of $M = 2$ Fock space in Jordan–Wigner representation. First column: base-2 qubit occupation string (i.e. qubit computational basis state). Second column: base-10 qubit occupation string (i.e. base-10 index for vector and matrix quantities). Third column: Slater determinantal configuration represented by this qubit computational basis state. Fourth column: number of α electrons in this configuration (always a proper eigenstate). Fifth column: number of β electrons in this configuration (always a proper eigenstate). Sixth column: is this configuration a valid CSF, i.e. a proper eigenstate of $\hat{S}^2$? Seventh column: if yes to the previous question, S eigenvalue for this simultaneous Slater determinant/CSF ($S = 0$—singlet, $S = 1$—doublet, $S = 2$—triplet, …).

Base 2	Base 10	Determinant	N_α	N_β	Is CSF?	S
$\|0000\rangle$	$\|\#0\rangle$		0	0	Y	0
$\|0001\rangle$	$\|\#1\rangle$		1	0	Y	1
$\|0010\rangle$	$\|\#2\rangle$		0	1	Y	1
$\|0011\rangle$	$\|\#3\rangle$		1	1	Y	0
$\|0100\rangle$	$\|\#4\rangle$		1	0	Y	1
$\|0101\rangle$	$\|\#5\rangle$		2	0	Y	2
$\|0110\rangle$	$\|\#6\rangle$		1	1	N	
$\|0111\rangle$	$\|\#7\rangle$		2	1	Y	1
$\|1000\rangle$	$\|\#8\rangle$		0	1	Y	1
$\|1001\rangle$	$\|\#9\rangle$		1	1	N	
$\|1010\rangle$	$\|\#10\rangle$		0	2	Y	2
$\|1011\rangle$	$\|\#11\rangle$		1	2	Y	1
$\|1100\rangle$	$\|\#12\rangle$		1	1	Y	0
$\|1101\rangle$	$\|\#13\rangle$		2	1	Y	1
$\|1110\rangle$	$\|\#14\rangle$		1	2	Y	1
$\|1111\rangle$	$\|\#15\rangle$		2	2	Y	0

Appendix C. The specific case of $M = 2$ in $\mathcal{F}(2^{2M})$

This section explicitly enumerates, for the special case of $M = 2$, the characteristics of the Jordan–Wigner computational basis functions (representing Fock space Slater determinants), S^2-pure CSF linear combinations thereof, and arbitrary linear combinations thereof which we will refer to as FCI states. This section is useful to develop the beginnings of a picture for the arbitrary M-spatial-orbital Fock space, and also directly leads to the FCI gate fabric of the next full section.

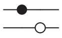
C.1. Slater determinants/Jordan–Wigner computational basis states

Table 1 enumerates the Slater determinants in the $M = 2$ case along with their corresponding Jordan–Wigner computational basis states, provides the N_α and N_β eigenvalues of each determinant (all determinants are proper eigenstates of the particle number operators), identifies which determinants are CSFs (only some determinants are also CSFs), and if a CSF, provides the S eigenvalue of the determinant/CSF.

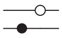
C.2. Quantum number operators

The $\hat{N}_\alpha$ and $\hat{N}_\beta$ operators are diagonal (as is always true in the Jordan–Wigner representation), and their diagonal values are depicted in the fourth and fifth columns of table 1, respectively.

The $\hat{S}^2$ operator is not diagonal (as is generally true in the Jordan–Winger representation). Instead, only 14 out of the 16 rows/columns of this operator are diagonal, and their diagonal entries are given in the seventh column of table 1. The non-diagonal contributions arise from the non-CSF determinants

$|0110\rangle \equiv$  $\equiv |\#6\rangle$

and,

$|1001\rangle \equiv$  $\equiv |\#9\rangle$

These two determinants form the seniority-2 coupling set (the set of determinants with 2× non-spin-paired electrons). In this restricted basis, the $\hat{S}^2$ operator is,

$$\hat{S}^2 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Table 2. CSF irreps for $M = 2$ Fock space. D refers to the irrep dimension. The listed elements are our particular convention for the CSF basis functions of each irrep.

N_α	N_β	S	D	Elements
0	0	0	1	$\vert\#0\rangle$
2	2	0	1	$\vert\#15\rangle$
2	0	2	1	$\vert\#5\rangle$
0	2	2	1	$\vert\#10\rangle$
1	1	2	1	$\vert\Phi^{S=2}_{Z=2}\rangle$
1	0	1	2	$\vert\#1\rangle$,$\vert\#4\rangle$
0	1	1	2	$\vert\#2\rangle$,$\vert\#8\rangle$
1	2	1	2	$\vert\#11\rangle$,$\vert\#14\rangle$
2	1	1	2	$\vert\#7\rangle$,$\vert\#13\rangle$
1	1	0	3	$\vert\#3\rangle$,$\vert\#12\rangle$,$\vert\Phi^{S=0}_{Z=2}\rangle$

C.3. Configuration state functions (CSFs)

CSFs are defined as sparse linear combinations of Slater determinants that provide proper eigenstates of $\hat{S}^2$. The 14 spin-pure Slater determinants discussed above are also CSFs, with corresponding eigenvalues S.

In the seniority-2 coupling set of the non-spin-pure Slater determinants $|\#6\rangle$ and $|\#9\rangle$, the eigenvectors of the $\hat{S}^2$ operator are,

$$\hat{V} \equiv \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and the corresponding eigenvalues are,

$$\hat{\Lambda} \equiv \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

e.g. the + combination yields an $S = 0$ singlet CSF, while the − combination yields an $S = 2$ triplet CSF.

Thus the symmetry-adapted CSFs for this seniority coupling set are,

$$|\Phi^{S=0}_{Z=2}\rangle \equiv \frac{1}{\sqrt{2}}\left(|0110\rangle + |1001\rangle\right)$$

and,

$$|\Phi^{S=2}_{Z=2}\rangle \equiv \frac{1}{\sqrt{2}}\left(|0110\rangle - |1001\rangle\right).$$

Therefore, we have a complete real, orthonormal set of 16 CSFs for $\mathcal{F}(2^{2*2})$: 5 singlets, 8 doublets, and 3 triplets. These CSFs are proper eigenfunctions of $\hat{N}_\alpha$, $\hat{N}_\beta$, and $\hat{S}^2$.

C.4. Quantum number irreps

Valid solutions to the time-dependent or time-independent Schrödinger equation for spin-1/2 fermions governed by spin-free Hamiltonian operators must be definite simultaneous eigenstates of the quantum number operators $(\hat{N}_\alpha, \hat{N}_\beta, \hat{S}^2)$ with definite target eigenvalues (N_α, N_β, S). We refer to the set of valid simultaneous eigenstates for a given set of target quantum numbers (N_α, N_β, S) as a quantum number irrep.

Table 2 enumerates the dimensionality and our particular convention for the CSF basis for each definite (N_α, N_β, S) irrep of the $M = 2$ Fock space. An arbitrary special orthogonal rotation within each irrep would also provide a faithful representation of the basis for that irrep.

C.5. Full configuration interaction (FCI) states

The restriction of physically valid solutions of the time-dependent or time-independent Schrödinger equation to a given target quantum number irrep severely constrains, but does not exactly determine the valid solution for most irreps. For instance, in the $(N_\alpha = 1, N_\beta = 0, S = 1)$ irrep, the 15-parameter generic solution,

$$|\Psi\rangle \overset{?}{=} \sum_{I=0}^{I=15} c_I|\#I\rangle : \sum_{I=0}^{I=15} |c_I|^2 = 1, \quad c_I \in \mathbb{R}$$

is invalid because it does not respect the quantum number symmetries, but the one-parameter solution,

$$|\Psi\rangle = \sum_{I\in\langle 1,4\rangle} c_I|\#I\rangle : \sum_{I\in\langle 1,4\rangle} |c_I|^2 = 1, \quad c_I \in \mathbb{R}$$

is valid due to the fact that the dimension of the target irrep is $D = 2 > 1$.

We generically refer to states which exactly lie within a given target quantum number irrep, but where the remaining flexibility in the state is determined by solving an auxiliary equation such as the time-dependent or time-independent Schrödinger equation, as 'FCI' states. The motivation for this naming is the set of states that emerge from exactly diagonalizing the spin-free electronic Hamiltonian within a given quantum number irrep, i.e. the classical FCI method, though the usage within this work should be understood to be generalized to solving any linear auxiliary equation governed by a spin-free operator which is simultaneously diagonalized by the three quantum number operators.

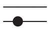
The question that arises at this point is how to construct special orthogonal operators that respect the quantum number symmetry but have the power to move from an arbitrary quantum-number-pure trial state to an FCI state within the same quantum number irrep. The simple answer is to construct complete special orthogonal operators acting on the CSF basis of each irrep, with the property that these operators commute with all three quantum number operators. This leads to the construction of 4× one-parameter $\mathcal{SO}(2)$ operators (simple Givens rotation matrices) acting within the $4 \times S = 1$ doublet irreps, and 1× three-parameter $\mathcal{SO}(3)$ operator acting within the $(N_\alpha = 1, N_\beta = 1, S = 0)$ irrep. This seems to imply that the parameter dimension of $\mathcal{F}(2^{2*2})$ is 7. However, further analysis reveals that to preserve $\hat{S}^2$ symmetry, the same operator must be applied in the $(N_\alpha = 1, N_\beta = 0, S = 1)$ and $(N_\alpha = 0, N_\beta = 1, S = 1)$ irreps and that the same operator must be applied in the $(N_\alpha = 1, N_\beta = 2, S = 1)$ and $(N_\alpha = 2, N_\beta = 1, S = 1)$ irreps. This is related to the fact that the spin-free Schrödinger equation is invariant under permutation of the α and β labels in the working equations. This reduces the total number of parameter of $\mathcal{F}(2^{2*2})$ to 5, and yields the highly structured special orthogonal operators that will be encountered as $M = 2$ FCI gate operators in the next section.

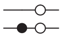
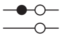
Appendix D. Gate fabric for $\mathcal{F}(2^{2M})$ via $M = 2$ FCI gates

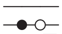
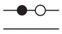
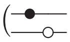
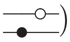
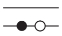
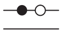
An early iteration of the gate fabric described in the main text was developed by constructing a gate fabric comprising a five-parameter four-qubit $\hat{F}$ gate universal for $M = 2$ FCI as detailed in figure 7. A fabric of these $\hat{F}$ gates was found to exactly preserve quantum number symmetry, to provide universality for $\mathcal{F}(2^{2M})$ for sufficient parameter depth, and to yield an expressive approximate representation at intermediate depths. The representation power and numerical convergence was found to be similar between $\hat{F}$ gate fabrics and the $\hat{Q}$ gate fabrics, and the latter is conceptually simpler, so we have elected to focus on the latter in the main text. In the following we describe this alternative gate fabric and additional variants and refer to and use the concepts and notation introduced in appendices B and C.

D.1. Decomposition of $\hat{F}$ into simple gate elements

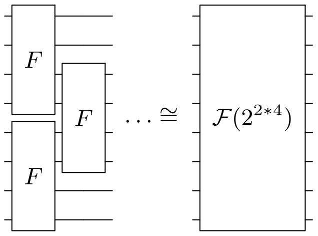
There are many different possible implementations of $\hat{F}$ into products of simpler (e.g. one-parameter) gate elements. However, the block diagonal nature and configuration constituency of $\hat{F}$ suggests the following pragmatic choice, leading to a decomposition with a simple decomposition all the way down to a standard two-qubit gate library. The text refers to the $\hat{F}$ gate matrix in the second line of figure 7.

The red block (matrix entries $c_{1p} := \cos(\theta_{1p}/2)$ and $s_{1p} := \sin(\theta_{1p}/2)$) corresponds to a Givens rotation between the one particle $(N_\alpha = 1, N_\beta = 0, S = 1)$ CSFs $|0001\rangle =:$  $= |\#1\rangle$ and $|0100\rangle =:$  $= |\#4\rangle$ and the same Givens rotation between the $(N_\alpha = 0, N_\beta = 1, S = 1)$ CSFs $|\#2\rangle$ and $|\#8\rangle$ (to preserve $\hat{S}^2$ symmetry). We call this operation the $\mathrm{QNP_{1p}}$ gate (quantum-number-preserving one-particle gate).

The blue block (matrix entries $c_{1h} := \cos(\theta_{1h}/2)$ and $s_{1h} := \sin(\theta_{1h}/2)$) implements a Givens rotation between the one hole $(N_\alpha = 1, N_\beta = 2, S = 1)$ CSFs  $= |\#11\rangle$ and $|1110\rangle :=$  $= |\#14\rangle$ and the same Givens rotation between the $(N_\alpha = 2, N_\beta = 1, S = 1)$ CSFs $|\#7\rangle$ and $|\#13\rangle$ (to preserve $\hat{S}^2$ symmetry). We call this operation the $\mathrm{QNP_{1h}}(\theta_{1h})$ gate (quantum-number-preserving one-hole gate).

The green block implements an $\mathcal{SO}(3)$ rotation between the three $(N_\alpha = 1, N_\beta = 1, S = 0)$ CSFs, $|0011\rangle :=$  $=: |\#3\rangle, |1100\rangle =:$  $= |\#12\rangle$, and $(|0110\rangle + |1001\rangle)/\sqrt{2} =:$ ( + ) $/\sqrt{2}$ =. There are three natural rotation gates (i.e. Euler-angle-like rotation gates) in this subspace: first, the $\mathrm{QNP_{PX}}$ gate (quantum-number-preserving pair exchange gate) implements a Givens rotation between the two closed shell CSFs $|0011\rangle =:$  $:= |3\rangle$ and $|1100\rangle =:$  $= |\#12\rangle$. Second and third the $\mathrm{QNP_{PBU}}$ and $\mathrm{QNP_{PBL}}$ (QNP pair-break upper/lower gates) rotate between the upper, respectively, lower closed shell CSF and the open-shell singlet CSF $(|0110\rangle + |1001\rangle)/\sqrt{2}$.

Explicit decompositions of the five gates $\mathrm{QNP_{1p}}$, $\mathrm{QNP_{1h}}$, $\mathrm{QNP_{PX}}$, $\mathrm{QNP_{PBU}}$, and $\mathrm{QNP_{PBL}}$ to elementary two-qubit gate operations are provided in appendix E.



$$\boxed{F} := \begin{bmatrix}
1 &&&&&&&&&&&&&&& \\
& \color{red}c_{1\mathrm{p}} &&& \color{red}s_{1\mathrm{p}} &&&&&&&&&&& \\
&& \color{red}c_{1\mathrm{p}} &&&&&& \color{red}s_{1\mathrm{p}} &&&&&&& \\
&&& \color{green}u_{00} &&& \color{green}u_{01} &&& \color{green}u_{02} &&& \color{green}u_{03} &&& \\
& \color{red}-s_{1\mathrm{p}} &&& \color{red}c_{1\mathrm{p}} &&&&&&&&&&& \\
&&&&& 1 &&&&&&&&&& \\
&&& \color{green}u_{10} &&& \color{green}u_{11} &&& \color{green}u_{12} &&& \color{green}u_{13} &&& \\
&&&&&&& \color{blue}c_{1\mathrm{h}} &&&&&& \color{blue}+s_{1\mathrm{h}} && \\
&& \color{red}-s_{1\mathrm{p}} &&&&&& \color{red}c_{1\mathrm{p}} &&&&&&& \\
&&& \color{green}u_{20} &&& \color{green}u_{21} &&& \color{green}u_{22} &&& \color{green}u_{23} &&& \\
&&&&&&&&&& 1 &&&&& \\
&&&&&&&&&&& \color{blue}c_{1\mathrm{h}} &&& \color{blue}+s_{1\mathrm{h}} & \\
&&& \color{green}u_{30} &&& \color{green}u_{31} &&& \color{green}u_{32} &&& \color{green}u_{33} &&& \\
&&&&&&& \color{blue}-s_{1\mathrm{h}} &&&&&& \color{blue}c_{1\mathrm{h}} && \\
&&&&&&&&&&& \color{blue}-s_{1\mathrm{h}} &&& \color{blue}c_{1\mathrm{h}} & \\
&&&&&&&&&&&&&&& 1
\end{bmatrix} \in \mathcal{F}(2^{2*2})$$

$$\begin{bmatrix} u_{00} & u_{01} & u_{02} & u_{03} \\ u_{10} & u_{11} & u_{12} & u_{13} \\ u_{20} & u_{21} & u_{22} & u_{23} \\ u_{30} & u_{31} & u_{32} & u_{33} \end{bmatrix} := \begin{bmatrix} 1 &&& \\ &&& 1 \\ & 1/\sqrt{2} & 1/\sqrt{2} & \\ & 1/\sqrt{2} & -1/\sqrt{2} & \end{bmatrix}^{\dagger} \begin{bmatrix} v_{00} & v_{01} & v_{02} & \\ v_{10} & v_{11} & v_{12} & \\ v_{20} & v_{21} & v_{22} & \\ &&& 1 \end{bmatrix} \begin{bmatrix} 1 &&& \\ &&& 1 \\ & 1/\sqrt{2} & 1/\sqrt{2} & \\ & 1/\sqrt{2} & -1/\sqrt{2} & \end{bmatrix}$$

$$c_{1\mathrm{p}} := \cos(\theta_{1\mathrm{p}}), \quad s_{1\mathrm{p}} := \sin(\theta_{1\mathrm{p}})$$
$$c_{1\mathrm{h}} := \cos(\theta_{1\mathrm{h}}), \quad s_{1\mathrm{h}} := \sin(\theta_{1\mathrm{h}})$$

$$\hat{v} := \exp(\hat{x}) : \ \hat{x} = -\hat{x}^{\dagger}, \ \hat{x} \in \mathbb{R}^3 \times \mathbb{R}^3$$

Figure 7. Gate fabric hypothesized to be universal for $\mathcal{F}(2^{2M})$ (sketched for $M = 4$). The spin orbitals in Jordan–Wigner representation are physically ordered in 'interleaved' ordering with even (odd) qubit indices denoting α (β) spin orbitals. The Jordan–Wigner strings are defined in 'α-then-β' order as defined in the main text. The gate fabric is a four-local-nearest-neighbor-tessellation of alternating even and odd spatial-orbital-pair five-parameter, four-qubit $\hat{F}$ gates, constructed to be universal for $M = 2$ FCI. Each $\hat{F}$ gate consists of: (A) a spin-adapted one-parameter $\mathcal{SO}(2)$ rotation in the one-particle ($N_\alpha + N_\beta = 1$) $S = 0$ block denoted in red (parameter θ_{1p}). (B) A spin-adapted one-parameter $\mathcal{SO}(2)$ rotation in the one-hole ($N_\alpha + N_\beta = 3$) $S = 0$ block denoted in blue (parameter θ_{1h}). (C) A spin-adapted three-parameter $\mathcal{SO}(3)$ rotation in the ($N_\alpha = 1, N_\beta = 0, S = 0$) irrep denoted in green (3 parameters in the unique upper triangle of $\hat{x}$). The decomposition of $\hat{u}$ into a transformation from the determinant basis of the Jordan–Wigner computational basis to the CSF basis, followed by the application of an $\mathcal{SO}(3)$ rotation in the $S = 0$ CSFs, followed by backtransformation to the determinant basis is depicted below the definition of $\hat{F}$. The factoring of $\hat{F}$ into a product over a set of 5× representative one-parameter four-qubit gates, and the explicit decomposition of these gates into physically realizable forms in the standard two-qubit gate library is discussed later in the appendix.

D.2. Simplifications of the $\hat{F}$ gate fabric

A natural question at this point is whether there exist gate fabrics for $\mathcal{F}(2^{2M})$ which are simpler than the $\hat{F}$-gate fabric described above. e.g. a simpler gate fabric might have fewer parameters per gate element, and/or fewer QNP product gates per gate element, while still preserving quantum number symmetry and numerical efficiency. For one explicit example, it is clear that the 3× QNP product gates in ($N_\alpha = 1, N_\beta = 1, S = 0$) irrep are redundant, as $\mathrm{QNP}_{\mathrm{PX}}$ and $\mathrm{QNP}_{\mathrm{PBU}}$ (or $\mathrm{QNP}_{\mathrm{PBL}}$) are sufficient to attain any desired action in the irrep. Further, repeated application of pairs of $\mathrm{QNP}_{\mathrm{PX}}$ and $\mathrm{QNP}_{\mathrm{PBU}}$ (or $\mathrm{QNP}_{\mathrm{PBL}}$) gates is sufficient to attain any desired operator in the irrep. So we can already reduce from a five-parameter $\hat{F}$ gate fabric to a four-parameter modified $\hat{F}'$ gate fabric.

Next, we can consider the one-hole and one-particle spaces. Depending on the target irrep, only one of these rotations is generally needed, e.g. for most irreps, a gate fabric of $\mathrm{QNP}_{1\mathrm{p}}$, $\mathrm{QNP}_{\mathrm{PX}}$, and, $\mathrm{QNP}_{\mathrm{PBU}}$ is universal. So, for most irreps, a three-parameter modified $\hat{F}''$ gate fabric is sufficient. A technical detail here is that the choice of $\mathrm{QNP}_{1\mathrm{p}}$ vs $\mathrm{QNP}_{1\mathrm{h}}$ required for universality is contingent on whether there are more particles or holes in the desired irrep of $\mathcal{F}(2^M)$ for extreme edge case irreps.

As we show in the main text, it is possible to reduce $\hat{F}''$ even further to a two-parameter $\hat{Q}$ gate fabric, where the $\hat{Q}$ fabric symmetrizes the rotations between the one-particle and one-hole irreps, and additional mixes rotations between the one-particle/hole irreps and the ($N_\alpha = 1, N_\beta = 1, S = 0$) irrep. To that end we

$$\boxed{\mathrm{QNP}_{\mathrm{OR}}(\varphi)} := \left[\begin{array}{cccccccccccccccc}
1 &&&&&&&&&&&&&&& \\
& c &&& +s &&&&&&&&&&& \\
&& c &&&&&& -s &&&&&&& \\
&&& c^2 &&& +cs &&& +cs &&& +s^2 &&& \\
& -s &&& c &&&&&&&&&&& \\
&&&&& 1 &&&&&&&&&& \\
&&& -cs &&& c^2 &&& -s^2 &&& +cs &&& \\
&&&&&&& c &&&&&& +s && \\
&& -s &&&&&& c &&&&&&& \\
&&& -cs &&& -s^2 &&& c^2 &&& +cs &&& \\
&&&&&&&&&& 1 &&&&& \\
&&&&&&&&&&& c &&& +s & \\
&&& +s^2 &&& -cs &&& -cs &&& c^2 &&& \\
&&&&&&& -s &&&&&& c && \\
&&&&&&&&&&& -s &&& c & \\
&&&&&&&&&&&&&&& 1
\end{array}\right] \in \mathcal{F}(2^{2*2})$$

$$c := \cos(\varphi), \quad s := \sin(\varphi)$$

Figure 8. Spin-adapted spatial orbital rotation gate between two adjacent spatial orbitals. The parameter φ is the argument of the Givens rotation between orbitals $|\phi_0\rangle$ and $|\phi_1\rangle$, with the same Givens rotation applied in the α and β spaces.

consider an alternative QNP gate which is already well-known in the literature, the spatial orbital rotation gate, which we describe in the following section.

D.3. Orbital rotations

A well-known operation in both classical and quantum electronic structure methods is the spin-adapted spatial orbital rotation gate, which implements,

$$|\phi'_p\rangle \equiv \sum_q V_{qp}|\phi_q\rangle \tag{D1}$$

for $V_{pq} \in \mathcal{SO}(M)$, and for the particular case of $M = 2$ adjacent spatial orbitals. If we take V_{pq} to be a 2×2 special orthogonal matrix, i.e. a Givens rotation matrix with parameter φ, then this one-parameter, four-qubit $\mathrm{QNP_{OR}}$ gate (quantum-number-preserving orbital rotation gate) is a special case of the five-parameter, four-qubit $\hat{F}$ gate from figure 7 with, $c := c_{p1} = c_{h1} = \cos(\theta/2)$, $s := s_{p1} = s_{h1} = \sin(\theta/2)$, and

$$u_{ij} = \begin{bmatrix} c^2 & cs & cs & +s^2 \\ -cs & c^2 & -s^2 & cs \\ -cs & -s^2 & c^2 & cs \\ +s^2 & -cs & -cs & c^2 \end{bmatrix}_{i,j}. \tag{D2}$$

This gate can be viewed as a simultaneous and symmetrical application of the $\mathrm{QNP_{1p}}$ and $\mathrm{QNP_{1h}}$ gates which also acts in a direct product manner in the $(N_\alpha = 1, N_\beta = 1, S = 0)$ irrep. The explicit action of the $\mathrm{QNP_{OR}}$ gate is depicted in figure 8.

It is well-known that a fabric of $\binom{M}{2}$ $\mathrm{QNP_{OR}}$ gates arranged in a rectangular or triangular gate fabric pattern can exactly implement an arbitrary orbital rotation within M spatial orbitals, with a classically tractable relationship between the V_{pq} orbital rotation matrix and the parameters $\{\phi_d\}$ of the fabric being possible through the QR decomposition of the orbital rotation matrix.

