Accelerating Quantum Chemistry through Compressed Hamiltonian Representation and Efficient Quantum Analytic Energy Gradient Algorithms

Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Oumarou Oumarou

Köln, January 17, 2026

Abstract

In ab-initio quantum chemistry, solving the electronic structure problem of a molecular system of interest to obtain an approximation of the ground and, in some cases, the excited state energies is a core and central task in various quantum chemistry processes. These include, but certainly not limited to, the calculation of the activation energy –that is the minimal energy required to launch a given chemical reaction–, the dipole and multipole moments which describe the polarity of the molecule, performing the optimization of the geometry of the molecule and searching for transition states. Furthermore, many quantum chemistry applications are predicated, not merely on the approximation of the ground state energy, but also on the corresponding first order derivatives. The body of literature of classical algorithms approximating the ground state energy and its first order derivative is extensively rich with a plethora of methods such as self-consistent mean-field methods, coupled-cluster methods, density matrix renormalization group and many more, each of which has intrinsic set of properties and characteristics that makes it suitable for specific tasks and cases. However, classical methods face a practically insurmountable challenge: the **efficient** computation of high accuracy approximation of ground state energies of large system sizes. This is fundamentally because the electronic structure problem provably belongs to the complexity class of QMA-hard problems. Which also means that it is out of reach for quantum algorithms as well in general. However, with certain assumptions that are expected to hold molecules of interest –that so far proved to be classically challenging because the required accuracy set by the standard of chemical accuracy threshold (generally held in quantum chemistry)– quantum algorithms can efficiently treat such cases.

While originally theorized by Richard Feynmann as a potentially efficient way of simulating some classically intractable many-body systems, quantum computing has since proven to have a stronger and richer potential paving the way for the inception of new algorithms such as Shor's algorithm for integer factorization, HHL for solving linear system of equation and Grover's algorithm for constrained binary search, just to name a few. These algorithms provably exhibit an improvement rate that goes respectively exponentially and quadratically in the time complexity. Quantum chemistry is considered a prominent candidate to benefit from quantum computing given the variety of quantum algorithms put forth to solve the electronic structure problem of particular systems of interest. These quantum algorithm proposals can be broadly categorized in three main categories that span the spectrum: from noisy intermediate scale quantum (NISQ) computers to early and full scale fault tolerant quantum ((E)FTQC) computers.

Firstly, the variational quantum eigensolver (VQE) family of quantum algorithms which variationally search for the ground state energy through the optimization of the energy of a parametrized state prepared in a quantum computer using an ansatz. Secondly, the quantum Krylov subspace diagonalization (QKSD) methods, which generate a subspace of the original Hilbert space and efficiently produce the lowest energy in said subspace as its output. Lastly, we have the quantum phase estimation (QPE) algorithms, which provably produce the ground state energy in Heisenberg scaling presenting hence a provable efficient solution.

In this thesis, we firstly set to work on the problem of efficient Hamiltonian linear combination of unitaries (LCU) representation. This is mainly motivated by the wide-range impact an efficient compression of the Hamiltonian can potentially have on several quantum algorithms such as VQE, QPE...etc. We designed a method that efficiently compress the Hamiltonian representation in a way that is beneficially impactful across the set of the aforementioned quantum algorithms. Namely, we motivate and present in Sections 1.3 and 1.4 respectively, the regularized compressed double-factorization (RC-DF) and show how it reduces the energy measurement complexity both in terms of number of distinct observables and number of samples. Moreover, it also reduces the total depth of the quantum phase estimation circuit thanks to its low normalization factor.

Secondly, given the importance of the energy derivatives, in Sections 2.1 and 2.2, we introduce and present an efficient quantum algorithm for the calculation of the analytic gradient of VQE energies of double-factorized Hamiltonians. Furthermore, in Section 2.1, we expand on the utility of RC-DF by presenting a quantum algorithm that determines the analytic gradient of energies measured with respect to RC-DF Hamiltonains.