D.4. Variants of the QNP fabrics

As with the gate fabric in the main text it can be interesting to prepend the parametrized gate elements $\hat{F}'$ or $\hat{F}''$ with a fixed gate like the $\hat{\Pi}$ gate as this may improve trainability of the fabric or even expressiveness at intermediate depths. In the main text we have explored the options $\hat{\Pi} \in \{\hat{I}, \mathrm{QNP_{OR}}(\pi)\}$. Another natural option, inspired by the concept of fermionic swap networks, would be to take $\hat{\Pi}$ to be an orbital wise fermionic swap gate. This gate is also QNP and we introduce it in the end of the following section.

Appendix E. Explicit decompositions of the quantum number preserving gates

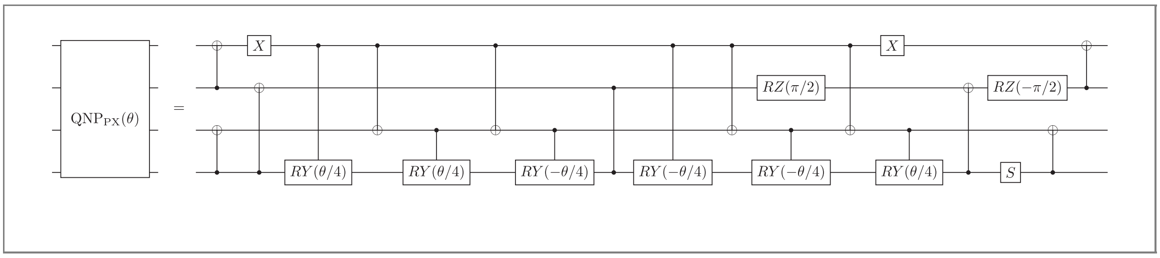
Here we provide explicit decompositions of all the QNP gates introduced in the main text, namely the gates: $\mathrm{QNP_{1p}}$, $\mathrm{QNP_{1h}}$, $\mathrm{QNP_{PX}}$, $\mathrm{QNP_{PBU}}$, $\mathrm{QNP_{PBL}}$, and $\mathrm{QNP_{OR}}$ as well as one additional gate OFSWAP. We call these gates number preserving gates because for any gate G from the above list it holds that

$$\left[\hat{G}(\theta),\hat{N}_\alpha\right]=0,\quad \left[\hat{G}(\theta),\hat{N}_\beta\right]=0,\quad \left[\hat{G}(\theta),\hat{S}^2\right]=0. \tag{E1}$$

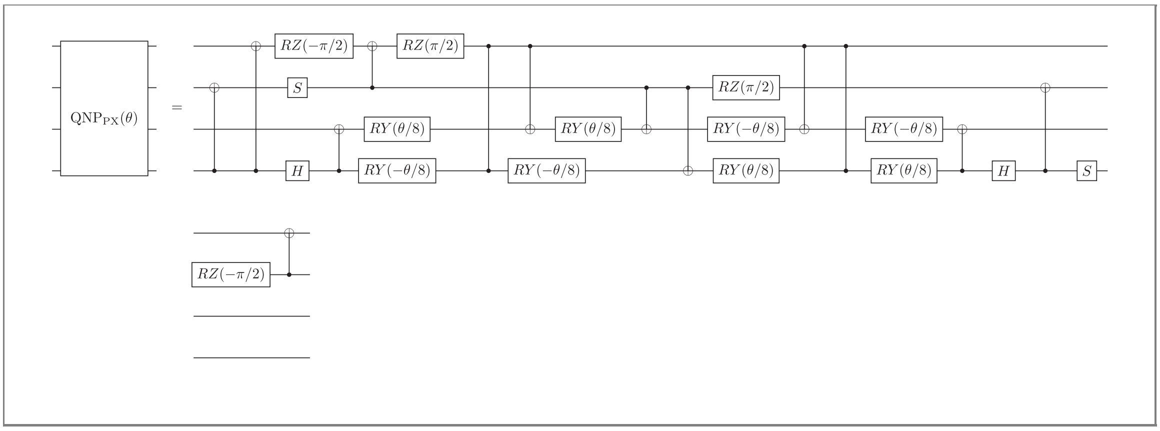
The decompositions of $\mathrm{QNP}_{1\mathrm{p}}(\theta)$ and $\mathrm{QNP}_{1\mathrm{h}}(\theta)$ are given in terms of decompositions of two gates, each acting on just the alpha or beta space such that $\mathrm{QNP}_{1\mathrm{p}}(\theta)=\mathrm{QNP}_{\mathrm{A0B1}}(\theta)\mathrm{QNP}_{\mathrm{A1B0}}(\theta)$ and $\mathrm{QNP}_{1\mathrm{h}}(\theta)=\mathrm{QNP}_{\mathrm{A2B1}}(\theta)\mathrm{QNP}_{\mathrm{A1B2}}(\theta)$. None of these four gates is individually QNP.

E.1. Pair exchange gate $\mathrm{QNP}_{\mathrm{PX}}(\theta)$

For the $\mathrm{QNP}_{\mathrm{PX}}(\theta)$ gate we present the following decomposition in terms of standard gates and controlled Y rotations:



Due to cancellations when expanding out the controlled Y rotations, a decomposition in terms of only standard gates has only slightly higher depth (15 → 18) and requires less two-qubit gates (17 → 13) even if the controlled Y rotation is a native operation:



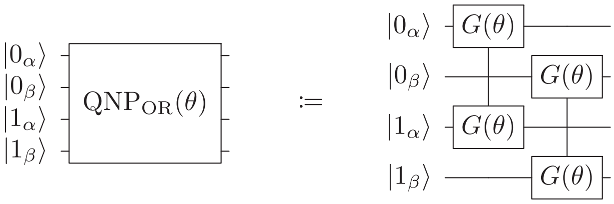
A similar gate was considered in [26, 50].

E.2. Two orbital givens rotation gate $\mathrm{QNP}_{\mathrm{OR}}$

We describe the construction of the Givens rotation gate in more detail. A Givens rotation is generally a rotation in a two dimensional subspace of the form

$$\begin{bmatrix}\phi_0'(\vec{r}_1)\\ \phi_1'(\vec{r}_1)\end{bmatrix}:=\begin{bmatrix}c & +s\\ -s & c\end{bmatrix}\begin{bmatrix}\phi_0(\vec{r}_1)\\ \phi_1(\vec{r}_1)\end{bmatrix}, \tag{E2}$$

where $c:=\cos(\theta/2)$ and $s:=\sin(\theta/2)$ for a continuous parameter θ. Under the Jordan–Wigner mapping a Givens rotation between the orbital bases can be implemented as pair of parallel Givens gates as follows:

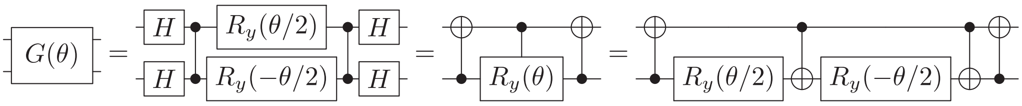


(E3)

In the two qubit space, the Givens rotation gate $G(\theta)$ has the action

$$\hat{G}(\theta) := \begin{bmatrix} 1 & & & \\ & c & +s & \\ & -s & c & \\ & & & 1 \end{bmatrix} \tag{E4}$$

and can be decomposed into elementary gates as follows:

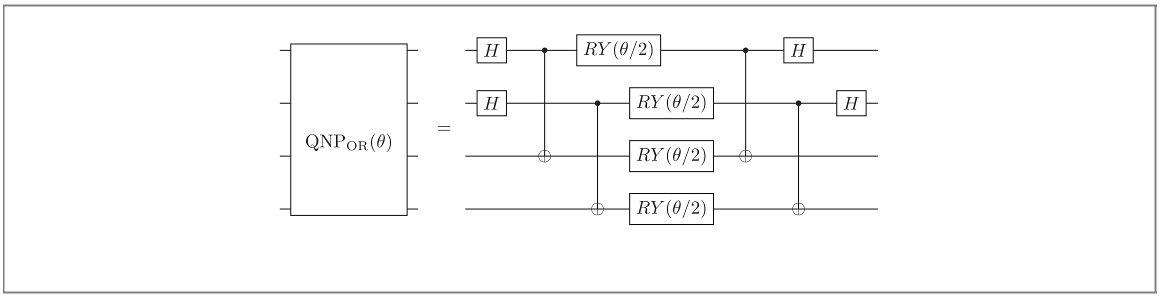


(E5)

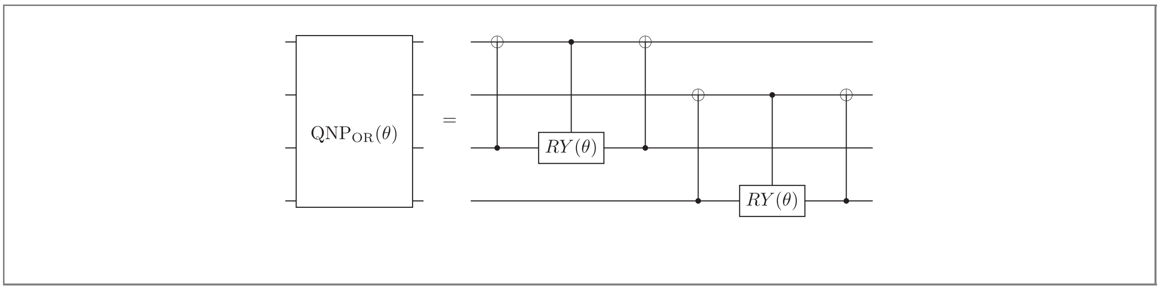
In the four-qubit Hilbert space, the two orbital Givens rotation gate $\mathrm{QNP}_{\mathrm{OR}}(\theta)$ has the action

$$\mathrm{QNP}_{\mathrm{OR}}(\theta) := \begin{bmatrix}
1 & & & & & & & & & & & & & & & \\
 & c & & & +s & & & & & & & & & & & \\
 & & c & & & & & & -s & & & & & & & \\
 & & & c^2 & & & +cs & & & +cs & & & +s^2 & & & \\
 & -s & & & c & & & & & & & & & & & \\
 & & & & & 1 & & & & & & & & & & \\
 & & & -cs & & & c^2 & & & -s^2 & & & +cs & & & \\
 & & & & & & & c & & & & & & +s & & \\
 & & -s & & & & & & c & & & & & & & \\
 & & & -cs & & & -s^2 & & & c^2 & & & +cs & & & \\
 & & & & & & & & & & 1 & & & & & \\
 & & & & & & & & & & & c & & & +s & \\
 & & & +s^2 & & & -cs & & & -cs & & & c^2 & & & \\
 & & & & & & & -s & & & & & & c & & \\
 & & & & & & & & & & & -s & & & c & \\
 & & & & & & & & & & & & & & & 1
\end{bmatrix}. \tag{E6}$$

When applied to two neighboring spatial orbitals, this gate also preserves all three quantum numbers and has the following decomposition with gate depth 5 and just 4 CNOT gates:

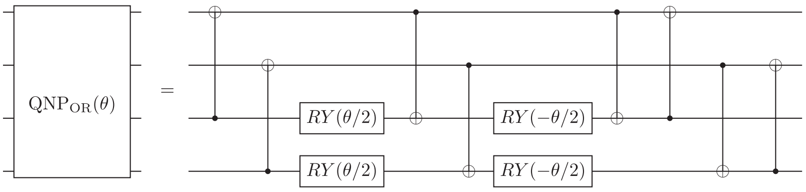


An alternative decomposition into controlled Y rotation gates is:



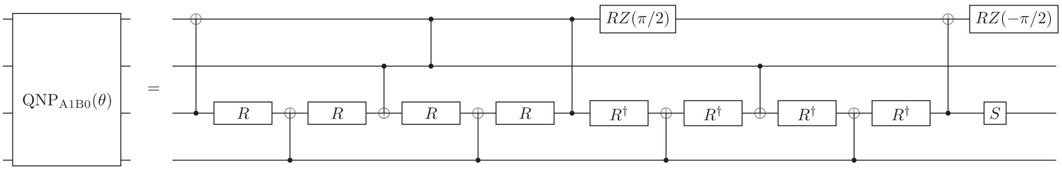
Note that if these gates are native, the two-qubit gate count is raised to 6 while reducing the depth to 3. Expanding out the controlled Y rotations yields:

Of course the two Givens rotations of the gate commute and can be performed at the same time, giving a gate depth of 6 and a CNOT count of 8. Depending on the preceding and following gates, it may furthermore be favourable to substitute the doubled CNOT gates with a single CNOT and an SWAP gate.

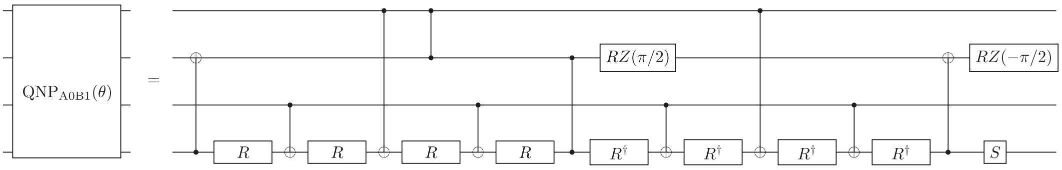


E.3. Single particle and single hole gates

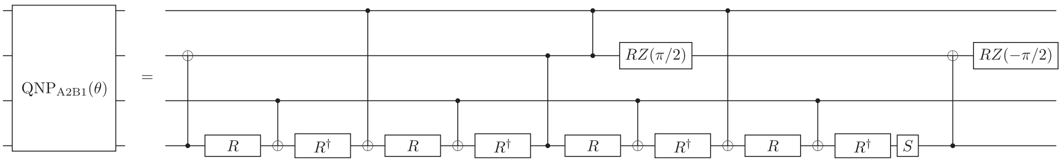
In the following we abbreviate $R := RY(\theta/8)$. The $\mathrm{QNP}_{\mathrm{A1B0}}(\theta)$ gate can be decomposed as follows:



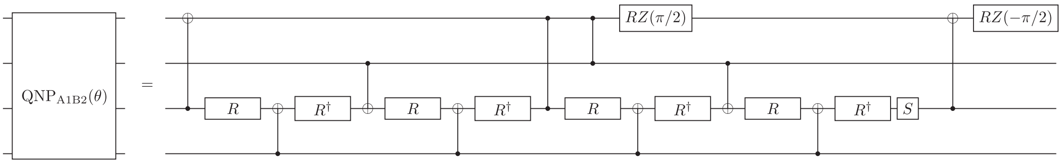
The $\mathrm{QNP}_{\mathrm{A0B1}}(\theta)$ gate can be decomposed as follows:



The $\mathrm{QNP}_{\mathrm{A2B1}}(\theta)$ gate can be decomposed as follows:

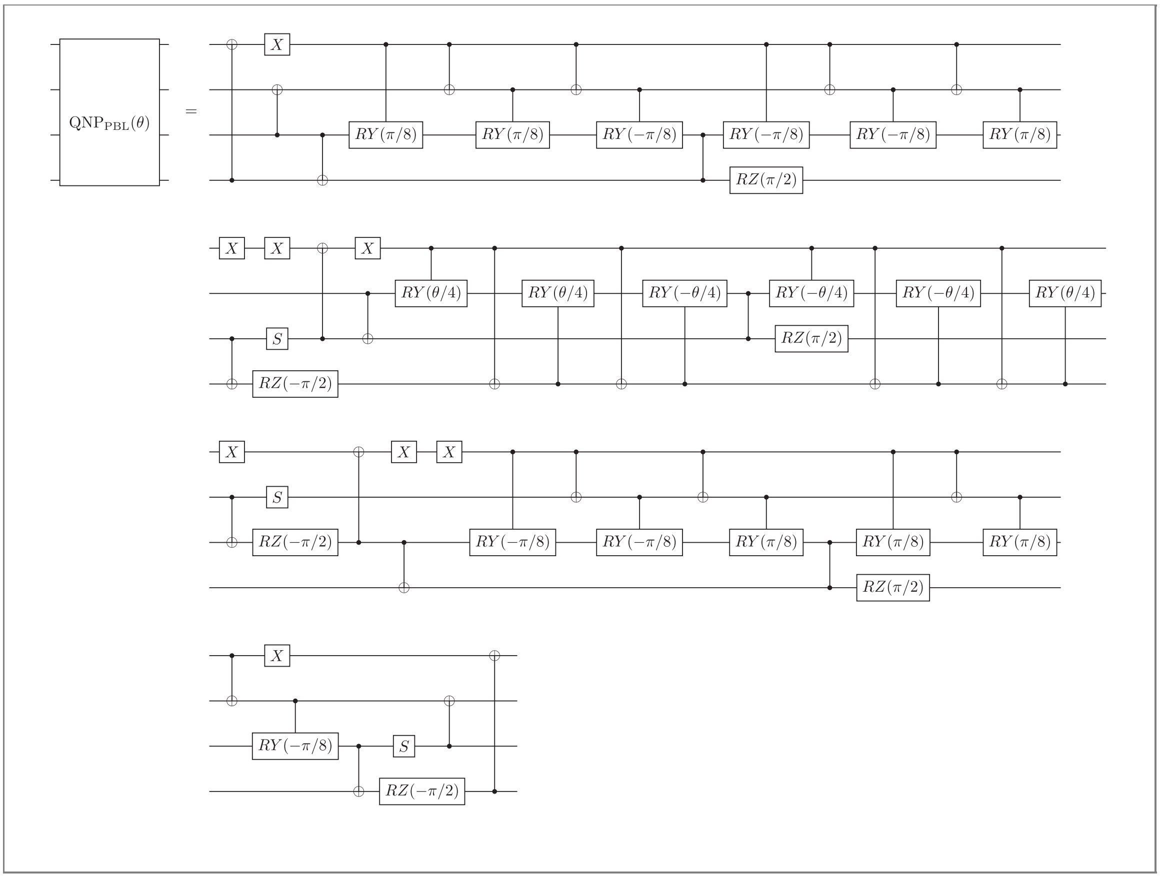


Finally, the $\mathrm{QNP}_{\mathrm{A1B2}}(\theta)$ gate can be decomposed as follows:

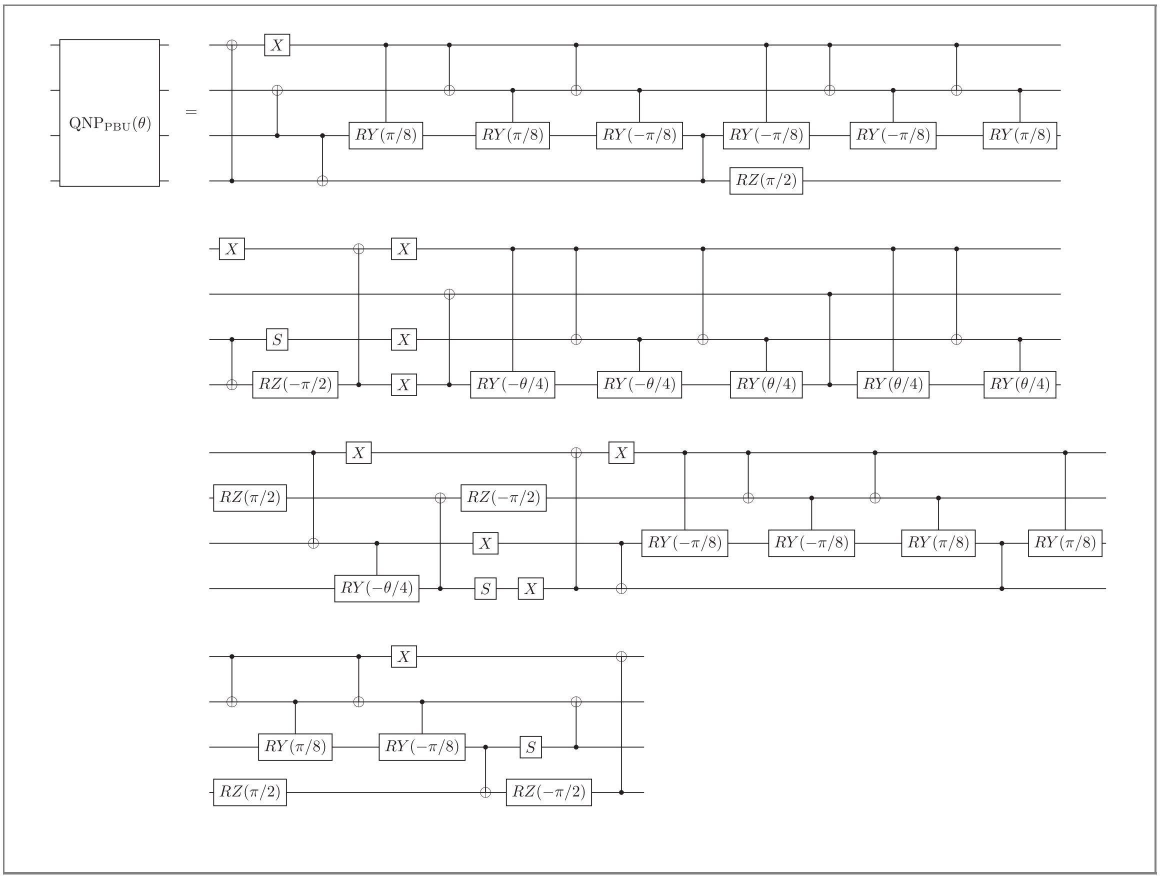


E.4. Pair breaking gates

For the pair breaking gates we present decompositions into standard gates and controlled Y rotations. The pair break low gate $\mathrm{QNP_{PBL}}$ has the decomposition:



While the pair break up gate $\mathrm{QNP_{PBU}}$ has the decomposition:



E.5. Fermionic orbital swap gate

Finally, the OFSWAP gate is an orbital wise fermionic swap gate, i.e. two fermionic swap gates (an SWAP gate followed by a controlled Z gate) with action

$$\mathrm{FSWAP} = \begin{bmatrix} +1 & & & \\ & & +1 & \\ & +1 & & \\ & & & -1 \end{bmatrix}, \tag{E7}$$

where the -1 in the lower right corner takes into account the sign of the fermionic anti-commutation relations, applied between the alpha and beta wires respectively. Curiously OFSWAP is only up to phases representable by an orbital rotation as we have OFSWAP $\hat{Z}_0\hat{Z}_{\bar{0}} = \mathrm{QNP}_{2\mathrm{OGR}}(\pi)$.

Appendix F. Generalized parameter-shift rules

In order to compute the derivative of expectation values with respect to quantum gate parameters, the so-called parameter-shift rule has been established as a tool to avoid finite difference derivatives, which become unstable under the influence of noise from both, measurements and circuit imperfections [51]. In addition to the original concept, multiple efforts have been made to analyze and generalize the parameter-shift rule [33, 35, 38, 52, 53]. In this appendix we introduce the concept of tuning the shift angle in parameter-shift rules for an algorithmic advantage (appendix F.1), a new four-term parameter-shift rule for gates with three distinct eigenvalues (appendix F.2) and exclude a further straightforward generalization of this type of parameter-shift rules (appendix F.2.2). We also compare our new four-term rule to the one recently presented in [37] (appendix F.2.1) and extend the variance minimization strategy from [38] to both four-term rules. This four term shift rule is applicable to all of the QNP gates introduced above, except for the spin adapted QNP$_{\mathrm{OR}}$ gate, which can be analytically differentiated by differentiating the individual G gates and using chain rule.

F.1. Shift tuning

We briefly recap the derivation of the standard parameter-shift rule without fixing the shift angle, leading to a free parameter in the rule. Consider a parametrized gate of the form

$$U(\theta) := \exp\left(-\mathrm{i}\frac{\theta}{2}P\right) \tag{F1}$$

where $P^2 = \mathbf{1}$, as is the case for example for Pauli rotation gates. In a circuit with an arbitrary number of parameters, let us single out the parameter of the gate U above and write our cost function of interest as

$$f(\theta) = \langle\psi(\theta)|B_0|\psi(\theta)\rangle := \langle\phi|U(\theta)^\dagger BU(\theta)|\phi\rangle, \tag{F2}$$

where the part of the circuit preparing $|\psi\rangle(\theta)$ from some initial state applied before the gate U has been absorbed into $|\phi\rangle$ and the part after U is absorbed in to B. Then the derivative is, by the product rule, given by

$$\frac{\partial}{\partial\theta}f(\theta) = \langle\phi|U(\theta)^\dagger\left(-\frac{\mathrm{i}}{2}[B,P]\right)U(\theta)|\phi\rangle. \tag{F3}$$

Now look at the conjugation of B by U at an arbitrary shift angle $\pm\alpha$:

$$\begin{aligned}\mathcal{U}(\pm\alpha)(B) &:= U(\pm\alpha)^\dagger BU(\pm\alpha)\\ &= U(\pm\alpha)^\dagger B\left(\cos\left(\frac{\alpha}{2}\right)\mathbf{1} \mp \mathrm{i}\,\sin\left(\frac{\alpha}{2}\right)P\right)\\ &= \cos\left(\frac{\alpha}{2}\right)^2 B + \sin\left(\frac{\alpha}{2}\right)^2 PBP \mp \frac{\mathrm{i}}{2}\,\sin(\alpha)[B,P].\end{aligned} \tag{F4}$$

Subtracting $\mathcal{U}(-\alpha)(B)$ from $\mathcal{U}(\alpha)(B)$ and excluding multiples of π as values for α, we obtain the generalized two-term parameter-shift rule

$$\mathcal{U}(\alpha)(B) - \mathcal{U}(-\alpha)(B) = -\mathrm{i}\,\sin(\alpha)[B,P] \tag{F5}$$

$$\Rightarrow \frac{\partial}{\partial\theta}f(\theta) = \frac{1}{2\,\sin(\alpha)}\left(f(\theta+\alpha) - f(\theta-\alpha)\right) \tag{F6}$$

where the original parameter-shift rule corresponds to choosing $\alpha = \pi/2$. We note that the concept of shift-tuning was independently discovered in [38] and introduced in the quantum computing software package PennyLane [36].

F.1.1. Reducing the gate count

In particular, the general form of equation (F5) allows us—provided that θ is not a multiple of π—to choose $\alpha = -\theta$, making the first of the cost function evaluations $f(0)$ and therefore reducing the gate count because $U(0) = \mathbf{1}$ can be skipped in the circuit. This may lead to an additional gate count reduction if the neighboring gates on both sides of U can be merged, which is true for example in circuits for the QAOA.

F.2. Four-term parameter-shift rule

Here we derive a four-term parameter-shift rule for gates that do not fulfill the two-term rule, e.g. controlled rotation gates like $CR_Z(\theta)$ or many of our QNP gates with one parameter.

To this end, consider a gate

$$U(\varphi) := \exp\left(-\mathrm{i}\frac{\varphi}{2}Q\right) \tag{F7}$$

with $Q^3 = Q$ but not necessarily $Q^2 = \mathbf{1}$, as is true for any gate with spectrum $\{-1, 0, 1\}$. Then the exponential series can be rewritten as

$$U(\varphi) = \mathbf{1} + \left(\cos\left(\frac{\varphi}{2}\right) - 1\right)Q^2 - \mathrm{i}\,\sin\left(\frac{\varphi}{2}\right)Q \tag{F8}$$

and a computation similar to the one above leads to

$$\begin{aligned}\mathcal{U}(\alpha)(B) - \mathcal{U}(-\alpha)(B) = &-2\mathrm{i}\,\sin\left(\frac{\alpha}{2}\right)[B,Q]\\ &- 2\mathrm{i}\,\sin\left(\frac{\alpha}{2}\right)\left(\cos\left(\frac{\alpha}{2}\right) - 1\right)[Q,QBQ].\end{aligned} \tag{F9}$$

We can then obtain the commutator by linearly combining this difference with itself for a second angle $\pm\beta$, so that

$$-\frac{\mathrm{i}}{2}[B, Q] = d_1\left(\mathcal{U}(\alpha)(B) - \mathcal{U}(-\alpha)(B)\right) \tag{F10}$$

$$- d_2\left(\mathcal{U}(\beta)(B) - \mathcal{U}(-\beta)(B)\right) \tag{F11}$$

which holds true if the angles α, β and the prefactors $d_{1,2}$ satisfy

$$\frac{1}{4} = d_1 \sin\left(\frac{\alpha}{2}\right) - d_2 \sin\left(\frac{\beta}{2}\right) \tag{F12}$$

$$\frac{1}{2} = d_1 \sin(\alpha) - d_2 \sin(\beta) . \tag{F13}$$

Therefore, we get the four-term parameter-shift rule

$$\frac{\partial}{\partial\varphi} f(\varphi) = d_1\left(f(\varphi+\alpha) - f(\varphi-\alpha)\right) - d_2\left(f(\varphi+\beta) - f(\varphi-\beta)\right), \tag{F14}$$

where we again can choose α or β such that one of the function evaluations skips the gate U. A particularly symmetric solution of equations (F12) and (F13) is

$$d_1 = \frac{1}{2}, \qquad d_2 = \frac{\sqrt{2}-1}{4}, \qquad \alpha = \frac{\pi}{2}, \qquad \beta = \pi. \tag{F15}$$

In general, any gate for which the spectrum of the generator is $\{-a+c, c, a+c\}$ obeys the four-term parameter-shift rule as the shift c can be absorbed into a global phase that does not contribute to the gradient and a can be absorbed into the variational parameter of the gate.

As an example, the four-term rule is applicable to (multi-)controlled Pauli rotations $CR_P(\varphi)$ for which Q is the zero matrix except for the Pauli operator P on the target qubit. For multiple control qubits and our QNP gates, this will lead to less circuit evaluations using the chain rule and applying the two-term rule to the gate decomposition.

In order to find out whether an n-qubit single-parameter gate U satisfies the four-term rule, one can compute

$$Q = \frac{\partial}{\partial\varphi} U(\varphi)\Big|_{\varphi=0}, \qquad \overline{Q} := Q - \frac{1}{2^n}\, tr(Q) \tag{F16}$$

and test if there is an $a \in \mathbb{R}$ such that $\overline{Q}^3 = a^2\,\overline{Q}$, which is a sufficient condition, as the only thing we needed for the four term rule to apply was this assumptions about the generator spectrum.

F.2.1. Relation to other four-term rule

Previous work showed the existence of a four-term parameter-shift rule [37] for gates of the form (F7), which is implemented with only one shift angle but requires the two additional gates

$$V_{\pm} = \exp\left(\mp\frac{\mathrm{i}\alpha\pi}{4} P_0\right) \quad \text{with } P_0 = \mathbf{1} - Q^2.$$

There are four relevant aspects when comparing this rule to the one in (F14): first, our four-term rule does not require any additional gates like $V_{\pm}$, which add overhead to the gradient evaluation circuits. While the authors bound the additional cost by the cost of the differentiated gate itself, it might more crucially be non-trivial to construct $V_{\pm}$ for gates that do not have an obvious fermionic representation like the gates considered in [37].

Second, the shift tuning technique for gate count reduction in (F1) can easily be extended to both, our four-term rule and the rule derived in [37], provided one has access to the parametrized versions of $V_{\pm}$. As the construction of $V_{\pm}$ for fermion-based gates is based on rotations, this access can be assumed for these gates whenever $V_{\pm}$ can be implemented.

Third, it was shown in [37] that their four-term rule reduces to a standard *two*-term rule up to the insertions of the $V_{\pm}$ operators whenever both the circuit of interest and the measured observables are purely real-valued. This is the case for virtually all molecular Hamiltonians and most of the circuits proposed for quantum chemistry problems—including the fabrics in this work—such that gradients of highly complex gates may be computed with just two circuit executions including the gates $V_{\pm}$ using the rule in [37].

Fourth, the variances of the derivative estimators given by the two rules can be minimized to the same value by choosing the shift angles optimally, as shown in appendix F.3. This means that for a given budget of circuit executions, the quality of the estimated derivative is the same, even though the number of distinct circuits differs.

In summary, the specialized two-term parameter-shift rule in [37] is preferable if the following three criteria hold: firstly, the circuit and observable need to be real-valued. Secondly, the auxiliary gates $V_\pm$ have to be available. Thirdly, the computation must happen on a simulation level in which the number of *distinct* circuits instead of the measurement budget is relevant, so that the reduction from four to two terms provides an advantage which is larger than the overhead of adding $V_\pm$. In all other scenarios the four-term rule equation (F14) with the optimal parameters in equations (F24) and (F25) requires slightly fewer gates and the same number of circuit executions, making it preferable in particular on quantum computers.

F.2.2. Impossibility of some further shift rules

One may wonder whether a three shift rule is possible for gates whose generators have just three distinct eigenvalues and whether shift rules exist for gates with more distinct eigenvalues. We present some insights on these questions in the following.

During the derivation of the four-term parameter-shift rule we chose to first linearly combine $\mathcal{U}(\pm\alpha)(B)$ and $\mathcal{U}(\pm\beta)(B)$ with the same prefactors, respectively. Alternatively one may try to combine $\mathcal{U}(\alpha_i)(B)$ at three shift angles $\{\alpha_i\}_{i\in\{1,2,3\}}$ linearly and demand the result to fulfill

$$\sum_{i=1}^{3} d_i\mathcal{U}(\alpha_i)(B) \overset{!}{=} -\frac{\mathrm{i}}{2}[B,Q]. \tag{F17}$$

This leads to the system of equations

$$\begin{aligned}
0 &= d_1[c_1-c_3]+d_2[c_2-c_3]\\
0 &= d_1\left[s_1^2-s_3^2\right]+d_2\left[s_2^2-s_3^2\right]\\
1 &= 2d_1[s_1-s_3]+2d_2[s_2-s_3]\\
1 &= d_1\left[\sin(\alpha_1)-\sin(\alpha_3)\right]+d_2\left[\sin(\alpha_2)-\sin(\alpha_3)\right]
\end{aligned}$$

with $c_i=\cos\left(\frac{\alpha_i}{2}\right)$ and $s_i=\sin\left(\frac{\alpha_i}{2}\right)$, which we conjecture to not have a solution.

Considering the generalization of the (standard) two-term shift rule to the four-term rule in (F14) and their requirement on the gate generator, i.e. $Q^2=\mathbf{1}$ and $Q^3=Q$, it seems a natural question whether further generalization is possible to gates that, e.g. fulfill $Q^5=Q$. We show next that this is not the case.

Consider the generalized condition $Q^m=Q^n$, $m\neq n$ for the generator of a d-dimensional one-parameter gate. We recall that we may absorb shifts and scaling prefactors of the spectrum of Q into a global phase gate and the variational parameter, respectively, which may be used to obtain gates satisfying the generalized condition $Q^m=Q^n$. In the eigenbasis of the Hermitian matrix Q, this condition becomes $\lambda_i^m=\lambda_i^n\ \forall 1\leqslant i\leqslant d$, which only ever is solved by -1, 0 and 1 over $\mathbb{R}$ (in which the spectrum of Q must be contained) with the additional condition $m-n \bmod 2=0$ for $\lambda_i=-1$. This means that Q already satisfies $Q^3=Q$, allowing for the four-term rule to be applied.

Consequently, a direct generalization of the four-term rule is not possible. Note that this does not exclude the existence of other schemes to compute the derivative of an expectation value w.r.t. parametrized states that are based on linear combinations of shifted expectation values.

F.3. Minimizing the variance

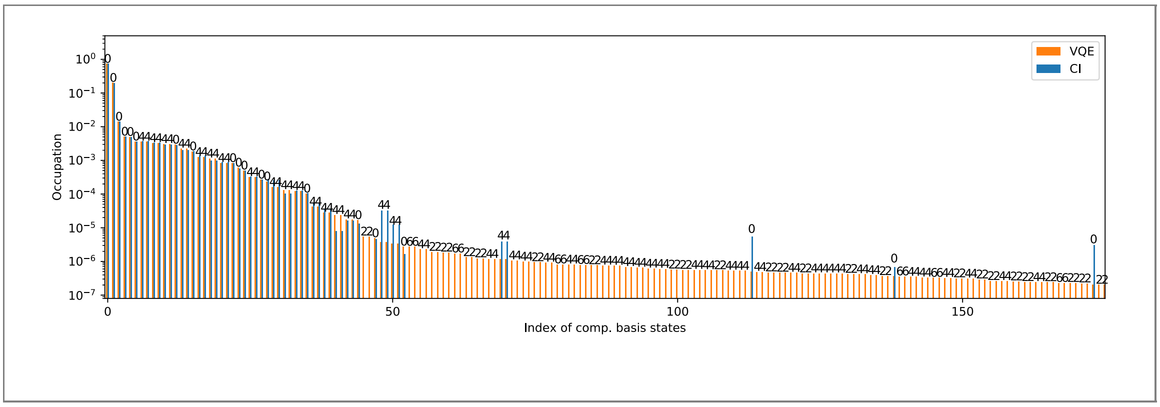
If we approximate the physical variance of the expectation value, V, to be independent of θ, the variance of measuring f at a given parameter for sufficiently many measurements N is V/N. The resulting variance of the two-term shift rule derivative for a budget of N measurements is

$$\sigma^2=\frac{V}{N\sin^2\alpha}, \tag{F18}$$

where we chose the optimal allocation of $N/2$ measurements to each of the two terms in the shift rule. We may optimize the shift angle in the two-term rule w.r.t. this variance which yields the standard choice $\pi/2$ for the shift because

$$\underset{\alpha}{\arg\min}\left\{\frac{V}{N\sin^2\alpha}\right\}=\frac{\pi}{2}. \tag{F19}$$

The variance can be reduced further by introducing a multiplicative bias to the estimator, as presented in [38]; the optimal choice of the prefactor depends on the value and the variance of the derivative and is



given by

$$\lambda^{*} = \left(1 + \frac{V}{N(\partial_\theta f)^2}\right)^{-1}. \tag{F20}$$

Note that λ^{*} has to be estimated because V and $\partial_\theta f$ are not known exactly. The optimal choice of the shift parameter remains $\frac{\pi}{2}$.

For the four-term rule in equation (F14), the optimal shot allocation is proportional to the prefactors $d_{1,2}$ and leads to the variance

$$\sigma^2 = 4(d_1 + d_2)^2 \frac{V}{N}. \tag{F21}$$

As for the two-term parameter-shift rule, we may minimize this variance w.r.t. α and β via d_1 and d_2, which are given via equations (F12) and (F13) by

$$d_2 = \frac{1}{4\,\sin(\beta/2)} \frac{\cos(\alpha/2) - 1}{\cos(\beta/2) - \cos(\alpha/2)} \tag{F22}$$

$$d_1 = \frac{1}{\sin(\alpha)} \left(\frac{1}{2} + d_2\,\sin(\beta)\right). \tag{F23}$$

This results in

$$d_1 = \frac{\sqrt{2}+1}{4\sqrt{2}}, \qquad \alpha = \frac{\pi}{2}, \tag{F24}$$

$$d_2 = \frac{\sqrt{2}-1}{4\sqrt{2}}, \qquad \beta = \frac{3\pi}{2} \tag{F25}$$

and three equivalent solutions based on the symmetries of equations (F12) and (F13).

The variance then is $\sigma^2 = V/N$ as for the optimal two-term rule and again it may be further reduced by introducing a bias via a multiplicative prefactor λ, with the same optimal λ^{*} as before.

For both, the four-term rule and the specialized two-term rule in [37], the minimal variance is $\sigma^2 = V/N$ as well, as the prefactors are equally large and sum to 1.

In conclusion, under the constant variance assumption, the variance for all discussed two- and four-term parameter shift rules is the same at a given measurement budget, showing that they are equally expensive on a quantum device, for which the number of measurements instead of the number of distinct circuits is relevant.

Appendix G. Additional numerical results

G.1. Computational basis state amplitudes

In figure 3(b) of the main text the individual ordering of each trace of computational basis states is ordered individually, which allows to view the shape of each tail but restricts comparability between single amplitudes. In figure 9 only the computational basis states of the true ground state and of one optimized VQE state at 110 parameters are plotted in consistent ordering. This allows for direct comparison between the amplitudes of the VQE and of FCI and demonstrates how our fabric finds a good approximation to most amplitudes while having far too few parameters to reproduce all amplitudes exactly.

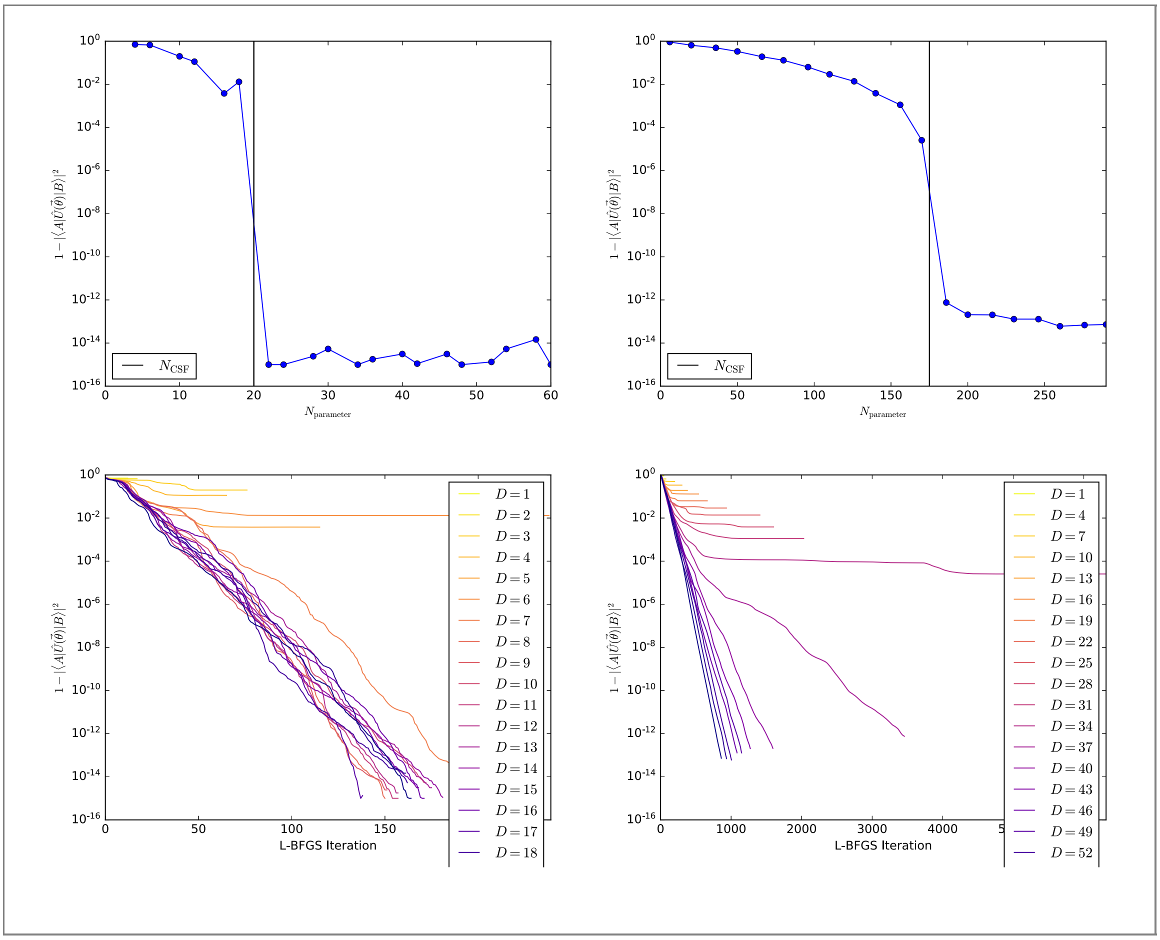


Table 3. Quantum number irreps for $M = 4$ for which the $\hat{Q}$-type QNP gates of the main text are not universal. Overall there are 35 unique irreps for $M = 4$ with total dimension $D \equiv 2^{2M} = 256$. The 6 irreps with total dimension 36 listed below are not universal due to high-spin constraints. All other irreps are numerically found to be universal to the essentially machine precision.

N_α	N_β	S	Dimension
0	2	2	6
1	1	2	6
2	0	2	6
2	4	2	6
3	3	2	6
4	2	2	6

G.2. Numerical universality demonstration for Haar random states

The test cases in real molecular systems in the main text are somewhat complicated by the specifics of the electronic structure Hamiltonian and especially by the spatial point group symmetry of the test molecules. One notable artifact is that some of the left-most gates in our gate fabrics in real molecules are ‘dead,’ as they perform orbital rotations and diagonal pair exchanges in the occupied or virtual subspaces of the HF starting state. The point group symmetry also seems to adversely affect the numerical convergence behavior of the VQE gate fabric parameter optimization, e.g. suggesting the $\hat{\Pi}$-gate pre-mixing initialization adopted in the main text. Noticeably better convergence behavior was observed when the molecules were perturbed from $D_{2\mathrm{h}}$ symmetry to C_1 by random Gaussian perturbations in XYZ coordinates.

This section is included to demonstrate the numerical universality properties of our proposed gate fabric for the artificial case of Haar random statevectors. Specifically, for a number of test case irreps $(M, N_\alpha, N_\beta, S)$, we form the full CSF basis, and then generate Haar random statevectors $|A\rangle$ and $|B\rangle$ within this irrep of $\mathcal{F}(2^{2M})$ by Gaussian random sampling and normalization of the statevector in the CSF basis, and then backtransformation to the standard Jordan–Wigner computational basis. We then optimize the VQE gate fabric parameters of the VQE entangler circuit $\hat{U}$ to maximize $|\langle A|\hat{U}|B\rangle|^2$ via L-BFGS with

Table 4. Quantum number irreps for $M = 6$ for which the $\hat{Q}$-type QNP gates of the main text are not universal. Overall there are 84 unique irreps for $M = 6$ with total dimension $D \equiv 2^{2M} = 4096$. The 24 irreps with total dimension 400 listed below are not universal due to high-spin constraints. All other irreps are numerically found to be universal to the essentially machine precision.

N_α	N_β	S	Dimension
0	2	2	15
0	3	3	20
0	4	4	15
1	1	2	15
1	2	3	20
1	3	4	15
2	0	2	15
2	1	3	20
2	2	4	15
2	6	4	15
3	0	3	20
3	1	4	15
3	5	4	15
3	6	3	20
4	0	4	15
4	4	4	15
4	5	3	20
4	6	2	15
5	3	4	15
5	4	3	20
5	5	2	15
6	2	4	15
6	3	3	20
6	4	2	15

noise-free analytical gradients. Note that we do not perform $\hat{\Pi}$-based pre-mixing convergence enhancement in this section.

The results are shown for the half-filled cases for $M = 4$ and $M = 6$ in figure 10. The top panels show the bulk convergence properties with respect to circuit depth D/number of parameters $N_{\text{parameter}}$ (roughly linearly proportional). The general finding here is roughly geometric convergence at low parameter depths, followed by a sharp drop to near the machine epsilon as the number of parameters crosses over the number of CSFs in the irrep, indicating the onset of universality in the action of the VQE entangler circuit. Quantum number symmetries are preserved to at least the machine epsilon for all intermediate and final parameter values. The lower panels show the numerical convergence behavior of the L-BFGS optimization procedure for each point in the top panel. There are several salient features in these plots: (1) the earliest convergence behavior appears to be roughly geometric, and self-similar between gate fabrics with different numbers of parameters (2) fabrics with smaller numbers of parameters deviate earlier from this geometric convergence and eventually 'flatline' at their non-universal terminal values (3) some minor plateaus are observed in the convergence behavior for small numbers of parameters (4) there is a distinct phase change as universality is crossed, with circuits with larger numbers of parameters than needed for universality exhibiting strongly geometric convergence behavior all the way to the machine epsilon.

Such tests are assuredly artificial, but are free from the external artifacts present in the molecular test cases, and serve to more-strongly indicate that the gate fabric developed in figure 2 are universal and quantum-number-symmetry-preserving for $\mathcal{F}(2^{2M})$.

G.3. Non-universal edge cases

It is important to note that while the $\hat{Q}$-type QNP gate fabrics of main text are numerically universal for the vast majority of quantum number irreps in the 'bulk' of the Hilbert space, there are a limited number of edge cases for which these gate fabrics are not universal. These cases constitute systems where, after high-spin constraints are accounted for, there are only holes or particles left in the remaining orbitals. In these cases, the QNP_{PX} gates have trivial action in the wholly hole or particle space, and are unable to explore new configurations within the space. More tangibly, for an irrep with dimensions $(M, N_\alpha, N_\beta, S)$, we first compute the 'unconstrained' irrep $(M - S, N_\alpha', N_\beta', 0)$ where $N_\alpha' + N_\beta' + S = N_\alpha + N_\beta$ and the larger of $N_\alpha' := N_\alpha$ or $N_\beta' := N_\beta$ is decremented first until $N_\alpha' = N_\beta'$, and then both N_α' and N_β' are decremented together (in this line, $:=$ is read as 'initialized to'). The resulting unconstrained irrep will always have $N_\alpha' = N_\beta'$. If the unconstrained irrep is all holes ($N_\alpha' = N_\beta' = 0$) or all particles ($N_\alpha' = N_\beta' = M - S$), then the $\hat{Q}$-type QNP gate fabric is not universal. A trivial exception is if only a single orbital with all holes or all particles remains in the unconstrained irrep, in which case universality is still preserved.

Note that the number of irreps in the Hilbert space is growing roughly as $\mathcal{O}(M^3)$, while the required constraints $N'_\alpha = N'_\beta = 0$ or $M - S$ seem to indicate that the number of non-universal irreps indicate that the number of irreps which are not universal with $\hat{Q}$-type QNP gate fabrics will grow as roughly $\mathcal{O}(M)$. Moreover, the non-universal irreps appear at the 'edge' of the Hilbert space, and consist of cases with severe high-spin constraints which are likely to be either polynomially tractable classically, physically uninteresting, or both. Interesting cases with roughly half-and-half filling of holes and particles and moderately low total spin number will almost surely fall into irreps which are universal with $\hat{Q}$-type QNP gate fabrics. Finally, it is worth noting that any issues with these edge cases can be completely obviated by instead working with the five-parameter $\hat{F}$ gate fabrics discussed in appendix D—these do not appear to exhibit any edge case non-universalities, and are numerically universal for all cases we have tested.

Tables 3 and 4 show explicitly the irreps for $M = 4$ and $M = 6$ that were found numerically to be non-universal with $\hat{Q}$-type QNP gate fabrics via numerical studies of the same type as the previous section. The non-universality behavior was immediately apparent as discrepancies of overlap of $1 - |\langle A|\hat{U}|B\rangle|^2$ of order of 10^{-2}, while the universal irreps exhibited maximum discrepancies of overlap of order of $< 10^{-13}$.