Third, in Section 3.1, we summarize the theoretical results of purification error-mitigation techniques, namely echo verification and virtual distillation. In Section 3.2, we report on the results of the experimental benchmark, of said techniques, performed on on Google's Sycamore device to validate the theoretical scaling of error suppression and extrapolate the resources needed for classically intractable system sizes.

Fourth, in Section 4 we establish the challenges of computing energy gradients in a quantum Krylov framework. To overcome the problem, we provide, in Section 4.3, an efficient quantum algorithm for calculating energy derivative obtained through QKSD procedures. Our approach shows how the QKSD ground state can be prepared using quantum signal processing (QSP) and reduces the number of observables to O(1) per reduced density matrix (RDM).

Lastly, in light of the importance of quantum RAMs due to their utility in a variety of quantum algorithms that require efficient initial state preparation, as explained in Section 5.5, we address the problematic exponential depth of the bucket-brigade QRAM (when decomposed). We present in Section 5.6 a very efficient Clifford+T decomposition of the bucket-brigade QRAM, which preserves the linear scaling of the depth of the circuit without introducing any additional ancilla.

Partial Publication

- **Oumarou, O.**, Scheurer, M., Parrish, R. M., Hohenstein, E. G., & Gogolin, C. (2024). Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization. Quantum, 8, 1371. CLOCKSS. https://doi.org/10.22331/q-2024-06-13-1371

- **Oumarou, O.**, Ollitrault, P. J., Cortes, C. L., Scheurer, M., Parrish, R. M., & Gogolin, C. (2025). Molecular Properties from Quantum Krylov Subspace Diagonalization. arXiv preprint arXiv:2501.05286.

- Hohenstein, E. G., **Oumarou, O.**, Al-Saadon, R., Anselmetti, G.-L. R., Scheurer, M., Gogolin, C., & Parrish, R. M. (2023). Efficient quantum analytic nuclear gradients with double factorization. The Journal of Chemical Physics, 158(11). https://doi.org/10.1063/5.0137167

- Paler, A., **Oumarou, O.**, & Basmadjian, R. (2020). Parallelizing the queries in a bucket-brigade quantum random access memory. Physical Review A, 102(3). https://doi.org/10.1103/physreva.102.032608

- Scheurer, M., Anselmetti, G.-L. R., **Oumarou, O.**, Gogolin, C., & Rubin, N. C. (2024). Tailored and Externally Corrected Coupled Cluster with Quantum Inputs. Journal of Chemical Theory and Computation, 20(12), 5068–5093. https://doi.org/10.1021/acs.jctc.4c00037

- Anselmetti, G. L., Scheurer, M., **Oumarou, O.**, Gogolin, C., Kiser, M., Vodola, D., & Streif, M. (2024, March). Numerical simulation of statistical properties of matchgate shadows. In APS March Meeting Abstracts (Vol. 2024, pp. T49-013).

- O'Brien, T. E., Anselmetti, G., Gkritsis, F., Elfving, V. E., Polla, S., Huggins, **O. Oumarou**, W. J., ... & Vidal, G. (2023). Purification-based quantum error mitigation of pair-correlated electron simulations. Nature Physics, 19(12), 1787-1792.

Contents

Introduction

In the first part of this section, we briefly introduce the electronic structure problem (ESP) to then show how it is mapped on a quantum computer. The second part focuses on introducing and motivating the importance of the first-order derivatives of the energies produced by the first set of methods addressing the ESP. This division stems from the nature of the work presented in this thesis. Expressly, we propose:

- new quantum algorithms for calculating the energy gradient on a quantum computer.
- new classical methods that optimize the resources of quantum algorithms used to solve the electronic structure problem.

We begin by stressing the point that the quantum chemistry methods and procedures used to generate the parameters defining the problem –such as the spin-orbitals and the active space size– are neither the focus nor of importance to the current thesis. However, a brief introduction will be provided with references for more in-depth explanation.