References

[1] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O'brien J L 2014 A variational eigenvalue solver on a photonic quantum processor *Nat. Commun.* **5** 4213

[2] McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 The theory of variational hybrid quantum-classical algorithms *New J. Phys.* **18** 023023

[3] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79

[4] McClean J R, Kimchi-Schwartz M E, Carter J and de Jong W A 2017 Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states *Phys. Rev.* A **95** 042308

[5] Parrish R M, Hohenstein E G, McMahon P L and Martínez T J 2019 Quantum computation of electronic transitions using a variational quantum eigensolver *Phys. Rev. Lett.* **122** 230401

[6] Nakanishi K M, Mitarai K and Fujii K 2019 Subspace-search variational quantum eigensolver for excited states *Phys. Rev. Res.* **1** 033062

[7] Urbanek M, Camps D, Van Beeumen R and de Jong W A 2020 Chemistry on quantum computers with virtual quantum subspace expansion *J. Chem. Theory Comput.* **16** 5425–31

[8] Ollitrault P J *et al* 2020 Quantum equation of motion for computing molecular excitation energies on a noisy quantum processor *Phys. Rev. Res.* **2** 043140

[9] Huggins W J, Lee J, Baek U, O'Gorman B and Whaley K B 2020 A non-orthogonal variational quantum eigensolver *New J. Phys.* **22** 073009

[10] Parrish R M and McMahon P L 2019 Quantum filter diagonalization: quantum eigendecomposition without full quantum phase estimation (arXiv:1909.08925)

[11] Stair N H, Huang R and Evangelista F A 2020 A multireference quantum Krylov algorithm for strongly correlated electrons *J. Chem. Theory Comput.* **16** 2236–45

[12] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 Barren plateaus in quantum neural network training landscapes *Nat. Commun.* **9** 4812

[13] Gard B T, Zhu L, Barron G S, Mayhall N J, Economou S E and Barnes E 2020 Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm *npj Quantum Inf.* **6** 10

[14] O'Malley P J J *et al* 2016 Scalable quantum simulation of molecular energies *Phys. Rev.* X **6** 031007

[15] Ryabinkin I G, Yen T-C, Genin S N and Izmaylov A F 2018 Qubit coupled cluster method: a systematic approach to quantum chemistry on a quantum computer *J. Chem. Theory Comput.* **14** 6317–26

[16] Lee J, Huggins W J, Head-Gordon M and Whaley K B 2018 Generalized unitary coupled cluster wave functions for quantum computation *J. Chem. Theory Comput.* **15** 311–24

[17] O'Gorman B, Huggins W J, Rieffel E G and Whaley K B 2019 Generalized swap networks for near-term quantum computing (arXiv:1905.05118)

[18] Matsuzawa Y and Kurashige Y 2020 Jastrow-type decomposition in quantum chemistry for low-depth quantum circuits *J. Chem. Theory Comput.* **16** 944–52

[19] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets *Nature* **549** 242–6

[20] Bian T, Murphy D, Xia R, Daskin A and Kais S 2019 Quantum computing methods for electronic states of the water molecule *Mol. Phys.* **117** 2069–82

[21] Grimsley H R, Economou S E, Barnes E and Mayhall N J 2019 An adaptive variational algorithm for exact molecular simulations on a quantum computer *Nat. Commun.* **10** 3007

[22] Elfving V E, Millaruelo M, Gámez J A and Gogolin C 2020 Simulating quantum chemistry in the seniority-zero space on qubit-based quantum computers (arXiv:2002.00035)

[23] Evangelista F A, Chan G K-L and Scuseria G E 2019 Exact parameterization of fermionic wave functions via unitary coupled cluster theory *J. Chem. Phys.* **151** 244112

[24] Ganzhorn M *et al* 2019 Gate-efficient simulation of molecular eigenstates on a quantum computer *Phys. Rev. Appl.* **11** 044092

[25] Xia R and Kais S 2020 Qubit coupled cluster singles and doubles variational quantum eigensolver ansatz for electronic structure calculations *Quantum Sci. Technol.* **6** 015001

[26] Yordanov Y S, Arvidsson-Shukur D R M and Barnes C H W 2020 Efficient quantum circuits for quantum computational chemistry *Phys. Rev.* A **102** 062612

[27] Salis G and Moll N 2019 Short-depth trial-wavefunctions for the variational quantum eigensolver based on the problem Hamiltonian (arXiv:1908.09533)
[28] Khamoshi A, Evangelista F A and Scuseria G E 2020 Correlating AGP on a quantum computer *Quantum Sci. Technol.* **6** 014004
[29] Nielsen M A and Chuang I 2002 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[30] Note that in a field-free lab frame, only the total particle number $\langle \hat{N}_T \equiv \hat{N}_\alpha + \hat{N}_\beta \rangle$ and the total spin $\langle \hat{S}^2 \rangle$ are distinguishable quantum numbers. The high-spin projection particle number $\langle \mathrm{N}_\Delta \equiv \mathrm{N}_\alpha - \mathrm{N}_\beta \propto \hat{S}_z \rangle$ is only resolvable in presence of external fields. However, as $\langle N_\alpha \rangle$ and $\langle N_\beta \rangle$ are relatively easy to separately constrain, we will require that the entangler circuits respect all three quantum number symmetries $\langle \hat{N}_\alpha \rangle$, $\langle \hat{N}_\beta \rangle$, and $\langle \hat{S}_2 \rangle$ (and thereby also $\langle N_\mathrm{T} \rangle$ and $\langle N_\Delta \rangle$) throughout this manuscript.
[31] Brandão F G S L, Harrow A W and Horodecki M 2016 Local random quantum circuits are approximate polynomial-designs *Commun. Math. Phys.* **346** 397–434
[32] Oszmaniec M, Sawicki A and Horodecki M 2020 Epsilon-nets, unitary designs and random quantum circuits (arXiv:2007.10885)
[33] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 Quantum circuit learning *Phys. Rev.* A **98** 032309
[34] Farhi E and Neven H 2018 Classification with quantum neural networks on near term processors (arXiv:1802.06002)
[35] Schuld M, Bergholm V, Gogolin C, Izaac J and Killoran N 2019 Evaluating analytic gradients on quantum hardware *Phys. Rev.* A **99** 032331
[36] Bergholm V *et al* 2018 Pennylane: automatic differentiation of hybrid quantum-classical computations (arXiv:1811.04968)
[37] Kottmann J S, Anand A and Aspuru-Guzik A 2021 A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers *Chem. Sci.* **12** 3497
[38] Mari A, Bromley T R and Killoran N 2021 Estimating the gradient and higher-order derivatives on quantum hardware *Phys. Rev.* A **103** 012405
[39] Keller S, Boguslawski K, Janowski T, Reiher M and Pulay P 2015 Selection of active spaces for multiconfigurational wavefunctions *J. Chem. Phys.* **142** 244104
[40] McClean J R *et al* 2020 OpenFermion: the electronic structure package for quantum computers *Quantum Sci. Technol.* **5** 034014
[41] Reck M, Zeilinger A, Bernstein H J and Bertani P 1994 Experimental realization of any discrete unitary operator *Phys. Rev. Lett.* **73** 58
[42] Wecker D, Hastings M B, Wiebe N, Clark B K, Nayak C and Troyer M 2015 Solving strongly correlated electron models on a quantum computer *Phys. Rev.* A **92** 062318
[43] Google AI Quantum and Collaborators 2020 Hartree–Fock on a superconducting qubit quantum computer *Science* **369** 1084–9
[44] Tsuchimochi T, Mori Y and Ten-no S L 2020 Spin-projection for quantum computation: a low-depth approach to strong correlation *Phys. Rev. Res.* **2** 043142
[45] Lacroix D 2020 Symmetry-assisted preparation of entangled many-body states on a quantum computer *Phys. Rev. Lett.* **125** 230502
[46] Arute F *et al* 2020 Hartree–Fock on a superconducting qubit quantum computer *Science* **369** 1084–9
[47] Barron G S, Gard B T, Altman O J, Mayhall N J, Barnes E and Economou S E 2020 Preserving symmetries for variational quantum eigensolvers in the presence of noise (arXiv:2003.00171)
[48] Grant E, Wossnig L, Ostaszewski M and Benedetti M 2019 An initialization strategy for addressing barren plateaus in parametrized quantum circuits *Quantum* **3** 214
[49] Setia K, Chen R, Rice J E, Mezzacapo A, Pistoia M and Whitfield J D 2020 Reducing qubit requirements for quantum simulations using molecular point group symmetries *J. Chem. Theory Comput.* **16** 6091–7
[50] Nam Y *et al* 2020 Ground-state energy estimation of the water molecule on a trapped-ion quantum computer *npj Quantum Inf.* **6** 33
[51] Li J, Yang X, Peng X and Sun C-P 2017 Hybrid quantum-classical approach to quantum optimal control *Phys. Rev. Lett.* **118** 150503
[52] Banchi L and Crooks G E 2021 Measuring analytic gradients of general quantum evolution with the stochastic parameter shift rule *Quantum* **5** 386
[53] Jakob Meyer J, Borregaard J and Eisert J 2020 A variational toolbox for quantum multi-parameter estimation (arXiv:2006.06303)

6 Optimizing the information extracted by a single qubit measurement

Bibliographic Information

Polla, S., **Anselmetti, G.-L. R.**, O'Brien, T. E (2022, July). Optimizing the information extracted by a single qubit measuremen. 10.48550/arXiv.2207.094. arXiv:2207.0947.

Summary

The error mitigation techniques discussed, simulated and benchmarked on hardware in the previous chapters are formulated in terms of the Hadamard test, which only extracts one bit of information per measurement. We optimize the estimation of the expectation value in this measurement settings of an operator by its linear decomposition into bitwise-measurable terms. We prove that optimal decompositions must be in terms of reflections with eigenvalues ±1. We find the optimal reflection decomposition of a fast-forwardable operator and construct a circuit representation of it using quantim signal processing. By running large scale numerical simulations of both the analytical terms and the decomposition into a quantum signal processing circuit we show a numerical improvement over a simple Pauli decomposition by a factor $N^{0.7}$.

Contributions

The author contributed the numerical implementation of the ideas of the work. (1) Parallelized code was written to simulate the measurement of different observables on up to 17 qubits and quantum signal processing depths of up to 20/40 repetitions in the circuit approximation of the decomposition and the analytical expectation values. (2) Simulations were run on an in house high-performance cluster, filling nodes with 32 CPUs and 670GB of RAM, investigating the behaviour outside of the conjectured optimal performance of observables with states with more support than two eigenvalues, extracting scaling parameters. (3) Furthermore writing the manuscript parts concerned with the numerical simulations. (4) An open-source implementation of the core part of the paper in Pennylane [89] to allow better access towards the ideas was developed by the author and is available at `https://github.com/PennyLaneAI/pennylane/pull/2852`.

Optimizing the information extracted by a single qubit measurement

Stefano Polla,[1, 2] Gian-Luca R. Anselmetti,[3] and Thomas E. O'Brien[1, 2]
[1] *Google Quantum AI, 80636 München, Germany*
[2] *Instituut-Lorentz, Universiteit Leiden, 2300 RA Leiden, The Netherlands*
[3] *Covestro Deutschland AG, Leverkusen 51373, Germany*
(Dated: July 21, 2022)

We consider a quantum computation that only extracts one bit of information per quantum state preparation. This is relevant for error mitigation schemes where the remainder of the system is measured to detect errors. We optimize the estimation of the expectation value of an operator by its linear decomposition into bitwise-measurable terms. We prove that optimal decompositions must be in terms of reflections with eigenvalues ±1. We find the optimal reflection decomposition of a fast-forwardable operator, and show a numerical improvement over a simple Pauli decomposition by a factor $N^{0.7}$.

The largest bottleneck in quantum algorithm design is the encoding and decoding of a quantum state. Although each full characterization of a quantum state requires an exponentially large amount of information, direct tomography of an N-qubit quantum state ρ extracts only N bits of information, and collapses ρ to a state described by those N bits alone — erasing any other information. Performing this repeatedly allows the estimation of an expectation value $\langle O\rangle := \mathrm{Tr}[O\rho]$ of any operator O that is diagonal in the measurement basis. The rate at which such a measurement converges is known as the standard quantum or shot noise limit - after M repeated preparations, $\langle O\rangle$ can be estimated with variance

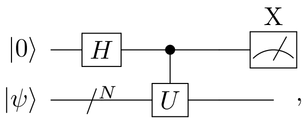
$$\mathrm{Var}[O] = M^{-1}\left(\langle O^2\rangle - \langle O\rangle^2\right). \tag{1}$$

Though this rate can be improved upon [1–4], doing so requires implementing long coherent circuits or performing large correlated measurements, which are not feasible in the current NISQ era [5].

Instead of using all N qubits to extract data from a quantum state, one may perform a partial measurement that extracts $< N$ bits, and use the remaining qubits to detect and mitigate errors [6–8]. Error mitigation is key in obtaining precise results from NISQ circuits, such as variational algorithms [9, 10], where the output of the quantum algorithm is a set of estimates of expectation values. Echo verification (EV - see Appendix B) [11–13] allows one to strongly mitigate errors in a wide class of algorithms, by recasting measurements as Hadamard tests. In each EV circuit, a single bit of information is extracted from the system register as a measurement, freeing up the remainder of the register for error detection/mitigation. One may combine results of multiple EV circuits (through classical post-processing) into an error-mitigated estimator of any target quantity. However, the stringent requirement that only one bit of information be extracted from the device further tightens the bottleneck of quantum-classical I/O.

In this paper we study how we can optimize information extraction from a quantum system to estimate the expectation value of an observable O, under the restriction that only a single bit of information is measured per state preparation. This matches the requirements of EV, the rest of the information being reserved for error mitigation. We do not focus in this work on the effectiveness of EV as an error mitigation strategy, and consider only the case of error-free quantum simulation. We define measurements with a single-bit outcome in terms of the Hadamard test, use these to construct an expectation value estimator for a more complicated operator via a linear decomposition, and calculate the variance of this resulting estimator. We prove necessary conditions for such a linear decomposition to be optimal; i.e. to minimize the cost of expectation value estimation. We construct a provably optimal (in some sense) decomposition for a fast-forwardable operator, and give a general (albeit expensive) method to implement this decomposition through quantum signal processing [14–16]. We analyse our methods numerically, comparing the variance of estimators based on our optimal method with other known approaches such as Pauli decompositions and the Dirichelet kernel measurements introduced in [17]. We find an asymptotic improvement between our optimal decomposition and a simple Pauli decomposition of a factor $N^{0.7}$, which at 13 qubits gives already an order of magnitude improvement.

The Hadamard test — To motivate this work, we give a brief outline of the Hadamard test (HT). An operation U on a N-qubit "system" register in the state $|\psi\rangle$ is controlled by a control qubit in the $|+\rangle$ state. As a quantum circuit this can be written



,

and the resulting state can be easily calculated to be

$$\frac{1}{\sqrt{2}}\Big(|0\rangle|\psi\rangle + |1\rangle U|\psi\rangle\Big). \tag{2}$$

Tracing out the system register then yields the following reduced density matrix on the control qubit,

$$\frac{1}{2}\begin{pmatrix} 1 & \langle U\rangle \\ \langle U\rangle^* & 1 \end{pmatrix}. \tag{3}$$

One may estimate $\langle U\rangle$ by measuring the control qubit in the X basis, which estimates $\langle\frac{1}{2}(U+U^\dagger)\rangle$.

Any generalized measurements that outputs a single bit of information is equivalent to a Hadamard test (in fact, the HT can be seen as the Naimark dilation of any binary-outcome POVM, see App. A). The direct (ancilla-based) measurement via the HT may often be replaced by an indirect measurement using an altered circuit [11, 18, 19], allowing control-free implementations of these single-bit measurements. Extracting a single bit of information allows to proceed in processing the quantum information remaining in the state register. For instance, inverting the unitary that prepared $|\psi\rangle$ and measuring in the computational basis yields a powerful error mitigation technique, echo verification [11–13]. (See App. B for details on this technique and its name.) In another example, the HT may be used to estimate the gradient of a cost function with respect to a variational term $\exp(iA\theta)$ in a circuit, as $\frac{d}{d\theta}\exp(iA\theta)=iA\exp(iA\theta)$ [20, 21]. Both these methods require operating on the system register after the HT (or its equivalent) is performed, preventing furhter information extraction. For the specific case of EV, we show in App. B 2 that extracting more than one bit of information is counterproductive.

The Hadamard test differs from the projective measurement of $\mathrm{Re}(U):=\frac{1}{2}(U+U^\dagger)$ (the Hermitian part of U). Each instance of the Hadamard test can only output $+1$ or -1, whereas the spectrum of $\mathrm{Re}(U)$ can have up to 2^N distinct eigenvalues in the range $[-1,1]$. This has a direct impact on the estimation uncertainty: performing the Hadamard test M times and measuring the control qubit in the X-basis yields an estimator of $\langle\mathrm{Re}(U)\rangle=\mathrm{Re}(\langle U\rangle)$ with a variance

$$\mathrm{Var}^*\left[\langle\mathrm{Re}(U)\rangle\right]=\frac{1-\langle\mathrm{Re}(U)\rangle^2}{M}, \tag{4}$$

which can be seen to be strictly larger than the variance one would obtain by performing a projective measurement of $\mathrm{Re}(U)$ on M copies of $|\psi\rangle$ [Eq. (1)],

$$\mathrm{Var}\left[\langle\mathrm{Re}(U)\rangle\right]\leq\mathrm{Var}^*\left[\langle\mathrm{Re}(U)\rangle\right], \tag{5}$$

as $\langle\mathrm{Re}(U)^2\rangle\leq 1$. Our goal is to optimize estimators of expectation values $\langle O\rangle$ of a given operator, which use data from multiple HTs with different unitaries U [each with the given variance Eq. (4)], and assuming one test per state preparation. We want to minimize the total number of state preparations (distributed over different choices of U) needed to achieve an estimator of $\langle O\rangle$ with error smaller than a fixed ϵ.

Operator decompositions — It is common in quantum computing to estimate the expectation value of a complex operator O by writing O as a linear combination of simpler terms (a.k.a. a *decomposition*) which have their expectation values estimated independently [9, 22, 23]. In this work, we make use of this method, and consider estimating these simpler terms via Hadamard tests. Let us fix a decomposition[1] X,

$$O=\sum_{x\in X}c_x\,\mathrm{Re}(U_x)\leftrightarrow\langle O\rangle=\sum_{x\in X}c_x\langle\mathrm{Re}(U_x)\rangle, \tag{6}$$

and consider estimating $\langle O\rangle$ by estimating each $\langle\mathrm{Re}(U_x)\rangle$ independently and summing the results. As $\mathrm{Re}(U_x)$ and O are Hermitian operators we may assume c_x to be real without loss of generality, and we may further assume $c_x\geq 0$ by absorbing a minus sign onto U_x. Note that the arrow in Eq. (6) points both ways as the set of expectation values on all states $|\psi\rangle$ uniquely defines an operator.

Once a suitable decomposition X of an operator O [Eq. (6)] has been chosen, to calculate the total cost of the algorithm we must allocate a number m_x of repeated single-shot HT experiments to estimate individual $\langle\mathrm{Re}(U_x)\rangle$. We assume a single-bit measurement per state preparation, i.e. each HT requires resetting the circuit and re-preparing $|\psi\rangle$, and the total number of re-preparations $M_X=\sum_{x\in X}m_x$ is the relevant cost of implementing our measurement scheme. If each $\langle\mathrm{Re}(U_x)\rangle$ is estimated independently, the variance on a final estimate of $\langle O\rangle$ can be calculated by standard propagation of variance

$$\mathrm{Var}^*_X\left[\langle O\rangle\right]=\sum_{x\in X}c_x^2\,\mathrm{Var}^*\left[\langle\mathrm{Re}(U_x)\rangle\right] \tag{7}$$

$$=\sum_{x\in X}\frac{c_x^2(1-\langle\mathrm{Re}(U)\rangle^2)}{m_x}. \tag{8}$$

Eq. (5) implies that under the same decomposition of O

$$\mathrm{Var}_X\left[\langle O\rangle\right]:=\sum_{x\in X}c_x^2\,\mathrm{Var}[\langle\mathrm{Re}(U_x)\rangle]\leq\mathrm{Var}^*_X\left[\langle O\rangle\right], \tag{9}$$

for all states ρ.

Adaptive shot allocation — Given a decomposition X and a total shot budget M_X, an optimal choice for the m_x may be found using Lagrange multiplier methods [24]

$$m_x=M_X\frac{c_x\sqrt{1-\langle\mathrm{Re}(U_x)\rangle^2}}{\sum_{y\in X}c_y\sqrt{1-\langle\mathrm{Re}(U_y)\rangle^2}}, \tag{10}$$

recalling that $c_x\geq 0$. This yields a bound on the required

---

[1] In a slight abuse of notation, throughout this work we will use the same label (e.g. X) to represent the entire linear decomposition defined by the set $\{c_x,U_x\}$ in Eq. 6, and the set of labels x that we sum over.

M_X to estimate $\langle O\rangle$ with $\mathrm{Var}^*\left[\langle O\rangle\right]=\epsilon^2$

$$M_X \geq \mathcal{M}_X := \epsilon^{-2}\left[\sum_{x\in X} c_x\sqrt{1-\langle\mathrm{Re}(U_x)\rangle^2}\right]^2. \qquad (11)$$

We call $\mathcal{M}_X$ the *cost* of the decomposition X. This may be compared to well-known results for measurement bounds using standard tomography methods [8, 23–25] by substituting Var^* for Var in Eq. (7). Though exact values of $\langle U_x\rangle$ will not be known in advance, these can be estimated using a small initial fraction of measurements before a final distribution of measurements is allocated.

Norm-preserving decompositions — we have shown above how to optimize measurement allocation given a linear decomposition X [Eq. (6)]. Let us now consider how to optimize X to minimize Eq. (11).

We first consider the effect of possible rescalings of $\mathrm{Re}(U_x)$. If any term $c_x\,\mathrm{Re}(U_x)$ has $\|\,\mathrm{Re}(U_x)\| < 1$,[2] one can find some unitary $U_{x'}$ for which $\mathrm{Re}(U_{x'}) = \mathrm{Re}(U_x)/\|\,\mathrm{Re}(U_x)\|$; substituting $U_x\rightarrow U_{x'}$ (and $c_x\rightarrow c_{x'}$ accordingly) will always improve the bound in Eq. (11). (For now we do not worry about how the unitaries may be implemented as quantum circuits; we will consider this issue later.)

One may next consider sub-dividing individual terms $\mathrm{Re}(U_x)$ of X, by writing

$$c_x\,\mathrm{Re}(U_x) = c_{x,0}\,\mathrm{Re}(U_{x,0}) + c_{x,1}\,\mathrm{Re}(U_{x,1}), \qquad (12)$$

where $U_{x,0}$ and $U_{x,1}$ are both unitary, and $c_x, c_{x,0}, c_{x,1} > 0$. As we can assume $\|\,\mathrm{Re}(U_x)\| = 1$, such a decomposition requires $c_{x,0}+c_{x,1}\geq c_x$, to preserve the spectral norm of $\mathrm{Re}(U_{x,0})$ and $\mathrm{Re}(U_{x,1})$. When this inequality is saturated, we call the sub-decomposition *norm-preserving*. It turns out that this condition is sufficient for the sub-decomposition to be non-increasing in the cost $\mathcal{M}$ of estimation [Eq. (11)], for all states $|\Psi\rangle$; formally:

Lemma 1 *Given a linear decomposition X of a target operator O [Eq. (6)], a sub-decomposition X' [Eq. (12)] that is norm-preserving has non-increasing cost, $\mathcal{M}_{X'}\leq\mathcal{M}_X$ [Eq. (11)], for any state $|\Psi\rangle$.*

We give a proof of this lemma in Appendix C 1

We would like to extend the above lemma to a statement that norm-increasing subdecompositions of a linear decomposition X are always suboptimal in some sense. To achieve this, note that as a corollary to lemma 1, we can improve on all terms $c_x\,\mathrm{Re}(U_x)$ in a linear decomposition X by a norm-preserving identity shift

$$c_x\,\mathrm{Re}(U_x) = c_x(1-\bar{\lambda}_x)\,\mathrm{Re}(U_{\tilde{x}}) + c_x\bar{\lambda}_x\mathbb{1}, \qquad (13)$$

where $\bar{\lambda}_x = \frac{1}{2}(\lambda_x^{\min}+\lambda_x^{\max})$, $\lambda_x^{\min}$ and $\lambda_x^{\max}$ are the lowest and highest eigenvalues of $\mathrm{Re}(U_x)$ respectively, and $\mathrm{Re}(U_{\tilde{x}})$ has the same eigenvectors of $\mathrm{Re}(U)$ (with its spectrum shifted and rescaled). We call the outcome decomposition $\tilde{X}$ of the procedure above the *center* of X. Though a norm-increasing subdecomposition of X may not be suboptimal relative to X, it is suboptimal relative to this center:

Lemma 2 *Let X be a linear decomposition of O with all $\|\mathrm{Re}(U_x)\| = 1$; let $\tilde{X}$ be the center of X and let X' be a strictly norm-increasing sub-decomposition. There exists at least one state $|\Psi\rangle$ for which the cost $\mathcal{M}_{\tilde{X}} < \mathcal{M}_{X'}$.*

We give a proof of this lemma in Appendix C 2.