In ab-initio quantum chemistry, a molecule is modeled by its set of nuclei and electrons (parametrized by their mass and charges) and their interactions. The variation of the state of the molecule –described by its wavefunction– over time is described by the fundamental equation of Schrödinger. Let us start by defining the one-electron and two electron integrals:

$$(p|\hat{h}|q) \coloneqq \int \phi_p^*(r)(-\frac{1}{2}\nabla^2(r) - \sum_m \frac{Z_m}{r - r_m})\phi_q(r)dr \tag{1}$$

$$(pq|rs) \coloneqq \iint \phi_p^*(r_1)\phi_q(r_2)\frac{1}{r_{12}}\phi^*(r_1)_r\phi(r_2)_s dr_1 dr_2 \tag{2}$$

where $\{\phi_i(\cdot), i = 0..n-1\}$, Z_m and r_m are respectively the spatial molecular orbitals in the chosen active space of size n, the charges and positions of nuclei m. As previously stated, the specific manner of generation of these molecular orbitals is of little relevance to the current manuscript. Suffice to say that, usually, the orbitals resulting from the Hartree-Fock method i.e. the Hartree-Fock orbitals are used to generate the Hilbert space in which the wavefunction reside. We briefly describe the Hartree-Fock method below and for further details the references [58, 21, 68] can be checked. We further denote the creation/annihilation operators, the single excitation operators and the modified one-electron integrals as $\hat{a}^\dagger_{i\sigma}/\hat{a}_{i\sigma}$, $\hat{E}_{pq} \coloneqq a^\dagger_{i\alpha}a_{i\alpha} + a^\dagger_{i\beta}a_{i\beta}$. The second-quantized molecular Hamiltonian can then be written as

$$\hat{\mathcal{H}} \coloneqq E_0 + \sum_{\sigma\in\{+,-\},p,q=0}^{n-1} (p|\hat{h}|q)\hat{a}^\dagger_{p\sigma}\hat{a}_{q\sigma} + \frac{1}{2}\sum_{\sigma\in\{+,-\},p,q,r,s=0}^{n-1} (pq|rs)(\hat{a}^\dagger_{p\sigma}\hat{a}^\dagger_{q\sigma}\hat{a}_{r\sigma}\hat{a}_{s\sigma}). \tag{3}$$

Exploiting the anti-commutation relation, we have

$$\hat{\mathcal{H}} = E_0 + \sum_{p,q=0}^{n-1} (p|\hat{h}|q)\hat{E}_{pq} + \frac{1}{2}\sum_{p,q,r,s=0}^{n-1} (pq|rs)(\hat{E}_{pq}\hat{E}_{rs} - \delta_{qr}\hat{E}_{ps}). \tag{4}$$

By defining $\mathcal{F}_{pq} = (p|\hat{h}|q) - \frac{1}{2}\sum_r (pr|qr)$, we have

$$\hat{\mathcal{H}} = E_0 + \sum_{p,q=0}^{n-1} \mathcal{F}_{pq}\hat{E}_{pq} + \frac{1}{2}\sum_{p,q,r,s=0}^{n-1} (pq|rs)\hat{E}_{pq}\hat{E}_{rs} \tag{5}$$

and the subsequent time-independent Schrödinger equation in the Born-Oppenheimer regime is written as

$$\hat{\mathcal{H}}|\Psi\rangle = E_\Psi|\Psi\rangle. \tag{6}$$

The ESP consists of solving the time independent Schrödinger equation for the lowest energy i.e. finding the pair of ground state $|\Psi_0\rangle$ and ground-state energy E_0 over the exponentially large Hilbert space of electronic configurations:

$$\hat{\mathcal{H}}|\Psi_0\rangle = E_0|\Psi_0\rangle. \tag{7}$$

Solving the ESP is an NP-hard problem [66]. There are, however, various classical methods that approximate the ground state energy such as wavefunction-based methods which parametrize the wavefunction in a way that helps ease the problem by narrowing down the dimension of the subspace to search. These methods are of special interest due to the similarity they share with some of the quantum algorithms we shall present below, namely

the variational quantum eigensolver and quantum Krylov subspace diagonalization. To list a few, we have the Hartree-Fock method [21, 58] as probably the most famous example. The Hartree-Fock method produces the lowest energy within the subspace of product states or the single slater determinant subspace. The Hartree-Fock result is often used as a stepping stone for the so-called post-Hartree-Fock methods to improve the estimates of the ground-state energy accuracy. For example, the couple clustered methods [14, 57] are often used as post-Hartree-Fock method. Other wavefunction-based methods include the Monte-Carlo family [4, 25], density matrix renormalization group (DMRG) [76, 13, 78].. etc

In addition to calculating the ground-state energy to high accuracies, determining the first-order energy derivative is a crucial task and of paramount importance, especially in chemistry. For instance, we have the nuclear gradient, which is the first-order energy derivative with respect to the cartesian coordinates of the nuclei, employed in methods such as molecular geometry optimization and molecular dynamics. Hence, it is also important to determine how to calculate the energy gradient in an efficient manner.

On the other hand, since the inception of quantum computing, the promise to solve eletronic structure problem of molecules of interest in chemistry to arbitrary accuracy in polynomial time complexity has been a key goal. The set of quantum computing methods that tackles the ESP can be laid on a spectrum that stretches from noisy intermediate-scale quantum (NISQ) to fault-tolerant quantum computing (FTQC) methods. For instance, the VQE family of methods is generally classified as NISQ procedures due to their built-in mitigation of coherent errors. For other types of errors, several error mitigation techniques have been developed to mitigate the noise effect. On the other hand, the QPE algorithm belongs to the other end of spectrum i.e. FTQC as it employs deep and resource intensive circuits. Error-correction is therefore required to execute the algorithm. Relatively recently, another hybrid era started to take shape where we have algorithms with error-corrected shallow circuits. This umbrella covers the set of statistical quantum phase estimation (SPE) and QKSD methods.

In addition, given the importance of first-order energy derivatives, methods for calculating said derivatives, of various methods of the overarching families of QPE and VQE, have been developed. In [52], the authors develop an analytical and efficient hybrid quantum method to calculate the nuclear gradient and more generally the relaxed-RDMs of ground and excited states produced by MC-VQE [53]. In [24], we present an analytical and efficient method for determining the relaxed RDMs of energies estimated with exact and truncated double-factorization (DF) measurement methods. Our method is detailed in Section 2.1. For the QPE, the methods presented in [44, 70, 64] propose methods to measure the force operator in a QPE-like circuit, where the unitary–by construction–encodes the force operator in its spectrum. Lastly, to our knowledge, we were the first to propose an analytical gradient approach of the QKSD methods. In [49] we present how we use QSP to efficiently determine the relaxed-RDMs necessary for the nuclear gradient for instance. We also present this work in Section 4.

In order to proceed further and present the results obtained in this thesis, we need to introduce a few preliminary concepts. First, the map that transform our problem from the second quantized fermionic presentation to qubit operators. For the rest of the manuscript we constrain ourselves to the Jordan-Wigner mapping. Each spin molecular orbital is mapped to a qubit where the states $|1\rangle/|0\rangle$ indicate that the orbital is occupied/non-occupied. There are two prominent qubit orderings that are conventionally used. Namely, the interleaved and sequential qubit orderings. Firstly, the qubits must be enumerated in some fashion that usually takes into account the properties of the device and the circuit to be executed, such as connectivity and the gates' layouts. Let us denote the set of $2n$ qubits by indexing them with the set $\{i \in [0, \ldots, 2n-1]\}$ of integers. In the interleaved ordering, the even and odd qubits represent the spin-up and spin-down fermions. Hence, each pair of adjacent qubits $(2i, 2i+1)$ belong to the same spatial orbital. For the rest of the thesis we will be using the interleaved qubit ordering. Moreover, another important notion is the ordering of the fermions, also referred to as the Jordan-Wigner mapping (which need not be identical to the qubit ordering). We will be using the sequential Jordan-Wigner ordering, i.e., the ordering where all the spin-up and spin-down fermions are adjacent and indexed, respectively, with the sets $\{i \in [0, \ldots, n-1]\}$ and $\{i \in [n, \ldots, 2n-1]\}$. We note the qubit ordering and the Jordan-Wigner fermions ordering by $f(\cdot)$ and $g(\cdot)$ respectively. The creation/annihilation operators $\hat{a}^\dagger_{i\sigma}/\hat{a}_{i\sigma}$ are mapped to