To recap, the above two lemmas show a) that norm-preserving sub-decompositions do not increase the cost of estimating expectation values via Hadamard tests on any given state, and b) norm-increasing sub-decompositions not only can increase expectation value estimation costs on some states, but are guaranteed to do so on at least one. This result is in direct contrast to standard expectation value estimation, where independent estimation of $\langle A\rangle$ and $\langle B\rangle$ is sub-optimal to joint estimation of $\langle A+B\rangle$ whenever the latter is possible. This suggests a path towards optimizing HT expectation value estimation, by repeatedly dividing terms $\mathrm{Re}(U_x)$ in a norm-preserving manner, until no further sub-decomposition can reduce the cost any state. It turns out that not all choices of division lead to the same end-point, however all end points of this procedure have one common property (proven in Appendix C 3):

Lemma 3 *A decomposition X of an operator O has no non-trivial norm-preserving sub-decompositions if and only if all operators $\mathrm{Re}(U_x)$ in X are reflections: $\mathrm{Re}(U_x)^2 = 1$.*

It should be no surprise that we find reflection operators $\mathrm{Re}(U_x)^2 = 1$ to be a crucial ingredient to optimize HT tomography, as these are the only operators that saturate the bound in Eq. (5) for all states $|\Psi\rangle$. We call a decomposition X that consists of reflection operators only a *reflection decomposition*. We give some simple examples of these in Appendix C 4.

Optimizing reflection decompositions — Above we demonstrated that, for a decomposition X of an operator O to be optimal with regards to the cost $\mathcal{M}_X$ of estimating expectation values on a set of states (Eq. 11), all terms in X must be reflection operators. Otherwise, we demonstrated a means of sub-dividing single terms in the distribution to generate a new distribution with lower cost. However, this is not to say that all reflection decompositions X have the same cost $\mathcal{M}_X$. (These two statements are consistent as we cannot transform between reflection decompositions using subdivision.) The set of reflection decompositions of O form a convex set that is 2^{2^N-N}-dimensional if all U_x are diagonal in the

---

[2] Unless stated otherwise, all norms in this work are the spectral norm.

eigenbasis of O. This raises two questions: is there an optimal decomposition amongst the set of reflection decompositions, and does it achieve the von Neumann bound [Eq. (9)]?

Lemma 4 *Let O be an operator and Π_j be projectors onto the eigenvalues of O; $O\Pi_j = \Pi_j O = \lambda_j \Pi_j$. The* Ξ-decomposition *of O, given by*

$$O = \frac{\lambda_0 + \lambda_J}{2}\mathbb{1} + \sum_{x=1}^{J-1} \frac{\delta\lambda_x}{2}\Xi_x \tag{14}$$

$$\Xi_x = \mathbb{1} - \sum_{j<x} 2\Pi_j\,, \quad \delta\lambda_x = \lambda_x - \lambda_{x-1}, \tag{15}$$

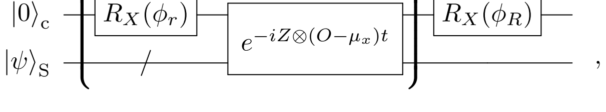
uniquely achieves the bound $\mathrm{Var}^*_\Xi[O] = \mathrm{Var}[O]$ *on all states $|\Psi\rangle$ with support on up to two eigenstates of O. No such decomposition achieves this bound on all states $|\Psi\rangle$ with support on three or more eigenstates of O.*

We prove this lemma in Appendix C 5. Note that the Ξ-decomposition can be immediately restricted to any subspace of the full-2^N-dimensional Hilbert space containing $|\Psi\rangle$ (i.e. if we knew that due to a symmetry or by virtue of being a low-energy state, $|\Psi\rangle$ had support only on such a space), and the optimality result still holds. This implies in turn that no linear decomposition X can achieve the von Neumann variance bound even for as small as a 3-dimensional subspace. This makes sense, as our restriction to measure one bit of information per state preparation forms a bottleneck with respect to the 3 nonzero-probability outcomes of a Von Neumann measurement on this space.

In order to realize the Ξ-decomposition estimator, we need to implement HT circuits that (approximately) estimate $\langle \Xi_x \rangle$. This may be achieved by realising that

$$\Xi_x = \mathrm{sgn}[O - \mu_x]\,, \quad \mu_x = \frac{\lambda_{x-1} + \lambda_x}{2}, \tag{16}$$

where sgn is the sign function. An approximation of this unitary operator can then be realized using quantum signal processing (QSP) [14–16] of the sign function [26], requiring only one additional ancillary qubit. The QSP circuit is given by



,

where $R_X(\phi_r) = e^{-i\frac{X}{2}\phi_r}$, implements a unitary block encoding $Q_{\boldsymbol{\phi}}$ of a degree-R trigonometric polynomial $S_{\boldsymbol{\phi}}$ of the operator $(O - \mu_x)t$:

$$\langle 1|_c Q_{\boldsymbol{\phi}} |0\rangle_c = \sum_{r=0}^{R} c_r(\boldsymbol{\phi}) e^{-ir(O-\mu_x)t} := S_{\boldsymbol{\phi}}[(O - \mu_x)t]. \tag{17}$$

Here, $\boldsymbol{\phi}$ is a vector containing the individual angles ϕ_r implemented during the QSP circuit. We can then sample $\langle \mathrm{Re}\{S_{\boldsymbol{\phi}}[(O - \mu_x)t]\}\rangle$ through HT (or EV), using another qubit controlling all gates in the QSP circuit. To approximate Eq. (16) with our block-encoded operator $S_{\boldsymbol{\phi}}$, we must choose $t < \frac{\pi}{\|O-\mu_x\|}$ to avoid aliasing, and find the optimal $\boldsymbol{\phi}$

$$\boldsymbol{\phi} = \underset{(\phi_r=-\phi_{R-r})}{\arg\min} \int_{0+\delta}^{\pi-\delta} d\omega \Big[\mathrm{sgn}(\omega) - \mathrm{Im}[S_\phi(\omega)]\Big]. \tag{18}$$

Here, the constraint $\phi_r = -\phi_{R-r}$ ensures $\mathrm{Im}[S_\phi(\omega)]$ is an odd function of ω. A resolution parameter $\delta \geq 0$ can be introduced to improve the approximation away from the nodes $\omega = \{0, \pm\pi\}$ of $S_\phi(\omega)$. In Appendix D we give further details of this decomposition, and analyse the approximation error numerically. We find that this error converges exponentially in the number of circuit blocks R.

Numerical results — To investigate performance of various decompositions on states that have support on more than two eigenstates of O, and therefore are not covered by Lemma 4, we perform numerical simulations using random variationally-generated states and a simple toy operator $O = \sum_j Z_j$. (In appendix F, we report this scaling for other systems.) We measure the variances on states generated by a hardware-efficient ansatz [27] with random input parameters using PennyLane [28]. For each datapoint 100 random states are generated. We consider estimating $\langle O \rangle$ in a realistic scenario where the $\langle \mathrm{Re}(U_x)\rangle$ values will not be known in advance to optimally choose m_x via Eq. (10). Instead, for each random state we generate a prior estimate of each $\langle \mathrm{Re}(U_x)\rangle$ from 10^5 measurements of the state, and use these to determine m_x (which are then only approximately optimal). This leaves the total shot count M_X as a free parameter; we resolve this in Fig. 1 by calculating $M_X \mathrm{Var}^*_X[\langle O \rangle]$. (This gives a quantity that is relevant regardless of the number of the shots actually used to estimate $\langle O \rangle$.)

An average of $M_X \mathrm{Var}^*_X[\langle O \rangle]$ over the 100 states is formed and plotted in Fig. 1 for each grouping method. This is compared to the Von Neumann measurement variance $\mathrm{Var}[O]$, which does not require any shot allocation, and sets a lower limit to the other estimators [see App. C 5, Eq. (C15)]. The Ξ-decomposition [orange, 'Ξ'] has the best asymptotic scaling of all decompositions, being suboptimal to $\mathrm{Var}[O]$ by a factor $\approx n^{1/3}$. The QSP approximation of Ξ, [teal, 'SGN'], has a slightly worse asymptotic scaling, which we associate to the error in approximating $\mathrm{sgn}(O - \mu_j)$. At the largest considered $N = 13$, these two decompositions suffer approximately a factor 2 penalty in their total cost compared to $\mathrm{Var}[\langle O \rangle]$. The generalized parameter-shift kernel decomposition [17] [green, 'GPSK', described in Appendix E] has the worst overall performance out of the investigated estimators, due to the constant factor. It has however a better asymptotic scaling than a simple Pauli decomposition $U_x = Z_j$ [red, 'Pauli', Appendix C 4]. In

Appendix F we investigate the scaling of different sets of observables. We observe that the order of the performance of the different decompositions remains consistent throughout, but the relative gains and losses in performance can be significantly different.

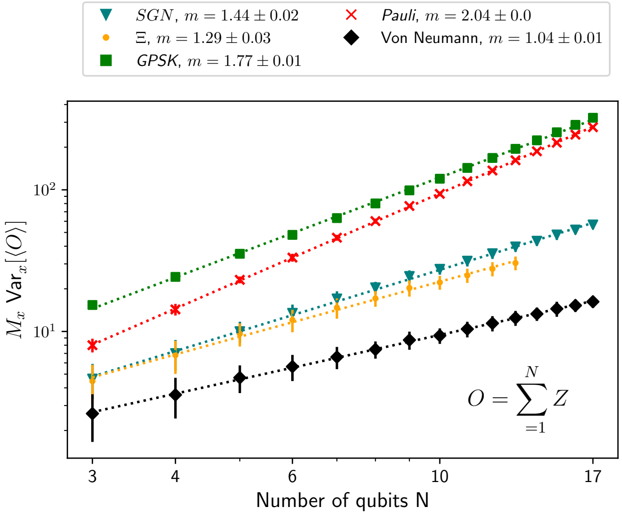


FIG. 1: Comparison study of variances of different decompositions on random states generated by a hardware-efficient ansatz (see text for details). Different colours correspond to different decompositions [Eq.(6)] of the target operator O (see text for the description of all decompositions). Dashed lines are power-law fits to the data (obtained exponents are given in legend).

Conclusion — In this work we studied the optimization of expectation value estimation for a quantum state in the case where we are only allowed to measure a single qubit per state preparation (e.g. through Hadamard tests, with relevant application to echo verification). We calculated the cost of estimating the expectation value of an operator O by linearly decomposing O into a linear combination of sub-unitary terms, assuming an optimal shot allocation. We demonstrated that this cost is strictly non-increasing when terms are further subdivided, under the constraint that this subdivision preserves the induced 1-norm of the term coefficients. We showed that the end-points of this procedure of repeated division are linear decompositions of O where all terms are reflection operators; a so-called 'reflection decomposition'. We identified one such decomposition, the Ξ-decomposition, as unique in its ability to estimate $\langle O \rangle$ with a variance matching the Von Neumann measurement limit on any linear combination of up to 2 eigenstates of O. We demonstrated how the Ξ-decomposition may be approximately implemented through quantum signal processing. Numerical results demonstrate that on simple systems, the Ξ-decomposition and its approximate couterpart demonstrate clear constant and asymptotic improvements over other reflection decompositions (in the cost of estimating $\langle O \rangle$ on random states), with up to a factor 10× improvement for estimation on 20 qubits.

Though these results are encouraging, the significant discrepancy between $\mathrm{Var}^*_\Xi[O]$ and $\mathrm{Var}[O]$ is worrying for NISQ algorithms that already incur a significant cost to tomograph complex Hamiltonians [6, 8, 23, 29–31]; either one incurs a large overhead for measurement due to the need to invoke quantum signal processing or incur the clear asymptotic scaling cost that comes with measuring single Pauli terms per state preparation. Given that echo verification has a sampling cost scaling as $1/F^2$ (for a circuit fidelity F) [11], this result adds to the unlikelihood of beyond-classical NISQ variational algorithms in chemistry. Finding reflection decompositions with lower circuit depth is a clear avenue for future work.

Acknowledgments

The authors wish to acknowledge Fotios Gkritsis for providing shot allocation code, and William Huggins, Ryan Babbush, Jordi Tura, Alicja Dutkiewicz, David Wierichs and Christian Gogolin for useful discussions and advice on this work. Covestro acknowledges funding from the German Ministry for Education and Research (BMBF) under the funding program quantum technologies as part of project HFAK (13N15630).

---

[1] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone, "Quantum-enhanced measurements: Beating the standard quantum limit," Science **306**, 1330–1336 (2004).
[2] B L Higgins, D W Berry, S D Bartlett, M W Mitchell, H M Wiseman, and G J Pryde, "Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements New Journal of Physics Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements," New Journal of Physics **11**, 73023 (2009).
[3] Emanuel Knill, Gerardo Ortiz, and Rolando D Somma, "Optimal quantum measurements of expectation values of observables," Physical Review A **75**, 012328 (2007).
[4] William J Huggins, Kianna Wan, Jarrod McClean, Thomas E. O'Brien, Nathan Wiebe, and Ryan Babbush, "Nearly Optimal Quantum Algorithm for Estimating Multiple Expectation Values," ArXiv:2111.09283 (2021).
[5] John Preskill, "Quantum Computing in the NISQ era and beyond," Quantum **2**, 79 (2018), arXiv:1801.00862v3.
[6] Xavier Bonet-Monroig, R Sagastizabal, M Singh, and T E O'Brien, "Low-cost error mitigation by symmetry verification," Physical Review A **98**, 62339 (2018).
[7] Sam McArdle, Xiao Yuan, and Simon Benjamin, "Error-

Mitigated Digital Quantum Simulation," Physical Review Letters **122**, 180501 (2019).
[8] William J Huggins, Jarrod R. McClean, Nicholas C. Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush, "Efficient and Noise Resilient Measurements for Quantum Chemistry on Near-Term Quantum Computers," npj Quant. Inf. **7** (2021).
[9] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor," Nature Communications **5**, 4213 (2014).
[10] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," New Journal of Physics **18**, 023023 (2016).
[11] Thomas E. O'Brien, Stefano Polla, Nicholas C. Rubin, William J Huggins, Sam McArdle, Sergio Boixo, Jarrod R. McClean, and Ryan Babbush, "Error Mitigation via Verified Phase Estimation," PRX Quantum **2** (2021).
[12] Zhenyu Cai, "Resource-efficient purification-based quantum error mitigation," ArXiv:2107.07279 (2021).
[13] Mingxia Huo and Ying Li, "Dual-state purification for practical quantum error mitigation," Physical Review A **105**, 022427 (2022).
[14] Guang Hao Low and Isaac L Chuang, "Optimal hamiltonian simulation by quantum signal processing," Physical review letters **118**, 010501 (2017).
[15] Guang Hao Low and Isaac L Chuang, "Hamiltonian simulation by qubitization," Quantum **3**, 163 (2019).
[16] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe, "Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (ACM, New York, NY, USA, 2019) pp. 193–204.
[17] David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin, "General parameter-shift rules for quantum gradients," Quantum **6**, 677 (2022).
[18] Kosuke Mitarai and Keisuke Fujii, "Methodology for replacing indirect measurements with direct measurements," Physical Review Research **1**, 013006 (2019).
[19] Aram W Harrow and John C Napp, "Low-depth gradient measurements can improve convergence in variational hybrid quantum-classical algorithms," Physical Review Letters **126**, 140502 (2021).
[20] Gian Giacomo Guerreschi and Mikhail Smelyanskiy, "Practical optimization for hybrid quantum-classical algorithms," ArXiv:1701.01450 (2017).
[21] Ying Li and Simon C Benjamin, "Efficient variational quantum simulator incorporating active error minimization," Physical Review X **7**, 021050 (2017).
[22] Gerardo Ortiz, J E Gubernatis, E Knill, and R Laflamme, "Quantum algorithms for fermionic simulations," Physical Review A **64**, 022319 (2001).
[23] Dave Wecker, Matthew B Hastings, and Matthias Troyer, "Progress towards practical quantum variational algorithms," Physical Review A **92**, 042303 (2015).
[24] Nicholas C. Rubin, Ryan Babbush, and Jarrod R. McClean, "Application of fermionic marginal constraints to hybrid quantum algorithms," New Journal of Physics **20**, 1–24 (2018).
[25] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F Izmaylov, "Measurement optimization in the variational quantum eigensolver using a minimum clique cover," J. Chem. Phys. **152**, 224109 (2020).
[26] Lin Lin and Yu Tong, "Near-optimal ground state preparation," Quantum **4**, 372 (2020).
[27] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," Nature **549**, 242–246 (2017).
[28] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, M. Sohaib Alam, Shahnawaz Ahmed, Juan Miguel Arrazola, Carsten Blank, Alain Delgado, Soran Jahangiri, Keri McKiernan, Johannes Jakob Meyer, Zeyue Niu, Antal Száva, and Nathan Killoran, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," (2018).
[29] Artur F. Izmaylov, Tzu-Ching Yen, and Ilya G. Ryabinkin, "Revising measurement process in the variational quantum eigensolver: Is it possible to reduce the number of separately measured operators?" Chem. Sci. **10**, 3746–3755 (2019).
[30] Ophelia Crawford, Barnaby van Straaten, Daochen Wang, Thomas Parks, Earl Campbell, and Stephen Brierley, "Efficient quantum measurement of pauli operators in the presence of finite sampling error," Quantum **5** (2021).
[31] Tzu-Ching Yen, Aadithya Ganeshram, and Artur F. Izmaylov, "Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates," ArXiv:2201.01471 (2022).

A. Hadamard tests are equivalent to binary POVMs

In this section we demonstrate the equivalence between postive-operator valued measurements (POVMs) with a binary outcome and Hadamard tests. We do this in two parts. We first find the positive operators representing a given HT. Then, for a given binary-outcome POVM, we explicitly construct the HT as a special case of Naimark dilation.

A Hadamard test on a system register s with state $|\psi\rangle_{\mathrm{s}}$, is defined by the control qubit c (initialized in the state $|+\rangle_{\mathrm{c}}$), a controlled unitary CU and a projective Pauli measurement X_{c} on the control qubit. We then define the measurement operators that represent the back-action of the measurement on the system register

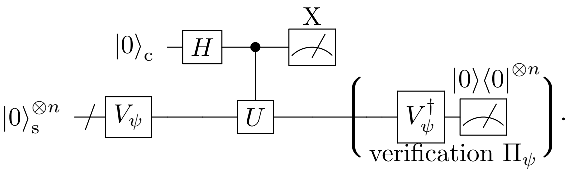
$$M_{\pm} = \langle \pm|_{\mathrm{c}} CU |+\rangle_{\mathrm{c}} = \frac{\mathbb{1} \pm U}{2}, \tag{A1}$$

and the relative positive operators $\Pi_{\pm} = M_{\pm}^{\dagger} M_{\pm}$ used to compute probabilities $p_{\pm} = \langle\psi|\Pi_{\pm}|\psi\rangle$ of measuring ±1 on the ancilla. The HT version is easily obtained by substituting $\langle\pm|$ for the relevant eigenstates of the chosen Pauli measurement.

Vice versa, any POVM with two outcomes (labeled ±) has positive operators $\Pi_+, \Pi_- > 0$ such that $\Pi_+ + \Pi_- = \mathbb{1}$. The operator $\Pi_+ - \Pi_-$, which is Hermitian and sub-unitary because both Π_+ and Π_- are Hermitian, positive, and sub-unitary ($\mathbb{1} - \Pi_\pm > 0$). An example of a unitary U satisfying $\mathrm{Re}(U) = \Pi_+ - \Pi_-$ is $\exp[i\arccos(O)]$. It is easy to check that the Hadamard test constructed from this unitary return the correct positive operators $\Pi_\pm$.

B. Echo verification

The name echo verification (EV) refers to a class of powerful error mitigation techniques [11–13], applicable in most algorithms that make use a Hadamard test to perform measurements on a system register. Originally this technique was called verified phase estimation [11], as it considered estimating expectation values of $\exp(iHt)$, with an archetypal application in the context to single-ancilla phase estimation. However, no phase estimation is necessary in the estimation of $\exp(iHt)$ (in our context, typically the phases are known), and we prefer the name echo verification due to the similarities to a Loschmidt echo. The key idea relies in exploiting for error mitigation the information left in the system register after the application of the controlled-unitary operator prescribed by the Hadamard test. This is done by "echoing" the preparation V_ψ of the system state $|\psi\rangle_\mathrm{s} = V_\psi|0\rangle_\mathrm{s}^{\otimes n}$, i.e. measuring on the N-qubit system register s whether the state remained unchanged or turned into an orthogonal state, using the following circuit



Let us denote the state of the combined system (system s plus control qubit c) after the controlled unitary as $|\Phi\rangle$, and let $\Pi_\psi = |\psi\rangle\langle\psi|$ be the projector on the state $|\psi\rangle_\mathrm{s}$ (which can be easily measured by inverting the state preparation V_ψ and verifying all system register bits return to $|0\rangle$). The estimate of $\langle\mathrm{Re}(U)\rangle$ can be obtained by measuring the operators $X^\mathrm{EV} := X \otimes \Pi_\psi$ on $|\Phi\rangle$, as opposed to $X_\mathrm{c} := X \otimes \mathbb{1}$. One can confirm that, in the absence of error, these operators have identical expectation values on the state at the end of the circuit [11]

$$\langle\Phi|X^\mathrm{EV}|\Phi\rangle = \langle\Phi|X_\mathrm{c}|\Phi\rangle = \langle\psi|\,\mathrm{Re}(U)|\psi\rangle. \tag{B1}$$

For an intuitive explanation, note that if the controlled unitary changes the state of the system register, the ancilla qubit must have been in the $|1\rangle$ state, and $\langle 1|X|1\rangle = 0$. This implies that the expectation value of $X \otimes (\mathbb{1}_N - \Pi_\psi)$ is 0.

1. Echo verification estimators

The estimator used for echo verification is not identical to the one studied in the main text, and so its variance is not quite identical. In particular, we have $(X^\mathrm{EV})^2 = \Pi_\psi$ which implies that the variance on an estimate of $\langle\Phi|X_\mathrm{EV}|\Phi\rangle$ is

$$\mathrm{Var}^*_\mathrm{EV}[\langle\mathrm{Re}(U)\rangle] = \frac{\langle\Phi|\mathbb{1}\otimes\Pi_\psi|\Phi\rangle - \langle\psi|\mathrm{Re}(U)|\psi\rangle^2}{M}. \tag{B2}$$

Clearly $|\langle\Phi|\psi\rangle|^2 \leq 1$, which implies $\mathrm{Var}^*_\mathrm{EV}[\mathrm{Re}(U)] \leq \mathrm{Var}^*[\mathrm{Re}(U)]$ [by comparison with Eq (4)]. In other words, the varianc of the EV estimator is always smaller or equal to the variance of the relative HT estimator. It is easy to calculate from the circuit above that

$$\langle\Phi|\Pi_\psi\otimes\mathbb{1}|\Phi\rangle = \frac{1}{2}|1 + \langle\psi|U|\psi\rangle|^2, \tag{B3}$$

(noting that $\langle\mathrm{Re}(U)\rangle = \mathrm{Re}(\langle U\rangle)$, which can be subtituted back into our variance estimate to obtain

$$\mathrm{Var}^*_\mathrm{EV}[\mathrm{Re}(U)] \geq \frac{1 - \langle\psi|\mathrm{Re}(U)|\psi\rangle^2}{2M}, \tag{B4}$$

Thus, we have

$$\mathrm{Var}^*[\langle\mathrm{Re}(U)\rangle] \geq \mathrm{Var}^*_\mathrm{EV}[\langle\mathrm{Re}(U)\rangle] \geq \frac{\mathrm{Var}^*[\langle\mathrm{Re}(U)\rangle]}{2}. \tag{B5}$$

This justifies our focus in the main text on optimizing the estimator from a standard Hadamard test; this estimator is simpler to analyse, more general, and differs from the EV estimator (that motivated this work) by at most a factor 2.

2. Parallelizing EV estimation

In absence of echo verification, we can trivially parallelize Hadamars tests measuring K commuting operators $\{\mathrm{Re}(U_0), ..., \mathrm{Re}(U_{K-1})\}$ using K ancillary qubits, one controlling each U_k. If each U_k is controlled by a separate ancillary qubit (labeled k, where C_kU_k represents the k-th unitary controlled by the k-th control qubit), the combined state of the system register s and ancillary qubits after all the unitaries are applied will be

$$\bigotimes_k C_kU_k|+\rangle_k|\psi\rangle. \tag{B6}$$

The probabilities of obtaining ±1 when measuring X on the j-th control qubit are

$$p_{j\pm} = \left\| \bigotimes_{k\neq j} C_k U_k |+\rangle_k \frac{1 \pm U_j}{2} |\psi\rangle \right\|^2 \tag{B7}$$

$$= \frac{1}{4} \langle\psi|(1 \pm U_j^\dagger)(1 \pm U_j)|\psi\rangle \tag{B8}$$

which coincides with the probabilities of a single Hadamard test with unitary U_j.

When performing echo verification, parallelization is more complicated. The result of verification (the measurement of $\Pi_\psi = |\psi\rangle\langle\psi|$ on the system register) is affected by all the controlled-U_k, and thus its result cannot be simply associated to one specific ancilla being in the state $|1\rangle$. To mitigate errors, all the cases in which the register is found in a state orthogonal to $|\psi\rangle$ should be considered as null towards all of the ancilla measurement results. The echo-verified probability of measuring the binary string $\vec{\sigma} = (\sigma_0, ..., \sigma_k)$, where each σ_k is ±1 corresponding to the state $|\pm\rangle$ measured on the k-th ancilla, is then

$$p_{\vec{\sigma}}^{\mathrm{EV}} = \left| \langle\psi| \prod_k \langle\sigma_k|_k C_k U_k |+\rangle_k |\psi\rangle \right|^2$$

$$= \frac{1}{4^K} \left| \langle\psi| \prod_k (1 + \sigma_k U_k) |\psi\rangle \right|^2 . \tag{B9}$$

The product in this equation can be then developed into a linear combination of 2^K expectation values (note that, as all U_k commute, the order does not matter). Under the assumption that all these expectation values are real [granted if $U_k = \mathrm{Re}(U_k)$] Eq. (B9) defines a quadratic system of 2^K equations with $2^K - 1$ unknowns[3]. Solving such system we find that the expectation value of a single $\mathrm{Re}(U_j)$ can be estimated by processing the sampled $p_{\vec{\sigma}}^{\mathrm{EV}}$ as

$$\langle \mathrm{Re}(U_j) \rangle = \underbrace{\left(\sum_{\vec{\sigma}:\sigma_j=+1} \sqrt{p_{\vec{\sigma}}^{\mathrm{EV}}} \right)^2}_{p_{j+}^{\mathrm{EV}}} - \underbrace{\left(\sum_{\vec{\sigma}:\sigma_j=-1} \sqrt{p_{\vec{\sigma}}^{\mathrm{EV}}} \right)^2}_{p_{j-}^{\mathrm{EV}}} , \tag{B10}$$

where we denoted $p_{j\pm}^{\mathrm{EV}}$ the terms that reproduce the probabilities that would be returned by a single, unparallelized EV experiment

$$p_{j\pm}^{\mathrm{EV}} = \frac{1}{4} \left| \langle\psi| 1 \pm U_j |\psi\rangle \right|^2 . \tag{B11}$$

We assume $p_{\vec{\sigma}}^{\mathrm{EV}}$ are sampled by averaging M shots of the parallel EV experiment. These are probabilities of mutually-exclusive measurements, thus the covariance matrix of the $p_{\vec{\sigma}}^{\mathrm{EV}}$ estimators is defined by

$$\mathrm{Var}[p_{\vec{\sigma}}^{\mathrm{EV}}] = \frac{1}{M} p_{\vec{\sigma}}^{\mathrm{EV}} (1 - p_{\vec{\sigma}}^{\mathrm{EV}}), \tag{B12}$$

$$\mathrm{Cov}[p_{\vec{\sigma}}^{\mathrm{EV}}, p_{\vec{\rho}}^{\mathrm{EV}}] = -\frac{1}{M} p_{\vec{\sigma}}^{\mathrm{EV}} p_{\vec{\rho}}^{\mathrm{EV}} \quad \text{if} \quad \vec{\sigma} \neq \vec{\rho}. \tag{B13}$$

We can then propagate the error through Eq. (B10) to obtain the variance on the parallel-EV (PEV) estimator of $\langle \mathrm{Re}(U_j) \rangle$

$$\begin{aligned} &M \, \mathrm{Var}^*_{\mathrm{PEV}}[\langle \mathrm{Re}(U_j) \rangle] = \\ &\quad \sum_{\vec{\sigma}} \frac{p_{j\sigma_j}^{\mathrm{EV}}}{p_{\vec{\sigma}}^{\mathrm{EV}}} p_{\vec{\sigma}}^{\mathrm{EV}} (1 - p_{\vec{\sigma}}^{\mathrm{EV}}) - \sum_{\vec{\sigma}\neq\vec{\rho}} \sigma_j \rho_j \frac{\sqrt{p_{j\sigma_j}^{\mathrm{EV}}}}{\sqrt{p_{\vec{\sigma}}^{\mathrm{EV}}}} \frac{\sqrt{p_{j\rho_j}^{\mathrm{EV}}}}{\sqrt{p_{\vec{\rho}}^{\mathrm{EV}}}} p_{\vec{\sigma}}^{\mathrm{EV}} p_{\vec{\rho}}^{\mathrm{EV}} \\ &= \sum_{\vec{\sigma}} p_{j\sigma_j}^{\mathrm{EV}} - \langle \mathrm{Re}(U_j) \rangle \\ &= 2^{K-1} (p_{j\sigma_+}^{\mathrm{EV}} + p_{j-}^{\mathrm{EV}}) - \langle \mathrm{Re}(U_j) \rangle. \end{aligned} \tag{B14}$$

which explodes exponentially with the size of the parallelization K.

More generally, we can compute the covariance matrix for all the $p_{j,\sigma_j}^{\mathrm{EV}}$ through error propagation

$$\mathrm{Var}[p_{j\sigma_j}^{\mathrm{EV}}] = p_{j\sigma_j}^{\mathrm{EV}} (2^{K-1} - p_{j\sigma_j}^{\mathrm{EV}}) \tag{B15}$$

$$\mathrm{Cov}[p_{j\sigma_j}^{\mathrm{EV}}, p_{k\rho_k}^{\mathrm{EV}}] = \delta_{j,k} \sqrt{p_{j\sigma_j}^{\mathrm{EV}} p_{k\rho_k}^{\mathrm{EV}}} - p_{j\sigma_j}^{\mathrm{EV}} p_{k\rho_k}^{\mathrm{EV}} \tag{B16}$$

[where the covariance assumes $(j, \sigma_j) \neq (k, \rho_k)$]. This shows that, increasing K, we effectively add to the covariance matrix a positive semi-definite term with a norm that scales exponentially in K. As all the decompositions Eq. (6) are ultimately to be estimated as linear combinations of the sampled probabilities $p_{j\sigma_j}^{\mathrm{EV}}$, parallelizing error verification is counterproductive.

C. Proof of decomposition optimality hierarchy

In this section we build up to the proof that the Ξ-decomposition is optimal in terms of cost (11), by proving the lemmas introduced in the main text. We first prove that a norm-preserving sub-decomposition has non-increasing cost with respect to its parent decomposition, for all states $|\psi\rangle$. We then prove that a sub-decomposition that does not have the norm-preserving property is always sub-optimal (i.e. it has strictly greater cost than an alternative norm-preserving sub-decomposition). The iteration of the norm-preserving sub-decomposition procedure leads to one of many alternative improving sequences of decompositions. The endpoint of each sequence is a norm-preserving linear decomposition of O for which all unitaries are reflection operators. Finally, we prove that one of such decompositions

---

[3] In the case of a more general $U = \mathrm{Re}(U) + i\,\mathrm{Im}(U)$, a similar system can be constructed by measuring each U_k and iU_k with $2K$ ancillas. Showing this is besides the scope of our work, and for the sake of simplicity we restrict ourselves to the case $U_k = Re(U_k)$.

(the Ξ-decomposition) achieves the Von-Neumann measurement variance bound on a certain set of states, and that no unbiased estimator based on single-qubit measurements can achieve this bound on a larger set of states.

1. Proof of Lemma 1, and corollaries

Given a linear decomposition X of an operator O [Eq. (6)], consider a norm-preserving sub-decomposition X' where a single term $x \in X$ is split according to Eq (12). The bound on the total number of shots Eq. (11) will then change:

$$\mathcal{M}_X \to \mathcal{M}_{X'} = \epsilon^{-2}\Bigg[\sum_{y\neq x} c_y\sqrt{1-\langle \mathrm{Re}(U_y)\rangle^2} + c_{x,0}\sqrt{1-\langle \mathrm{Re}(U_{x,0})\rangle^2} + c_{x,1}\sqrt{1-\langle \mathrm{Re}(U_{x,1})\rangle^2}\Bigg]^2. \tag{C1}$$

[with the change with respect to Eq.(11) being the second row]. This results in a reduction of the cost, as can be seen by calculating

$$\begin{aligned}
&c_x^2\left[1-\langle \mathrm{Re}(U_x)\rangle^2\right]\\
&= (c_{x,0}+c_{x,1})^2 - \big(c_{x,0}\langle \mathrm{Re}(U_{x,0})\rangle + c_{x,1}\langle \mathrm{Re}(U_{x,1})\rangle\big)^2\\
&= c_{x,0}^2\left[1-\langle \mathrm{Re}(U_{x,0})\rangle^2\right] + c_{x,1}^2\left[1-\langle \mathrm{Re}(U_{x,1})\rangle^2\right]\\
&\quad + 2c_{x,0}c_{x,1}\big(1-\langle \mathrm{Re}(U_{x,0})\rangle^2\langle \mathrm{Re}(U_{x,1})\rangle^2\big)\\
&\geq c_{x,0}^2\left[1-\langle \mathrm{Re}(U_{x,0})\rangle^2\right] + c_{x,1}^2\left[1-\langle \mathrm{Re}(U_{x,1})\rangle^2\right]\\
&\quad + 2c_{x,0}c_{x,1}\sqrt{\left[1-\langle \mathrm{Re}(U_{x,0})\rangle^2\right]\left[1-\langle \mathrm{Re}(U_{x,1})\rangle^2\right]}\\
&= \left[c_{x,0}\sqrt{1-\langle \mathrm{Re}(U_{x,0})\rangle^2} + c_{x,1}\sqrt{1-\langle \mathrm{Re}(U_{x,1})\rangle^2}\right]^2,
\end{aligned} \tag{C2}$$

where, in the center inequality we have used the fact that for $0 \leq a, b \leq 1$,

$$1 - ab \geq \sqrt{(1-a^2)(1-b^2)}. \tag{C3}$$

As a corollary and example, we look at identity shifts of a term $x \in X$. For $\mathrm{Re}(U_x)$ with unit norm, we can assume without loss of generality the largest eigenvalue is $\lambda_{\max} = 1$, and the smallest is $\lambda_{\min}$. We can then perform the simple norm-preserving decomposition

$$c_x \mathrm{Re}(U_x) = c_x(1-\bar\lambda)\mathrm{Re}(U_{x'}) + c_x\bar\lambda\mathbb{1} \tag{C4}$$

with $\bar\lambda = \frac{1}{2}(\lambda_{\min}+\lambda_{\max})$. The resulting $\mathrm{Re}(U_{x'})$ has maximum eigenvalue $+1$ and minimum eigenvalue -1, thus it does not admit non-trivial identity shift.

A norm-preserving sub-decomposition Eq. (12) of a term with $|\mathrm{Re}(U_x)| = 1$ will only admit terms with $|\mathrm{Re}(U_{x,i})| = 1$. (This can be checked by taking the expectation value of both sides of Eq. (12) on the eigenstate on which $|\langle \mathrm{Re}(U_x)\rangle| = 1$.) By the same reasoning, terms with $\mathrm{Re}(U_x)$ having maximum eigenvalue $+1$ and minimum eigenvalues -1 [like those obtained by the identity shifts Eq. (C4)] only admit sub-decompositions whose terms have the same property.

2. Proof of Lemma 2

In this appendix we compare the costs of two decompositions derived by an original decomposition X: the center $\tilde X$ where all terms are transformed according to Eq. (13), and the norm-increasing subdecomposition X' where a term $x \in X$ is changed according to Eq. (12) assuming $c_{x,0} + c_{x,1} > c_x$. Remembering that all coefficients are positive $c_y > 0$, the cost of each decomposition Eq. (11) is the square of a sum of positive values; the terms in this sum for $y \neq x$ do not change for $X \to X'$, and have a non-increasing value for $X \to \tilde X$. We thus focus only on the term $x \in X$ and the derived ones, highlighted here

$$\mathcal{M}_{X'} = \epsilon^{-2}\Bigg[\overbrace{\sum_{j\in\{0,1\}} c_{x,j}\sqrt{1-\langle \mathrm{Re}(U_{x,j})\rangle}}^{m'} + \ldots\Bigg]^2, \tag{C5}$$

$$\mathcal{M}_{\tilde X} = \epsilon^{-2}\Bigg[\underbrace{c_x(1-\bar\lambda_x)\sqrt{1-\langle \mathrm{Re}(U_{\tilde x})\rangle^2}}_{\tilde m} + \ldots\Bigg]^2. \tag{C6}$$

We now prove there exists a state $|\Psi\rangle$ for which $\tilde m < m'$, which implies $\mathcal{M}_{\tilde X} < \mathcal{M}_{X'}$.

Let $|\psi_+\rangle$ and $|\psi_-\rangle$ be eigenvectors of $\mathrm{Re}(U_{\tilde x})$ with eigenvalue $+1$ and -1 respectively. We consider three cases:

1. $|\langle\psi_\sigma|\mathrm{Re}(U_{x,j})|\psi_\sigma\rangle| < 1$ for at least one combination of $\sigma \in \{+,-\}$ and $j \in \{0,1\}$. In this case, on the state $|\Psi\rangle = |\psi_\sigma\rangle$ we get $\tilde m = 0 < m' \neq 0$.
2. $\langle\psi_\sigma|\mathrm{Re}(U_{x,j})|\psi_\sigma\rangle = \sigma$ for all combinations of $\sigma \in \{+,-\}$ and $j \in \{0,1\}$. By combining Eq. (12) and Eq. (13) and taking the expectation value on $|\psi_\sigma\rangle$ we obtain $\sigma[c_{x,0} + c_{x,1} - c_x(1-\bar\lambda_x)] = c_x\bar\lambda_x$, which implies $c_{x,0} + c_{x,1} = c_x$, violating one of the hypotheses of the lemma.
3. $\langle\psi_\sigma|\mathrm{Re}(U_{x,j})|\psi_\sigma\rangle = (-1)^j\sigma$ for all combinations of $\sigma \in \{+,-\}$ and $j \in \{0,1\}$. We define the state $|\Psi\rangle = \frac{|\psi_+\rangle+|\psi_-\rangle}{\sqrt{2}}$, on which $\langle \mathrm{Re}(\tilde U_x)\rangle = \langle \mathrm{Re}(\tilde U_{x,0})\rangle = \langle \mathrm{Re}(\tilde U_{x,1})\rangle = 0$. On this state, the costs are $\mathcal{M}_{\tilde X} = \epsilon^2 c_x^2(1-\bar\lambda_x)^2$ and $\mathcal{M}_{X'} = \epsilon^2(c_{x,0}+c_{x,1})^2$. As $\bar\lambda_x \geq 0$ and $c_{x,0} + c_{x,1} > c_x$, $\mathcal{M}_{X'} < \mathcal{M}_{\tilde X}$.

3. Proof of Lemma 3

In this appendix, we prove that the end-point of norm-preserving decomposition sequences are reflection operators. In other terms, if $\mathrm{Re}(U_x)$ is a reflection operator, it only admits a norm-preserving sub-decomposition [Eq. (12)] if $\mathrm{Re}(U_{x,0}) = \mathrm{Re}(U_{x,1}) = \mathrm{Re}(U_x)$.

To prove this, consider a state $|\psi\rangle$ in the $+1$ eigenspace of $\mathrm{Re}(U_x)$. For a norm-preserving decomposition, we must have

$$\begin{aligned} c_{x,0} + c_{x,1} = c_x &= c_x \langle\psi| \mathrm{Re}(U_x)|\psi\rangle \\ &= c_{x,0}\langle\psi| \mathrm{Re}(U_{x,0})|\psi\rangle + c_{x,1}\langle\psi| \mathrm{Re}(U_{x,1})|\psi\rangle. \end{aligned} \tag{C7}$$

As $\| \mathrm{Re}(U_{x,0})\|, \| \mathrm{Re}(U_{x,1})\| \leq 1$, this equality can only be satisfied if $|\psi\rangle$ is also a $+1$ eigenstate of both $U_{x,0}$ and $U_{x,1}$. A similar argument holds for all -1 eigenstates of U_x, and so $U_{x,0}, U_{x,1}$ and U_x share the same eigenstates and eigenvalues and must be equal. Taking such a sub-decomposition has no effect on the estimator of $\langle O\rangle$, as the same HT are performed and the total number of shots doesn't change, i.e. $\mathcal{M}_{X'} = \mathcal{M}$ in Eq. (C1).

4. Examples of reflection decompositions

The simplest example of a reflection-based decomposition is a decomposition in terms of Pauli operators

$$O = \sum_j^J c_j Z_j, \tag{C8}$$

with $c_j \geq 0$. We could be tempted to measure $\langle O\rangle$ with a single HT circuit (assuming access to a block-encoding of $\frac{O}{\|O\|}$, which is optimal). In this case, as $O = \|O\| \mathrm{Re}(U)$, the bound Eq. (11) is

$$M \geq \epsilon^{-2}\|O\|^2 \left[1 - \frac{\langle O\rangle^2}{\|O\|^2}\right]. \tag{C9}$$

To improve on this, we can estimate each $\langle Z_j\rangle$ separately, each with a Hadamard test with $U_j = \langle Z_j\rangle$ (a binary operator). As the spectral norm of O is equal to the induced 1-norm $\|O\|_1 = \sum_j^J c_j$, Eq. (C8) is a norm-preserving decomposition. The bound Eq. (11) then becomes

$$M \geq \epsilon^{-2} \left[\sum_j c_j \sqrt{1 - \langle Z_j\rangle^2}\right]^2, \tag{C10}$$

which is always smaller or equal than Eq. (C9) [easily proven through Eq. (C3)]. This inequality is only saturated when the considered state ρ has support only on the $\|O\|^2$-eigenvalue subspace of O^2; the operator O projected on this subspace is effectively a binary operator.

Norm-preserving decompositions do not need to involve only mutually commuting Pauli operators. As a practical example, we consider the two-qubit operator $O = XX + YY$, which appears commonly in quantum Hamiltonians. As $O = 2\,\mathrm{Im}[\mathrm{iSWAP}]$, this operator can be measured with a single Hadamard test circuit. Furthermore, in the context of electronic structure Hamiltonians, O preserves particle number, so in general a control-free scheme using the vacuum as reference state can be employed for the measurement. This operator has three eigenvalues $\{0, \pm 1\}$, which means we can improve its measurement by decomposing it in binary operators. We propose three decompositions $O = \mathrm{Re}\, U_0 + \mathrm{Re}\, U_1$ The obvious Pauli decomposition $U_0 = XX, U_1 = YY$ has the downside of not conserving particle number. To fix this, we can take

$$U_j = \frac{1}{2}[(XX + YY) + (-1)^j (Z\mathbb{1} + \mathbb{1}Z)]. \tag{C11}$$

These are particle-number preserving, reflection operators and can be easily implemented by combining iSWAP with single-qubit $e^{\pm iZ\pi/4}$ rotations on both qubits. The last decomposition,

$$U_j = \frac{1}{2}[(XX + YY) + (-1)^j (ZZ + \mathbb{1}\mathbb{1})], \tag{C12}$$

uses particle-preserving reflection operators with different eigenvalue multiplicities: unlike Pauli operators, the ± 1-eigenvalue subspaces of U_j have unequal dimension 1 and 3. For any state in the 0-eigenvalue subspace, spanned by $\{|00\rangle, |11\rangle\}$, the estimate variance $\mathrm{Var}^*[\langle \mathrm{Re}(U_j)\rangle] = 0$ for decomposition Eq. (C12). This is not true for the other two decompositions, which indicates that not all decompositions in binary operators are born equal. We will deal with this in the next section. Another example of a few-qubit reflection operator that is a sum of non-commuting Pauli operators is the three-spin all-to-all Heisenberg coupling

$$O = \frac{1}{3}\sum_{l=1}^{2}\sum_{m=0}^{l-1} X_m X_l + Y_m Y_l + Z_m Z_l, \tag{C13}$$

which appears e.g. in the Kagome-Heisenberg Hamiltonian.

5. Proof of Lemma 4

In this appendix we prove Lemma 4, which formally states the optimality and uniqueness of the Ξ-decomposition. To do this, we first define a variance bound for a class of estimators of $\langle O\rangle$ on a state $|\psi\rangle$. We prove that the bound is achieved on all eigenstates of O if all the sampled operators $\mathrm{Re}(U_x)$ are diagonal in the eigenbasis of O. We then construct the Ξ-decomposition, and prove that the related estimator saturates the bound on the set S_2 of all states with support on at most two eigenstates of O. Finally, we prove no other decomposition satisfies this requirement (i.e. the Ξ-decomposition is unique), and no decomposition satisfies the bound on a superset $S \supset S_2$.

A decomposition X [Eq. (6)] of an operator O is optimal on a state $|\psi\rangle$ if no other decomposition produces an estimator with lower cost [Eq. (11)] for that state. Optimality can be defined for a set S of states: X is optimal on S if, for each $|\psi\rangle \in S$, no decomposition X' has lower cost $M_{X'} < M_X$. (Note that this can be readily generalized to mixed state, without changing any of our next results.) Lemmas 1-3 imply a necessary condition for optimality on the whole Hilbert space: X can only be optimal on all states if it has the form

$$O = \bar{\lambda}_O \mathbb{1} + \sum_{x\in X} c_x \operatorname{Re}(U_x), \tag{C14}$$
$$c_x > 0, \quad |\bar{\lambda}_O| + \sum_{x\in X} c_x = \|O\|, \quad \operatorname{Re}(U_x)^2 = \mathbb{1}$$

where $\bar{\lambda}_O$ is the average of the largest and smalles eigenvalues of O. In other words, X is a norm-preserving decomposition of the center of O where all sampled terms are reflection operators. This condition is not sufficient: as many non-equivalent instances of such decompositions exist, as exemplified in Appendix C 4.

We now construct a bound on the variance of the estimator of $\langle O\rangle$ based on the decomposition X: saturating this bound on all $|\psi\rangle \in S$ implies optimality of X on S. [The cost of the decomposition Eq. (11) is defined as the minimum value of M required to achieve target variance ϵ^2, so minimum variance at fixed M implies minimum cost at fixed ϵ.]

$$\mathrm{Var}^*_X[\langle O\rangle] = \frac{1}{M}\left[\sum_x c_x\sqrt{1-\langle \operatorname{Re}(U_x)\rangle^2}\right]^2 \geq \frac{\mathrm{Var}[O]}{M}. \tag{C15}$$

This bound is implied by Eq. (5) and Eq. (8), with the choice of optimal shot allocation Eq. (10). It physical interpretation is rooted in the following observation: a Von Neumann measurement of O is the lowest-variance unbiased estimator of $\langle O\rangle$ when given access to a single state preparation. Thus, given M independent experiments each with a single state preparation, the mean of Von Neumann measurements is the lowest-variance unbiased estimator.

We first consider the set S_1 of all eigenstates of O. For any $|\phi\rangle \in S_1$, the value of the bound in Eq. (C15) becomes $\mathrm{Var}[O] = 0$. The bound is thus saturated only if we choose all reflection operators $\operatorname{Re}(U_x)$ diagonal in any eigenvector basis of O, i.e. $[U_x, O] = 0$ and $U_x|\phi\rangle = \pm|\phi\rangle$ for any $|\phi\rangle \in S_1$. For any decomposition of this form, we can write all U_x in terms of the eigenspace projectors of O:

$$U_x = \sum_{j=0}^{J-1} \xi_{x,j}\Pi_j\,, \quad \xi_{x,j} \in \{\pm 1\}, \tag{C16}$$

where Π_j is the projector on the (eventually degenerate) λ_j-eigenspace of O, J is the number of distinct eigenvalues $\{\lambda_j\}$ of O, and without loss of generality we assume $\lambda_j > \lambda_{j-1}$. The coefficients will then have to satisfy the relation $\lambda_j = \sum_x c_x \xi_{x,j}$.

We define the Ξ-decomposition based on Eq. (C16), by choosing $\xi_{x,j} = -1$ if $j < x$, and $+1$ otherwise. The resulting decomposition is presented in Lemma 4, Eq. (14). The operators Ξ_x are reflections by definition, and it is easy to check that the decomposition satisfied the necessary condition Eq. (C14). Note that $c_0 = (\lambda_0 + \lambda_J)/2$ defines the optimal identity shift (producing the center of O) and the $c_x = (\lambda_x - \lambda_{x-1})/2$ complete the decomposition.

We now prove that the Ξ-decomposition is optimal on the set S_2 of states with support on two eigenstates of O,

$$S_2 = \left\{ \frac{\alpha|\lambda_m\rangle + \beta|\lambda_n\rangle}{\sqrt{\alpha^2+\beta^2}} : |\lambda_m\rangle, |\lambda_n\rangle \in S_1 \right\}. \tag{C17}$$

On a general state $|\psi\rangle$ with eigenspace occupations $a_j = \langle\psi|\Pi_j|\psi\rangle$, the estimator based on the Ξ-decomposition has variance

$$\mathrm{Var}^*_\Xi[\langle O\rangle] = \frac{1}{M}\left[\sum_j^{J-1} \frac{\delta\lambda_j}{2}\sqrt{4\Big(\sum_{i<j} a_i\Big)\Big(\sum_{i\geq j} a_i\Big)}\right]^2. \tag{C18}$$

For a state $|\phi\rangle \in S_2$, only two occupations are nonzero $a_m, a_n \neq 0$ (we assume w.l.g. $m < n$), thus the term under square root is reduced to $4a_m a_n$ if $m < j \leq n$ and 0 otherwise. The resulting variance

$$\begin{aligned}\mathrm{Var}^*_\Xi[\langle\phi|O|\phi\rangle] &= \frac{1}{M}\left[\frac{\lambda_n - \lambda_m}{2}\sqrt{4a_m a_n}\right]^2 \\ &= \frac{1}{M} a_n a_m (\lambda_n - \lambda_m) = \mathrm{Var}[\langle O\rangle]\end{aligned} \tag{C19}$$

thus saturating the bound Eq. (8).