$$\hat{a}^\dagger_{i\sigma} \coloneqq \left(\prod_{j=0}^{g(i,\sigma)-1} \hat{Z}_j\right) \frac{\hat{X}_{f(i,\sigma)} + i\hat{Y}_{f(i,\sigma)}}{2} \tag{8}$$

$$\hat{a}_{i\sigma} \coloneqq \left(\prod_{j=0}^{g(i,\sigma)-1} \hat{Z}_j\right) \frac{\hat{X}_{f(i,\sigma)} - i\hat{Y}_{f(i,\sigma)}}{2} \tag{9}$$

where $\mathcal{P} \coloneqq \{\hat{X}_i, \hat{Y}_i, \hat{Z}_i, \hat{I}_i\}$ is the set of Pauli matrices defined as

$$\hat{X}_i \coloneqq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \hat{Y}_i \coloneqq \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \hat{Z}_i \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } \hat{I}_i \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{10}$$

The product of $\hat{Z}_i$ matrices in Eq (9) is to enforce the anti-commutation relation. It should be noted that the Bravyi-Kitaev mapping [10] is also another prominent mapping option. Moreover, a special operator that is relevant for later sections is the diagonal operator $\hat{E}_{kk} = \hat{a}_k^\dagger \hat{a}_k$ which is mapped to

$$\hat{E}_{kk} = \sum_\sigma \left(\left(\prod_{j=0}^{g(k)-1} \hat{Z}_j \right) \frac{\hat{X}_{f(k,\sigma)} + i\hat{Y}_{f(k,\sigma)}}{2} \left(\prod_{j=0}^{g(k)-1} \hat{Z}_j \right) \frac{\hat{X}_{f(k,\sigma)} - i\hat{Y}_{f(k,\sigma)}}{2} \right) \tag{11}$$

$$= \hat{I} - \frac{\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)}}{2} \tag{12}$$

where we used $\left(\prod_{j=0}^{k-1} \hat{Z}_j\right)^2 = \hat{I}$ and $\frac{\hat{X}_k + i\hat{Y}_k}{2} \frac{\hat{X}_k - i\hat{Y}_k}{2} = \frac{I - Z_k}{2}$ to transition between the first and second equations.

Lastly, we introduce the relaxed one- and two-body RDM of a given general state $|\Psi\rangle$, with energy $\langle\psi|\hat{\mathcal{H}}|\psi\rangle$. The relaxed one- and two-body RDMs are defined as

$$\gamma_{pq} \coloneqq \frac{d\langle\psi|\hat{\mathcal{H}}|\psi\rangle}{d(p|\hat{h}|q)} \tag{13}$$

$$\Gamma_{pqrs} \coloneqq \frac{d\langle\psi|\hat{\mathcal{H}}|\psi\rangle}{d(pq|rs)}. \tag{14}$$

The importance of these quantities is that, using the chain-rule, the energy derivative can be obtained by contracting these relaxed RDMs with the first-order derivatives of the one- and two-electron integrals. The latter is obtained classically, whereas the relaxed RDMs are obtained through quantum algorithms. The relaxed RDMs are independent of the differentiation parameter. Thus, we can obtain, from the same relaxed RDMs, all energy gradients that are of interest. Including but not limited to all local or global extrema in the potential energy surface.

Outline

As this thesis follows a cumulative format, the results are presented across chapters, each of which consists of a paper preceded by introductory sections.