We now prove that the only optimal decomposition on S_2 is the Ξ-decomposition (or equivalent up to relabeling and trivial subdecompositions). First of all, $S_1 \subset S_2$, so the terms of the decomposition need to be of the form of Eq. (C16). Consider a family of states $\sqrt{a_m}|\lambda_m\rangle + \sqrt{a_n}|\lambda_n\rangle$ for any $n > m$, with only two nonzero eigenstate occupations $a_m + a_n = 1$. On such a state,

$$\begin{aligned}\mathrm{Var}^*_X[O] &= \frac{1}{M}\left[\sum_x c_x\sqrt{1-[a_m\xi_{x,m} + a_n\xi_{x,n}]^2}\right]^2 \\ &= \frac{a_m a_n}{M}\left[\sum_x 2c_x\frac{1-\xi_{x,m}\xi_{x,n}}{2}\right]^2.\end{aligned} \tag{C20}$$

The bound Eq. (C15) is then saturated when

$$\left[\sum_x 2c_x\frac{1-\xi_{x,m}\xi_{x,n}}{2}\right]^2 = \lambda_n - \lambda_m, \tag{C21}$$

where we simplified out the free parameter $\frac{a_m a_n}{M}$. This can be rewritten as

$$\sum_x c_x \xi_{x,n}(\xi_{x,n} - \xi_{x,m}) = \sum_x c_x(\xi_{x,n} - \xi_{x,m}) \tag{C22}$$

using the condition on the decomposition coefficients $\lambda_j = \sum_x c_x \xi_{x,j}$. This implies that, if $\xi_{x,n} = -1$ then $(\xi_{x,m} - \xi_{x,n}) = 0$ (recall that $c_x > 0$), i.e. $\xi_{x,m} = -1$. Thus the only U_x that can appear in this decomposition, are of the same form as the operators in the Ξ decomposition ($\xi_{j,m} = -1, \xi_{j,n} = +1$ for $m < j \leq n$), and thus X is either Ξ or a trivial sub-decomposition of it.

We now show that the Ξ-decomposition does not saturate the bound Eq. (C15) for a state $|\psi\rangle$ with three non-zero occupations, $a_m, a_n, a_p \neq 0$ ($m < n < p$). On this state we can write

$$\mathrm{Var}^*_{\Xi}[\langle O\rangle] = \frac{1}{M}\Bigg[(\lambda_n - \lambda_m)\sqrt{a_m(a_n + a_p)} \tag{C23}$$

$$+ (\lambda_p - \lambda_n)\sqrt{a_p(a_m + a_n)}\Bigg]^2 . \tag{C24}$$

Subtracting from this $\mathrm{Var}[\langle O\rangle]$, expanding and then collecting terms we get

$$\begin{aligned} &\mathrm{Var}^*_{\Xi}[\langle O\rangle] - \mathrm{Var}[\langle O\rangle] = \\ &\quad = [(\lambda_n - \lambda_p)(\lambda_n - \lambda_m)] \cdot \\ &\quad \cdot \left[a_m a_p - \sqrt{a_m a_p (a_m + a_n)(a_p + a_n)}\right] > 0, \end{aligned} \tag{C25}$$

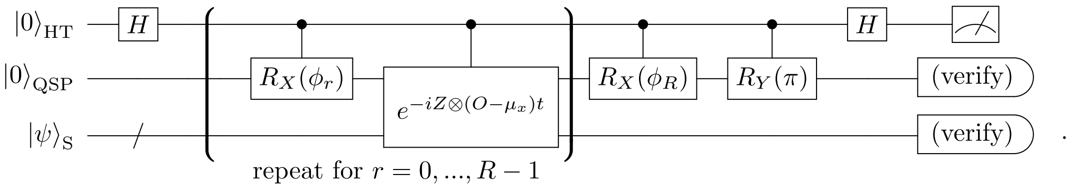
as both the terms in square brackets are strictly smaller than zero. This (along with the uniqueness of Ξ as the optimal estimator on S_2) implies that no HT-based estimator can saturate the bound Eq. (C15) for arbitrary states.

In fact, the bound can only be saturated on states in S_2: on these states the Von Neumann measurement has only two possible outcomes (λ_m and λ_n) with nonzero probability. The adaptive shot allocation scheme then ensures (for a large enough M) that most of the measurements we take (Ξ_x with $m \leq x < n$) reproduce the statistics of the Von Neumann measurement, with the single bit we sample in every experiment always distinguishing between λ_m and λ_n. On any state $|\psi\rangle \in \overline{S_2}$, the Von Neumann measurement has three or more outcomes with non-zero probability, and we cannot repoduce its statistics by sampling a single qubit per experiment. This, along with the uniqueness of Ξ, implies that no decomposition can satisfy the sufficient condition for optimality on a superset $S \supset S_2$. The numerical results presented in this paper quantify the increase in variance with respect to the bound, along with confirming the Ξ-decomposition outperforms other decompositions on all states.

D. Implementation of the Ξ decomposition via quantum signal processing

Verifiable samping of QSP polynomials — To measure the operators in the Ξ decomposition Eq. (16), we implement a Hadamard test (or EV) on trigonometric polynomials of $(H - \mu_x)t$ generated by the quantum signal processing. We tune the QSP coefficients such that the polynomials approximate the sign function in a suitable range. In this section we display and analyse this technique.

The full circuit we use to achieve this is:



The first control qubit (labeled HT) takes care of the Hadamard test. The second ancilla (labeled QSP) manages the quantum signal processing subroutine, extended through the sign-controlled evolution $e^{-iZ\otimes(O-\mu_x)t}$ to implements a quantum signal processing (QSP) on the operator $e^{(O-\mu_x)t}$. We now describe how the measurement scheme works, and how to select the ϕ parameters to approximate a measurement of $\mathrm{sgn}[(O-\mu_x)t]$ in the interval $[-\pi, \pi]$.

First, we analyze the QSP routine. Let us assume $|\psi\rangle$ to be an eigenstate of $(O - \mu_x)t$ with eigenvalue $\omega \in (-\pi, \pi)$, and only consider the effect of the controlled gates (removing the HT qubit). Then, we can reduce the cicuit to an effective single-qubit gate on the QSP qubit, with action

$$\begin{aligned} Q_\phi(\omega) &= e^{-i\frac{Y}{2}\pi} e^{-i\frac{X}{2}\phi_R}\left[\prod_{r=1}^{R} e^{-i\frac{Z}{2}2\omega} e^{-i\frac{X}{2}\phi_{R-r}}\right] \\ &= \begin{pmatrix} S_\phi(\omega) & \cdot \\ \cdot & \cdot \end{pmatrix} \end{aligned} \tag{D1}$$

which is a block encoding of $S_\phi(\omega)$, a degree-R trigonometric polynomial of ω. For the sake of simplicity we inserted the final gate $e^{-i\frac{Y}{2}\frac{\pi}{2}} = -iY$, shifting the polynomial of interest S from the block $\langle 1|Q|0\rangle$ to $\langle 0|Q|0\rangle$.

We ensure $S_\phi(\omega)$ is real and odd by constraining

$$\phi_r = -\phi_{R-r} \implies S(\omega) = -S(-\omega) \in \mathbb{R}. \tag{D2}$$

Re-introducing the system register, i.e. taking a general $|\psi\rangle_{\mathrm{S}}$, can be done by linearity taking $Q(\omega) \mapsto Q[(O-\mu_x)t]$ and recovering the circuit above.

The result of the verified Hadamard test (or EV) is obtained by measuring on the output state of the circuit the expectation value of Z_{HT} (or $Z_{\mathrm{HT}} \otimes |0\rangle\langle 0|_{\mathrm{QSP}} \otimes |\psi\rangle\langle\psi|_{\mathrm{S}}$). (In the absence of noise these two expectation values are equal. In the presence of noise, an additional measurement at $t=0$ can be taken to mitigate errors. For more details on the technique we refer the reader to the original work on EV [11].)

Approximating the sign function — To approximate the operators Eq. (16) that make up the Ξ decomposition, we need to choose the QSP parameter $\boldsymbol{\phi}$ such that $S_{\boldsymbol{\phi}}(\omega)$ in Eq. (D1) approximates $\mathrm{sgn}[\omega]$. The polynomial $S_{\boldsymbol{\phi}}(\omega)$ is odd, real, and 2π-periodic – thus having nodes $S_{\boldsymbol{\phi}}(0) = S_{\boldsymbol{\phi}(\pm\pi)} = 0$. To account for the approximation error in the neighborhood of these nodes, we introduce a resolution parameter $\delta \geq 0$, and request the approximation to be effective only in the $[\delta, \pi-\delta]$ interval. Choosing $\delta > 0$ implies accepting a larger error in approximating the sign function close to zero. For example, we know the eigenvalues of $(O-\mu_x)t$ closest to zero have absolute value $\frac{\delta\lambda_x}{2}t$, we can use this knowledge to choose δ.

We define a loss function to characterize the quality of the approximation: the average error

$$\mathcal{L}_\delta(\boldsymbol{\phi}) = \frac{1}{\pi-2\delta}\int_\delta^{\pi+\delta} d\omega \Big[\mathrm{sgn}(\omega) - \mathrm{Im}[S_{\boldsymbol{\phi}}(\omega)]\Big]. \tag{D3}$$

To choose the optimal parameters $\boldsymbol{\phi}$, we minimize this loss under the constraints (D2). Although an analytical approach to this problem is possible building on the techniques described in [16], we take the numerical route to this approximation (which is efficient, scalable and easy to implement). The integral is thus substituted with a sum on a grid with a number of points much larger than the degree of the trigonometric polynomial. We plot in Fig 2 the minimized cost function, as a function of the approximation's order R and of the resolution parameter δ. We find that the loss always decays exponentially with an increasing order R, with a decay rate depending on δ.

E. The generalized parameter-shift kernel decomposition of a diagonal operator with ladder spectrum

In [17] the authors propose techniques to estimate derivatives $\langle \frac{d}{dt}U(t)\rangle$ of a unitary $U(t) = e^{iOt}$ generated by O, by sampling $\langle \sin(Ot_l)\rangle = \langle \mathrm{Re}[U(t_l)]\rangle$ at a discrete set of points $\{t_l\}$. This technique can be used to estimate expectation values of O, as $\langle O\rangle = \langle [-i\frac{d}{dt}e^{iOt}]_{t=0}\rangle$, and it is clearly compatible with Hadamard test or EV measurements (as it only requires sampling $\langle \mathrm{Re}\, U(t_l)\rangle$).

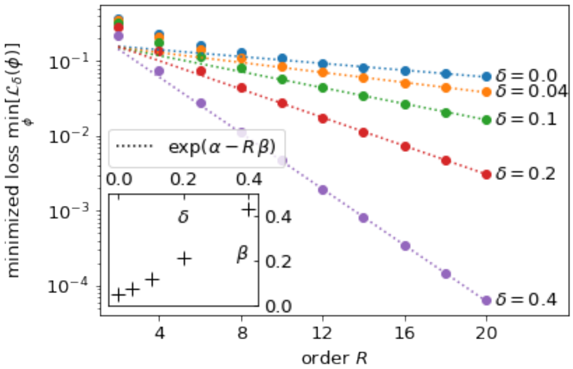


FIG. 2: Loss Eq (D3) for the optimal choice of QSP parameters $\boldsymbol{\phi}$, as a function of the order R (number of QSP layers) and resolution parameter δ. The dotted lines are log-lin fits for $R > 10$. The dependence of the fit parameter β on the resolution δ is shown in the inset.

For an operator O with equispaced eigenvalues $\Omega, 2\,\Omega, ..., R\,\Omega$ (commonly referred to as a "ladder spectrum"), the authors give a choice of $\{t_l\}$ and explicit coefficients $c_l(t)$ for the linear combination $\langle \frac{d}{dt}U(t)\rangle = \sum_l c_l \langle \mathrm{Re}[-iU(t_l)]\rangle$. Assuming $\Omega = 1$ (which can be considered a choice of units for the energy), the time points are chosen as $\{t_l = \frac{2l}{2R+1}\pi\}$. We can then define a modified version of the Dirichelet kernel,

$$\tilde{D}_l(t) = \frac{1}{R}\cos(t_l)\left[\frac{1}{2}\sin(Rt) + \sum_{j=1}^{R-1}\sin(jt)\right], \tag{E1}$$

which satisfies $\tilde{D}_l(t_{l'}) = \delta_{ll'}$. This is a linear combination of the R basis functions $\{\sin(jt)\}_{j=1,\ldots,R}$, like $\langle \sin(Ot)\rangle$. Thus, as the equality

$$\langle \sin(Ot)\rangle = \sum_{l=1}^{R} \langle \sin(Ot_l)\rangle \tilde{D}_l(t) \tag{E2}$$

holds for all $\{t_l\}_{l=1,\ldots,R}$, it must to hold for all t. We can then differentiate the kernel rather than the expectation value itself. Evaluating $[\frac{\partial}{\partial t}\tilde{D}_l(t)]_{t=0}$ and combining the equations above we obtain

$$\langle O\rangle = \sum_{l=1}^{R} \frac{(-1)^{l-1}}{2R\sin^2(\frac{1}{2}t_l)}\langle \sin(Ot_l)\rangle \tag{E3}$$

$$= \sum_{l=1}^{R} c_l \langle \mathrm{Re}[-iU(t_l)]\rangle. \tag{E4}$$

This matches the form of decompositions Eq. (6). We call thi the generalized parameter shift kernel (GPSK) decomposition. Under the optimal shot allocation choice [Eq. (10)], the shot-variance of the estimator based on this decomposition is

$$M\,\mathrm{Var}^*_{\mathrm{GPSK}} = \left[\sum_{l=1}^{R} \frac{\sqrt{1-\langle \sin(Ot_l)\rangle^2}}{|2R\sin^2(\frac{1}{2}t_l)|}\right]^2 \tag{E5}$$

F. Details on numerical simulations and further numerical results

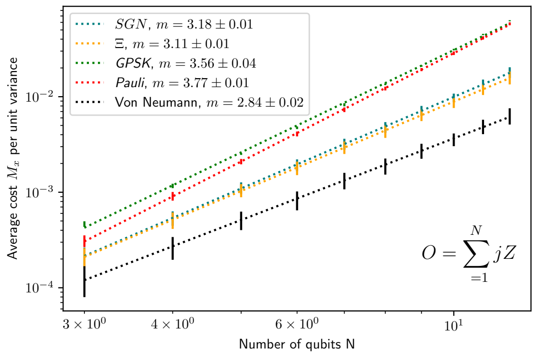


FIG. 3: Comparison study of variances of different decompositions on random states generated by a hardware-efficient ansatz (see text for details). Different colours correspond to different decompositions [Eq.(6)] of the target operator O (see text for the description of all decompositions). Dashed lines are exponential fits ($a\exp(mN+b)$) to the data (the parameter m is given in legend).

We measure the variances on random states generated by hardware-efficient ansatzes using PennyLane [28]. For each value of N, 100 random set of parameters (and therefore 100 random states) are generated and measured for all decompositions. For each decomposition X, we first use 10^5 shots (allocated proportionally to the weight of each term) to obtain a rough estimate of the expectation value of each term $\langle \mathrm{Re}(U_x)\rangle$ for $x \in X$. These values are plugged in Eq. (10) to get an estimate of the optimal shot allocation ratios $r_x = \frac{m_x}{M_x}$. The variance of each term $\mathrm{Var}^*[\langle \mathrm{Re}(U_x)\rangle]$ is obtained by Eq. (4) (or by sampling in the case of the QSP-approximation decomposition 'SGN'). With these we compute the final shot-variance $M_X\,\mathrm{Var}^*_X[\langle O\rangle] = \sum_{x\in X} r_x^{-1}\,\mathrm{Var}^*[\langle \mathrm{Re}(U_x)\rangle$. Finally, we average the values of $M_X\,\mathrm{Var}^*_X[\langle O\rangle]$ obtained for each random state. This average is the quantity reported in Fig. 1, Fig. 3 and Fig. 4.

The terms Ξ_x are constructed as per Eq. (14) using the known eigenvectors of O, and projectively measured on the prepared state (as these are reflection operators, Hadamard test samples match projective measurement samples). The terms in the Pauli decomposition are also directly measured on the prepared state. The GPSK-decomposition is constructed as described in E and measured through a Hadamard test. The Von Neumann variance $\mathrm{Var}[O]$ is computed analytically.

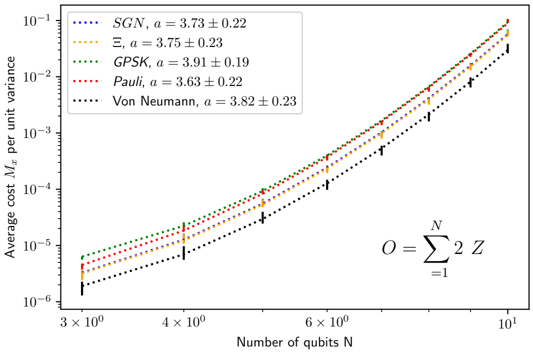


FIG. 4: Comparison study of variances of different decompositions on random states generated by a hardware-efficient ansatz (see text for details). Different colours correspond to different decompositions [Eq.(6)] of the target operator O (see text for the description of all decompositions). Dashed lines are exponential-power-law fits ($\exp(aN^2 + bN + c)$) to the data (the dominant scaling parameter a is given in the legend).

The QSP approximation of Ξ (denoted SGN from the sign term approximation) is implemented as described in Appendix D for $R = 20$ and $\delta = 0$. For fair comparison with the other methods, echo verification is not used. The comparison between the Ξ and SGN decomposition shows how the approximation increased the final variance. (The approximation also introduces a bias, see Appendix D.

All the simulations assume Hadamard-test-based measurement in an ideal circuit simulation: no circuit-level noise is considered and EV is not implemented.

We additionally report scaling results for the shot-variances of two other observables, $O = \sum_j jZ_j$ and $O = \sum_j 2^j Z_J$. The overall scaling of all decompositions matches the scaling of the operator norm $\|O\|$. Similarly to the case of Fig. 1, the Ξ decomposition performs best, the SGN approximation has a relatively small effect on the shot-variance, and the Pauli decomposition shows the worst scaling.

7 Discussion

This thesis has advanced the understanding of various aspects of the variational quantum eigensolver (VQE) for chemistry simulations, with a focus on superconducting NISQ hardware with error mitigation. Optimal strategies to estimate expectation values in the resulting single qubit measurement scenarios were found and implemented. Two more hardware accessible chemistry ansatzes were proposed, a more dense formulation of an ansatz approximating chemistry in a paired approximation and a new gate fabric that only parametrizes the relevant subspace with the correct quantum numbers. Purification based error mitigation frameworks have been tested in simulation and on hardware testing if accurate energy estimates can be extracted from noisy quantum hardware.

After almost a decade of VQE, this gives us a moment to reflect on the current state of the algorithm. Purification error mitigation methods have shown in theory and in experiment to give accurate results and to be able mitigate most of the incoherent error present in the quantum circuit, albeit still at an exponential scaling in circuit fidelity of the given quantum circuit. The purification methods make use of the exponentially growing Hilbert space, meaning as the scale of the experiments grows bigger the set of errors that pass verification decreases, as was shown in the scaling of the mean absolute error with qubit number in Section 4. However, very high shot budgets are needed to achieve these accurate results. In the optimal purification, when one is able to postselect on states where no error has occured, the number of shots that need to be taken and therefore the computational time grows exponentially with the infidelity of the circuit, hinting at an unsustainable path of scaling these experiments up to relevant sizes. New paradigms in quantum superconducting hardware [92] bringing down error rates, almost rivaling error rates in Ion-Traps, could achieve fidelities and sampling rates more compatible with these techniques. Parallel execution on system with many qubits or separate chips in the same fridge could also help increase shot counts.

Faced with the task of estimating expectation values in these error mitigated settings most optimized measurement settings like grouping of the $\mathcal{O}(N^4)$ Pauli-terms or double-factorization rely on the extraction of multiple bits per shot and are at first glance not compatible with these purification techniques. The optimal strategy of estimating Pauli-words one at a time recovers the original unfavorable scaling and the optimal scaling derived in parts in this work using quantum signal processing leads to deep circuits not in line with the NISQ paradigm. However, providing a framework with the reflection-decomposition nature of optimal decomposition setting, shorter decompositions with slightly worse variance are not ruled out and shadow purification techniques that do not rely on the Hadamard-test are not affected. Circuit depth is a major factor in achieving reasonable fidelities in non-error corrected scenarios, so the compatibility of the ansatz in use with the native gate set plays a key role in making this possible. Just using native gate sets like the hardware efficient ansatz comes with convergence problems. The first half of the gate set used in the quantum number preserving ansatz, orbital rotations, is already a native gate for the superconducting hardware and as the other gate (Pair exchange) is also physically motivated, it is well conceivable that in some platforms native implementations of this four qubit gate should be possible.

With all these limitations and possible solutions in mind, possibly the biggest challenge to overcome is the one most large VQE implementations omit - optimization. The global optimization of circuit parameters is one step all variational algorithms, even ADAPT, have in common. From the point of view of the experience made through this work it seems unlikely that on-device optimization at large scale is feasible with the current available methods and optimizers to the accuracy needed to rival classical computational chemistry methods and even in simulation with analytical gradients convergence seems difficult. To overcome this, finding better optimization techniques or being able to find optimal or at least good enough initial parameters classically (pre-initialisation) is paramount for VQE to find use. Other paths forward are techniques that do not rely on optimization (and extracting chemically accurate energies from the quantum hardware, but this to a lesser extent) [42]. Mid-circuit measurements (measurement on parts of the quantum circuit and being able to conditionally apply gates determined by the outcome) have also received increased attention over the last year as most no-go theorems for classical simulatability and less than exponential sample overhead error mitigation do not apply when allowing for mid-circuit measurements [93]. Creative ways of using these measurements [94], combining measurement based and gate based quantum computation, could also lead to more optimal preparation schemes of chemically interesting states. Otherwise, it pushes chemically relevant use-cases for quantum computers into the error-corrected regime.

8 Availability of primary data

The unavoidable nature of an Industrial-PhD is that some parts of the work remain confidential, however an effort was made to provide as much data and open-source implementation of ideas of the work presented in this thesis. Data from the Purification-experiment in with Google Quantum AI in Section 4 is available at `https://doi.org/10.5281/zenodo.7225821`. The QNP-ansatz from Section. 5 has been integrated into the Pennylane-suite at `https://docs.pennylane.ai/en/stable/code/api/pennylane.GateFabric.html`, although not by the author due to IP filed with the paper. Shown data from the large scale VQE optimization has been made available on Github at `https://github.com/gianans/PhDErrorbenchmark`. The implementation of the optimal decomposition has been integrated by the author into the Pennylane-suite at `https://github.com/PennyLaneAI/pennylane/pull/2852` to allow for reproduction of plots and further development by other researchers. It misses the feature of allocating shots ideally to the independent measurements as Pennylane did not support shot distribution at the moment of implementation of the feature, so it does not recover the true scaling of the decomposition after allocating shots optimally. The code generating the data and plots for the benchmark in Section 3.6 and 3.7 has been made available on Github at `https://github.com/gianans/PhDErrorbenchmark`.

List of publications

Anselmetti, G. L. R., Martinez, E. A., Ménard, G. C., Puglia, D., Malinowski, F. K., Lee, J. S., … Marcus, C. M. (2019). End-to-end correlated subgap states in hybrid nanowires. *Physical Review B*, *100*(20), 205412. (Publisher: APS)

Ménard, G., Anselmetti, G., Martinez, E., Puglia, D., Malinowski, F., Lee, J., … Higginbotham, A. (2020, January). Conductance-Matrix Symmetries of a Three-Terminal Hybrid Device. *Physical Review Letters*, *124*(3), 036802. Retrieved 2023-05-31, from `https://link.aps.org/doi/10.1103/PhysRevLett.124.036802` (Publisher: American Physical Society) doi: 10.1103/PhysRevLett.124.036802

Anselmetti, G.-L. R., Wierichs, D., Gogolin, C., & Parrish, R. M. (2021, November). Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems. *New Journal of Physics*, *23*(11), 113010. Retrieved 2023-04-06, from `https://dx.doi.org/10.1088/1367-2630/ac2cb3` (Publisher: IOP Publishing) doi: 10.1088/1367-2630/ac2cb3

O'Brien, T. E., Anselmetti, G., Gkritsis, F., Elfving, V. E., Polla, S., Huggins, W. J., … Rubin, N. C. (2022, October). *Purification-based quantum error mitigation of pair-correlated electron simulations.* arXiv. Retrieved 2023-05-31, from `http://arxiv.org/abs/2210.10799` (arXiv:2210.10799 [quant-ph]) doi: 10.48550/arXiv.2210.10799

Polla, S., Anselmetti, G.-L. R., & O'Brien, T. E. (2022, July). *Optimizing the information extracted by a single qubit measurement.* arXiv. Retrieved 2023-05-31, from `http://arxiv.org/abs/2207.09479` (arXiv:2207.09479 [quant-ph]) doi: 10.48550/arXiv.2207.09479

Parrish, R. M., Anselmetti, G.-L. R., & Gogolin, C. (2021, October). *Analytical Ground- and Excited-State Gradients for Molecular Electronic Structure Theory from Hybrid Quantum/Classical Methods.* arXiv. Retrieved 2023-05-31, from `http://arxiv.org/abs/2110.05040` (arXiv:2110.05040 [quant-ph]) doi: 10.48550/arXiv.2110.05040

Hohenstein, E. G., Oumarou, O., Al-Saadon, R., Anselmetti, G. L. R., Scheurer, M., Gogolin, C., & Parrish, R. M. (2023, March). Efficient quantum analytic nuclear gradients with double factorization. *The Journal of Chemical Physics*, *158*(11). Retrieved from `https://doi.org/10.1063/5.0137167` doi: 10.1063/5.0137167

References

[1] J. C. Slater. Atomic Shielding Constants. *Physical Review*, 36(1):57–64, July 1930. Publisher: American Physical Society.