In Section 1 we introduce the VQE method where the anasatz and optimization modules are presented in Section 1.2. In Section 1.3 we present the measurement module with a special focus on double-factorizion methods and layout the particular issues of high variances and number of terms of RC-DF and exact double-factorization (XDF) respectively. In Section 1.4 we present our solution RC-DF which reduces both number of observables to measure and their variances.

Next, we outline, in Section 2.1, our main result to compute the gradient of double-factorized Hamiltonian energies. Namely, the full Lagrangian equation of the energy at stationary point. The proofs and the numerical benchmarking are subsequently compiled in Section 2.2.

In Section 3, we start with a preliminary chapter in Section 3.1 that introduce echo-verification and virtual distillation. Then in Section 3.2 we present the numerical results of experimentally benchmarking both error-mitigation on VQE circuits facilitating a comparative evaluation of their performance and a projective prediction of the sampling overhead needed for scaling up the size of the systems.

In Section 4 we examine the computation of gradients of QKSD energies. Initially, Section 4.1 presents the QKSD algorithm with the Chebychev polynomials. Following this, we give a concise summary of key results as well as our quantum algorithm for determining the gradient of QKSD energies. In Section 4.3 the results are detailed with proofs and numerical results.

We present preliminaries on state-of-the art QPE including the block-encoding, qubitization in Sections 5.2 and 5.3, and conclude in Section 5.4 with walk-operator-based QPE. In Section 5.5, we summarize the main results of our linear depth Clifford+T decomposion of bucket brigade QRAM.

We note that the preliminary material on QPE presented in this chapter is also pertinent for the discussion in Section 6, particularly in relation to the effect of RC-DF on QPE.

Finally, in Section 6 we provide a compact overview of all the results put forth in this thesis. We also discuss the feasibility of achieving quantum advantage in quantum chemistry in light of these findings.

1 Hamiltonian Compressing

In this section, we shall introduce the challenging overhead that molecular Hamiltonian present within the framework of VQE and therefore the necessity of efficient Hamiltonian compressing methods to surmount this issue. We begin by setting up the context with the introduction of the general and common workflow of VQE.

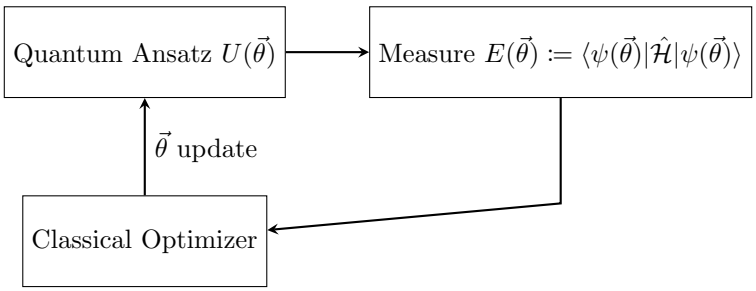


Figure 1: VQE workflow

We present the main constituents of VQE. We shall particularly focus on the measurement module where the aforementioned overhead arises. We then present our solution RC-DF to overcome this problem.

1.1 Variational Quantum Eigensolver & Energy Gradient

The variational quantum eigensolver is a cluster of methods whose main components consist of: (1) an ansatz, (2) energy measurement method and (3) an optimization procedure. Starting from an initial state $|\psi_0\rangle$, the VQE procedure then seeks to find the ground state and the ground-state energy through gradual and incremental minimization of the energy of the wavefunction prepared by the ansatz by changing the parameters according to the optimization procedure. See Algorithm 1 and FIG 1.1.