[2] J. C. Slater. A Simplification of the Hartree-Fock Method. *Physical Review*, 81(3):385–390, February 1951. Publisher: American Physical Society.

[3] Theresa Sperger, Italo A. Sanhueza, and Franziska Schoenebeck. Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities. *Accounts of Chemical Research*, 49(6):1311–1319, June 2016. Publisher: American Chemical Society.

[4] W. Kohn, A. D. Becke, and R. G. Parr. Density Functional Theory of Electronic Structure. *The Journal of Physical Chemistry*, 100(31):12974–12980, January 1996. Publisher: American Chemical Society.

[5] Attila Szabo and Neil S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Dover Publications, Mineola, N.Y, 1996.

[6] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular electronic-structure theory*. Wiley, Chichester ; New York, 2000.

[7] M. G. Evans and M. Polanyi. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.*, 31(0):875–894, 1935. Publisher: The Royal Society of Chemistry.

[8] Henry Eyring. The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics*, 3(2):107–115, 1935. _eprint: https://pubs.aip.org/aip/jcp/article-pdf/3/2/107/11062599/107_1_online.pdf.

[9] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of Superconductivity. *Physical Review*, 108(5):1175–1204, December 1957. Publisher: American Physical Society.

[10] Vincent E. Elfving, Marta Millaruelo, José A. Gámez, and Christian Gogolin. Simulating quantum chemistry in the seniority-zero space on qubit-based quantum computers. *Physical Review A*, 103(3):032605, March 2021. Publisher: American Physical Society.

[11] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas P. D. Sawaya, Sukin Sim, Libor Veis, and Alán Aspuru-Guzik. Quantum Chemistry in the Age of Quantum Computing. *Chemical Reviews*, 119(19):10856–10915, October 2019. Publisher: American Chemical Society.

[12] Guang Hao Low. Classical shadows of fermions with particle number symmetry, August 2022. arXiv:2208.08964 [quant-ph].

[13] Erik Koch. 2 Mean-Field Theory : Hartree-Fock and BCS. 2016.

[14] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, 92(1):015003, March 2020. Publisher: American Physical Society.

[15] Jiří Čížek. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *jcp*, 45(11):4256–4266, December 1966.

[16] Chi Chen. Fitting the parameters in STO-LG. `https://chc273.github.io/fitting-parameters-in-sto-lg/`, January 2022.

[17] Robert Ditchfield. Self-consistent perturbation theory of diamagnetism. *Molecular Physics*, 27(4):789–807, 1974. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00268977400100711.

[18] Jr. Dunning, Thom H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, January 1989. _eprint: https://pubs.aip.org/aip/jcp/article-pdf/90/2/1007/15358102/1007_1_online.pdf.

[19] David Deutsch and Michael Josza. Rapid solution of problems by quantum computation | Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences, December 1992.

[20] Peter W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997. _eprint: https://doi.org/10.1137/S0097539795293172.

[21] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing*, STOC '96, pages 212–219, New York, NY, USA, July 1996. Association for Computing Machinery.

[22] Daniel S. Abrams and Seth Lloyd. Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors. *Physical Review Letters*, 83(24):5162–5165, December 1999. Publisher: American Physical Society.

[23] Daniel A. Lidar and Haobin Wang. Calculating the thermal rate constant with exponential speedup on a quantum computer. *Phys. Rev. E*, 59(2):2429–2438, February 1999. Publisher: American Physical Society.

[24] G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme. Quantum algorithms for fermionic simulations. *Physical Review A*, 64(2):022319, July 2001. Publisher: American Physical Society.

[25] Hefeng Wang, Sabre Kais, Alán Aspuru-Guzik, and Mark R. Hoffmann. Quantum algorithm for obtaining the energy spectrum of molecular systems. *Physical Chemistry Chemical Physics*, 10(35):5388–5393, August 2008. Publisher: The Royal Society of Chemistry.

[26] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. Simulated Quantum Computation of Molecular Energies. *Science*, 309(5741):1704–1707, 2005. _eprint: https://www.science.org/doi/pdf/10.1126/science.1113479.

[27] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer. Gate-count estimates for performing quantum chemistry on small quantum computers. *Physical Review A*, 90(2):022305, August 2014. Publisher: American Physical Society.

[28] Ryan Babbush, Jarrod McClean, Dave Wecker, Alán Aspuru-Guzik, and Nathan Wiebe. Chemical basis of Trotter-Suzuki errors in quantum chemistry simulation. *Phys. Rev. A*, 91(2):022311, February 2015. Publisher: American Physical Society.

[29] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. Quantum computing enhanced computational catalysis. *Physical Review Research*, 3(3):033055, July 2021. Publisher: American Physical Society.

[30] Guang Hao Low and Isaac L. Chuang. Hamiltonian Simulation by Qubitization. *Quantum*, 3:163, July 2019. Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.

[31] S. A. Moses, C. H. Baldwin, M. S. Allman, R. Ancona, L. Ascarrunz, C. Barnes, J. Bartolotta, B. Bjork, P. Blanchard, M. Bohn, J. G. Bohnet, N. C. Brown, N. Q. Burdick, W. C. Burton, S. L. Campbell, J. P. Campora III, C. Carron, J. Chambers, J. W. Chan, Y. H. Chen, A. Chernoguzov, E. Chertkov, J. Colina, M. DeCross, J. M. Dreiling, C. T. Ertsgaard, J. Esposito, B. Estey, M. Fabrikant, C. Figgatt, C. Foltz, M. Foss-Feig, D. Francois, J. P. Gaebler, T. M. Gatterman, C. N. Gilbreth, J. Giles, E. Glynn, A. Hall, A. M. Hankin, A. Hansen, D. Hayes, B. Higashi, I. M. Hoffman, B. Horning, J. J. Hout, R. Jacobs, J. Johansen, T. Klein, P. Lauria, P. Lee, D. Liefer, S. T. Lu, D. Lucchetti, A. Malm, M. Matheny, B. Mathewson, K. Mayer, D. B. Miller, M. Mills, B. Neyenhuis, L. Nugent, S. Olson, J. Parks, G. N. Price, Z. Price, M. Pugh, A. Ransford, A. P. Reed, C. Roman, M. Rowe, C. Ryan-Anderson, S. Sanders, J. Sedlacek,

P. Shevchuk, P. Siegfried, T. Skripka, B. Spaun, R. T. Sprenkle, R. P. Stutz, M. Swallows, R. I. Tobey, A. Tran, T. Tran, E. Vogt, C. Volin, J. Walker, A. M. Zolot, and J. M. Pino. A Race Track Trapped-Ion Quantum Processor, May 2023. arXiv:2305.03828 [quant-ph].

[32] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, October 2019. Number: 7779 Publisher: Nature Publishing Group.

[33] Rajeev Acharya, Igor Aleiner, Richard Allen, Trond I. Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Juan Atalaya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Joao Basso, Andreas Bengtsson, Sergio Boixo, Gina Bortoli, Alexandre Bourassa, Jenna Bovaird, Leon Brill, Michael Broughton, Bob B. Buckley, David A. Buell, Tim Burger, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Ben Chiaro, Josh Cogan, Roberto Collins, Paul Conner, William Courtney, Alexander L. Crook, Ben Curtin, Dripto M. Debroy, Alexander Del Toro Barba, Sean Demura, Andrew Dunsworth, Daniel Eppens, Catherine Erickson, Lara Faoro, Edward Farhi, Reza Fatemi, Leslie Flores Burgos, Ebrahim Forati, Austin G. Fowler, Brooks Foxen, William Giang, Craig Gidney, Dar Gilboa, Marissa Giustina, Alejandro Grajales Dau, Jonathan A. Gross, Steve Habegger, Michael C. Hamilton, Matthew P. Harrigan, Sean D. Harrington, Oscar Higgott, Jeremy Hilton, Markus Hoffmann, Sabrina Hong, Trent Huang, Ashley Huff, William J. Huggins, Lev B. Ioffe, Sergei V. Isakov, Justin Iveland, Evan Jeffrey, Zhang Jiang, Cody Jones, Pavol Juhas, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Tanuj Khattar, Mostafa Khezri, Mária Kieferová, Seon Kim, Alexei Kitaev, Paul V. Klimov, Andrey R. Klots, Alexander N. Korotkov, Fedor Kostritsa, John Mark Kreikebaum, David Landhuis, Pavel Laptev, Kim-Ming Lau, Lily Laws, Joonho Lee, Kenny Lee, Brian J. Lester, Alexander Lill, Wayne Liu, Aditya Locharla, Erik Lucero, Fionn D. Malone, Jeffrey Marshall, Orion Martin, Jarrod R. McClean, Trevor McCourt, Matt McEwen, Anthony Megrant, Bernardo Meurer Costa, Xiao Mi, Kevin C. Miao, Masoud Mohseni, Shirin Montazeri, Alexis Morvan, Emily Mount, Wojciech Mruczkiewicz, Ofer Naaman, Matthew Neeley, Charles Neill, Ani Nersisyan, Hartmut Neven, Michael Newman, Jiun How Ng, Anthony Nguyen, Murray Nguyen, Murphy Yuezhen Niu, Thomas E. O'Brien, Alex Opremcak, John Platt, Andre Petukhov, Rebecca Potter, Leonid P. Pryadko, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Negar Saei, Daniel Sank, Kannan Sankaragomathi, Kevin J. Satzinger, Henry F. Schurkus, Christopher Schuster, Michael J. Shearn, Aaron Shorter, Vladimir Shvarts, Jindra Skruzny, Vadim Smelyanskiy, W. Clarke Smith, George Sterling, Doug Strain, Marco Szalay, Alfredo Torres, Guifre Vidal, Benjamin Villalonga, Catherine Vollgraff Heidweiller, Theodore White, Cheng Xing, Z. Jamie Yao, Ping Yeh, Juhwan Yoo, Grayson Young, Adam Zalcman, Yaxing Zhang, Ningfeng Zhu, and Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, 614(7949):676–681, February 2023. Number: 7949 Publisher: Nature Publishing Group.

[34] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.

[35] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A Quantum Approximate Optimization Algorithm, November 2014. arXiv:1411.4028 [quant-ph].

[36] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):4213, July 2014. Number: 1 Publisher: Nature Publishing Group.

[37] Boaz Barak, Ankur Moitra, Ryan O'Donnell, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Aravindan Vijayaraghavan, David Witmer, and John Wright. Beating the random assignment on constraint satisfaction problems of bounded degree, August 2015. arXiv:1505.03424 [cs].

[38] Robert M. Parrish, Gian-Luca R. Anselmetti, and Christian Gogolin. Analytical Ground- and Excited-State Gradients for Molecular Electronic Structure Theory from Hybrid Quantum/Classical Methods, October 2021. arXiv:2110.05040 [quant-ph].

[39] Robert M. Parrish and Peter L. McMahon. Quantum Filter Diagonalization: Quantum Eigendecomposition without Full Quantum Phase Estimation, September 2019. arXiv:1909.08925 [quant-ph].

[40] Nicholas H. Stair, Renke Huang, and Francesco A. Evangelista. A Multireference Quantum Krylov Algorithm for Strongly Correlated Electrons. *Journal of Chemical Theory and Computation*, 16(4):2236–2245, April 2020. Publisher: American Chemical Society.

[41] Cristian L. Cortes and Stephen K. Gray. Quantum Krylov subspace algorithms for ground- and excited-state energy estimation. *Physical Review A*, 105(2):022417, February 2022. Publisher: American Physical Society.

[42] William J. Huggins, Bryan A. O'Gorman, Nicholas C. Rubin, David R. Reichman, Ryan Babbush, and Joonho Lee. Unbiasing fermionic quantum Monte Carlo with a quantum computer. *Nature*, 603(7901):416–420, March 2022. Number: 7901 Publisher: Nature Publishing Group.

[43] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, September 2021. Number: 9 Publisher: Nature Publishing Group.

[44] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H. Booth, and Jonathan Tennyson. The Variational Quantum Eigensolver: A review of methods and best practices. *Physics Reports*, 986:1–128, 2022.

[45] P. Jordan and E. Wigner. Über das Paulische Äquivalenzverbot. *Zeitschrift für Physik*, 47(9):631–651, September 1928.

[46] Sergey B. Bravyi and Alexei Yu. Kitaev. Fermionic Quantum Computation. *Annals of Physics*, 298(1):210–226, May 2002.

[47] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Daniel Eppens, Edward Farhi, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Matthew P. Harrigan, Alan Ho, Sabrina Hong, Trent Huang, William J. Huggins, Lev Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Erik Lucero, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Hartmut Neven, Murphy Yuezhen Niu, Thomas E. O'Brien, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Doug Strain, Kevin J. Sung, Marco Szalay, Tyler Y. Takeshita, Amit Vainsencher, Theodore White, Nathan Wiebe, Z. Jamie Yao, Ping Yeh, and Adam Zalcman. Hartree-Fock on a superconducting qubit quantum computer. *Science*, 369(6507):1084–1089, August 2020. arXiv:2004.04174 [physics, physics:quant-ph].

[48] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus. *PRX Quantum*, 3(1):010313, January 2022. Publisher: American Physical Society.

[49] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets | Nature.

[50] Harper R. Grimsley, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications*, 10(1):3007, July 2019. Number: 1 Publisher: Nature Publishing Group.

[51] Panagiotis Kl. Barkoutsos, Jerome F. Gonthier, Igor Sokolov, Nikolaj Moll, Gian Salis, Andreas Fuhrer, Marc Ganzhorn, Daniel J. Egger, Matthias Troyer, Antonio Mezzacapo, Stefan Filipp, and Ivano Tavernelli. Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions. *Physical Review A*, 98(2):022322, August 2018. Publisher: American Physical Society.

[52] Harper R. Grimsley, Daniel Claudino, Sophia E. Economou, Edwin Barnes, and Nicholas J. Mayhall. Is the Trotterized UCCSD Ansatz Chemically Well-Defined? | Journal of Chemical Theory and Computation.

[53] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, November 2018. Number: 1 Publisher: Nature Publishing Group.

[54] Ho Lun Tang, V.O. Shkolnikov, George S. Barron, Harper R. Grimsley, Nicholas J. Mayhall, Edwin Barnes, and Sophia E. Economou. Qubit-ADAPT-VQE: An Adaptive Algorithm for Constructing Hardware-Efficient Ans\"atze on a Quantum Processor. *PRX Quantum*, 2(2):020310, April 2021. Publisher: American Physical Society.

[55] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F. Izmaylov. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *The Journal of Chemical Physics*, 152(12), March 2020. _eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5141458/15573883/124114_1_online.pdf.

[56] Edward G. Hohenstein, Robert M. Parrish, and Todd J. Martínez. Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory. *The Journal of Chemical Physics*, 137(4), July 2012. _eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4732310/13843839/044103_1_online.pdf.

[57] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush. Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction. *PRX Quantum*, 2(3):030305, July 2021. Publisher: American Physical Society.

[58] Dominic W. Berry, Craig Gidney, Mario Motta, Jarrod R. McClean, and Ryan Babbush. Qubitization of Arbitrary Basis Quantum Chemistry Leveraging Sparsity and Low Rank Factorization. *Quantum*, 3:208, December 2019. Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.

[59] Edward G. Hohenstein, O. Oumarou, Rachael Al-Saadon, G. L. R. Anselmetti, Maximilian Scheurer, Christian Gogolin, and Robert M. Parrish. Efficient quantum analytic nuclear gradients with double factorization. *The Journal of Chemical Physics*, 158(11), March 2023.

[60] Jeffrey Cohn, Mario Motta, and Robert M. Parrish. Quantum Filter Diagonalization with Compressed Double-Factorized Hamiltonians. *PRX Quantum*, 2(4):040352, December 2021. Publisher: American Physical Society.

[61] Oumarou Oumarou, Maximilian Scheurer, Robert M. Parrish, Edward G. Hohenstein, and Christian Gogolin. Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization, May 2023. arXiv:2212.07957 [quant-ph].

[62] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, October 2020. Number: 10 Publisher: Nature Publishing Group.

[63] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. Fermionic Partial Tomography via Classical Shadows. *Physical Review Letters*, 127(11):110504, September 2021. Publisher: American Physical Society.

[64] Kianna Wan, William J. Huggins, Joonho Lee, and Ryan Babbush. Matchgate Shadows for Fermionic Quantum Simulation, July 2022. arXiv:2207.13723 [physics, physics:quant-ph].

[65] T. E. O'Brien, G. Anselmetti, F. Gkritsis, V. E. Elfving, S. Polla, W. J. Huggins, O. Oumarou, K. Kechedzhi, D. Abanin, R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, K. Anderson, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, D. Bacon, J. C. Bardin, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, J. Campero, Y. Chen, Z. Chen, B. Chiaro, D. Chik, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, S. Demura, I. Drozdov, A. Dunsworth, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, V. S. Ferreira, L. Flores Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, R. Gosula, A. Grajales Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. Hansen, M. P. Harrigan, S. D. Harrington, P. Heu, J. Hilton, M. R. Hoffmann, S. Hong, T. Huang, A. Huff, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, J. Kelly, T. Khattar, M. Khezri, M. Kieferová, S. Kim, P. V. Klimov, A. R. Klots, R. Kothari, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. T. Lill, W. Liu, W. P. Livingston, A. Locharla, E. Lucero, F. D. Malone, S. Mandra, O. Martin, S. Martin, J. R. McClean, T. McCourt, M. McEwen, A. Megrant, X. Mi, A. Mieszala, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, R. Movassagh, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, S. Omonije, A. Opremcak, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, C. Rocque, P. Roushan, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, N. Shutty, V. Shvarts, J. Skruzny, V. Smelyanskiy, W. C. Smith, R. Somma, G. Sterling, D. Strain, M. Szalay, D. Thor, A. Torres, G. Vidal, B. Villalonga, C. Vollgraff Heidweiller, T. White, B. W. K. Woo, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, N. Zhu, N. Zobrist, C. Gogolin, R. Babbush, and N. C. Rubin. Purification-based quantum error mitigation of pair-correlated electron simulations, October 2022. arXiv:2210.10799 [quant-ph].

[66] Lennart Bittel and Martin Kliesch. Training Variational Quantum Algorithms Is NP-Hard. *Physical Review Letters*, 127(12):120502, September 2021. Publisher: American Physical Society.

[67] David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. General parameter-shift rules for quantum gradients. *Quantum*, 6:677, March 2022. Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.

[68] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, March 2019. Publisher: American Physical Society.

[69] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017. arXiv:1412.6980 [cs].

[70] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. *ACM Trans. Math. Softw.*, 23(4):550–560, December 1997. Place: New York, NY, USA Publisher: Association for Computing Machinery.

[71] Javier Gil Vidal and Dirk Oliver Theis. Calculus on parameterized quantum circuits, December 2018. arXiv:1812.06323 [quant-ph].

[72] Mateusz Ostaszewski, Edward Grant, and Marcello Benedetti. Structure optimization for parameterized quantum circuits. *Quantum*, 5:391, January 2021. Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.

[73] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. Quantum Error Mitigation, October 2022. arXiv:2210.00921 [quant-ph].

[74] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error Mitigation for Short-Depth Quantum Circuits. *Physical Review Letters*, 119(18):180509, November 2017. Publisher: American Physical Society.

[75] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, March 2019. Number: 7749 Publisher: Nature Publishing Group.

[76] Suguru Endo, Simon C. Benjamin, and Ying Li. Practical Quantum Error Mitigation for Near-Future Applications. *Physical Review X*, 8(3):031027, July 2018. Publisher: American Physical Society.

[77] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information.* Cambridge University Press, Cambridge ; New York, 10th anniversary ed edition, 2010.

[78] Sergey Bravyi, Jay M. Gambetta, Antonio Mezzacapo, and Kristan Temme. Tapering off qubits to simulate fermionic Hamiltonians, January 2017. arXiv:1701.08213 [quant-ph].

[79] Kanav Setia, Richard Chen, Julia E. Rice, Antonio Mezzacapo, Marco Pistoia, and James D. Whitfield. Reducing Qubit Requirements for Quantum Simulations Using Molecular Point Group Symmetries. *Journal of Chemical Theory and Computation*, 16(10):6091–6097, October 2020. Publisher: American Chemical Society.

[80] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien. Low-cost error mitigation by symmetry verification. *Physical Review A*, 98(6):062339, December 2018. Publisher: American Physical Society.

[81] Thomas E. O'Brien, Stefano Polla, Nicholas C. Rubin, William J. Huggins, Sam McArdle, Sergio Boixo, Jarrod R. McClean, and Ryan Babbush. Error Mitigation via Verified Phase Estimation. *PRX Quantum*, 2(2):020317, May 2021. Publisher: American Physical Society.

[82] Adriano Barenco, André Berthiaume, David Deutsch, Artur Ekert, Richard Jozsa, and Chiara Macchiavello. Stabilization of Quantum Computations by Symmetrization. *SIAM Journal on Computing*, 26(5):1541–1557, October 1997. Publisher: Society for Industrial and Applied Mathematics.

[83] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, March 2019. Number: 7747 Publisher: Nature Publishing Group.

[84] Juan Carlos Garcia-Escartin and Pedro Chamorro-Posada. swap test and Hong-Ou-Mandel effect are equivalent. *Physical Review A*, 87(5):052330, May 2013. Publisher: American Physical Society.

[85] William J. Huggins, Sam McArdle, Thomas E. O'Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. Virtual Distillation for Quantum Error Mitigation. *Physical Review X*, 11(4):041036, November 2021. Publisher: American Physical Society.

[86] Bálint Koczor. Exponential Error Suppression for Near-Term Quantum Devices. *Physical Review X*, 11(3):031057, September 2021. Publisher: American Physical Society.

[87] Mingxia Huo and Ying Li. Dual-state purification for practical quantum error mitigation. *Physical Review A*, 105(2):022427, February 2022. Publisher: American Physical Society.

[88] Zhenyu Cai. Resource-efficient Purification-based Quantum Error Mitigation, July 2021. arXiv:2107.07279 [quant-ph].

[89] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M. Sohaib Alam, Guillermo Alonso-Linaje, B. AkashNarayanan, Ali Asadi, Juan Miguel Arrazola, Utkarsh Azad, Sam Banning, Carsten Blank, Thomas R. Bromley, Benjamin A. Cordier, Jack Ceroni, Alain Delgado, Olivia Di Matteo, Amintor Dusko, Tanya Garg, Diego Guala, Anthony Hayes, Ryan Hill, Aroosa Ijaz, Theodor Isacsson, David Ittah, Soran Jahangiri, Prateek Jain, Edward Jiang, Ankit Khandelwal, Korbinian Kottmann, Robert A. Lang, Christina Lee, Thomas Loke, Angus Lowe, Keri McKiernan, Johannes Jakob Meyer, J. A. Montañez-Barrera, Romain Moyard, Zeyue Niu, Lee James O'Riordan, Steven Oud, Ashish Panigrahi, Chae-Yeun Park, Daniel Polatajko, Nicolás Quesada, Chase Roberts, Nahum Sá, Isidor Schoch, Borun Shi, Shuli Shu, Sukin Sim, Arshpreet Singh, Ingrid Strandberg, Jay Soni, Antal Száva, Slimane Thabet, Rodrigo A. Vargas-Hernández, Trevor Vincent, Nicola Vitucci, Maurice Weber, David Wierichs, Roeland Wiersema, Moritz Willmann, Vincent Wong, Shaoming Zhang, and Nathan Killoran. PennyLane: Automatic differentiation of hybrid quantum-classical computations, July 2022. arXiv:1811.04968 [physics, physics:quant-ph].

[90] Jarrod R. McClean, Zhang Jiang, Nicholas C. Rubin, Ryan Babbush, and Hartmut Neven. Decoding quantum errors with subspace expansions. *Nature Communications*, 11(1):636, January 2020.

[91] Nobuyuki Yoshioka, Hideaki Hakoshima, Yuichiro Matsuzaki, Yuuki Tokunaga, Yasunari Suzuki, and Suguru Endo. Generalized Quantum Subspace Expansion. *Physical Review Letters*, 129(2):020502, July 2022. Publisher: American Physical Society.

[92] Leon Ding, Max Hays, Youngkyu Sung, Bharath Kannan, Junyoung An, Agustin Di Paolo, Amir H. Karamlou, Thomas M. Hazard, Kate Azar, David K. Kim, Bethany M. Niedzielski, Alexander Melville, Mollie E. Schwartz, Jonilyn L. Yoder, Terry P. Orlando, Simon Gustavsson, Jeffrey A. Grover, Kyle Serniak, and William D. Oliver. High-Fidelity, Frequency-Flexible Two-Qubit Fluxonium Gates with a Transmon Coupler, April 2023. arXiv:2304.06087 [quant-ph].

[93] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The Complexity of NISQ, October 2022. arXiv:2210.07234 [quant-ph].

[94] Kevin C. Smith, Eleanor Crane, Nathan Wiebe, and S.M. Girvin. Deterministic Constant-Depth Preparation of the AKLT State on a Quantum Processor Using Fusion Measurements. *PRX Quantum*, 4(2):020315, April 2023. Publisher: American Physical Society.