Let us define the set of parameters of the ansatz gates as $\vec{\theta} \coloneqq (\theta_0, \theta_1, ..\theta_d)$ and $U(\vec{\theta})$ as the unitary of the ansatz. Subsequently, we define the state prepared by the ansatz with parameters $\vec{\theta}$ as $|\psi(\vec{\theta})\rangle \coloneqq U(\vec{\theta})|\psi_0\rangle$. The general algorithm of VQE is the following:

Algorithm 1 Variational Quantum Eigensolver

Require: $|\psi_0\rangle, \vec{\theta}_0$ ▷ Initial quantum state and initial parameters
Ensure: ϕ_{opt}, E_{opt} ▷ Optimized parameters and Optimal energy
1: $|\psi(\vec{\theta})\rangle \leftarrow U(\vec{\theta})|\psi_0\rangle$
2: $E(\vec{\theta}) \leftarrow \langle\psi(\vec{\theta})|\hat{\mathcal{H}}|\psi(\vec{\theta})\rangle$
3: $\vec{\theta} \leftarrow update(\vec{\theta})$
4: Repeat until convergence criteria is met

In the following we shall present each module of the VQE framework: **ansatz**, (2)**energy measurement** and (3)**Optimization** with a special focus on the energy measurement module, seeing that it is optimized by our RC-DF method and we shall present some state of the art methods in that regard and comment on the performance gains of our method.

1.2 VQE Ansatz & Optimization

The edge that a VQE method holds over its classical counterparts is in the expressivity of its ansatz. The expressivity of an ansatz is defined as the Hilbert subspace of states that the ansatz can prepare. From this definition, we can see that it is crucial that an ansatz can reach the ground-state energy or be within δ of it, energy-wise. A universal ansatz is an ansatz that can reach every state within the Hilbert space, i.e. prepare every possible wavefunction. For instance, it is well established that a universal ansatz can be built such as the coupled-cluster ansatz with all the excitations [54]. However, there are a set of characteristics and constraints that make the task of designing an expressive ansatz a challenging task. For practical reasons, a VQE ansatz needs to be shallow, meaning that the ansatz needs to have a depth that is polynomial in the system size (active space size).

We list a set of prominent VQE ansätze proposals with brief description, seeing that our work when it comes to VQE focuses essentially on the measurement module of the VQE framework. We have the **QNP fabric** [3] that employs a set of chemically inspired gates: the Givens rotations and pair exchanged gates. These gates conserve symmetries, namely $\hat{N}_\alpha$, $\hat{N}_\beta$ and $\hat{S}^2$ which are the number of particles α, β and spin-squared operators. The ansatz has a fabric layout, meaning a set of repeating alternating patterns whereby the parametrized gates act only locally and hence no long interactions that would otherwise require expensive swap gates. The benchmark results show that the QNP fabric achieves chemical accuracy with relatively shallow depth.

1.3 VQE Energy Measurement

The VQE energy measurement is a crucial component that drives the cost of the overall procedure. As presented in algorithm 1, the energy measurement is performed in each iteration. Therefore, performing this step in an efficient way is of paramount importance. We have in Section introduced the Jordan-Wigner mapping as a function that maps fermionic operators acting on fermions to Pauli operators acting on qubits. Directly substituting in the Hamiltonian equation, we obtain a "qubit Hamiltonian" as a LCU

$$\hat{\mathcal{H}} = \sum_{k=0}^{K} c_k \hat{P}_k \tag{15}$$

where the prefactors are real numbers $c_k \in \mathrm{R}$ and is a Pauli string $\hat{P}_k \in \bigotimes_{i=0}^{2n-1} \mathcal{P}$. To then calculate the energy of the state $|\psi(\vec{\theta})\rangle$, the most straight-forward method is to measure the expectation value of each term in the LCU Eq (15) and sum them using the linearity of the expectation value operation

$$\langle\psi(\vec{\theta})|\hat{\mathcal{H}}|\psi(\vec{\theta})\rangle = \sum_{k=0}^{K} c_k \langle\psi(\vec{\theta})|\hat{P}_k|\psi(\vec{\theta})\rangle. \tag{16}$$

Now, to determine the expectation values $\langle\psi(\vec{\theta})|\hat{P}_k|\psi(\vec{\theta})\rangle$, we need to prepare $|\psi(\vec{\theta})\rangle = U(\vec{\theta})|\psi_0\rangle$, rotate into the diagonal basis of $\hat{P}_k$ (which is the tensor product of each Pauli in the Pauli string $\hat{P}_k$), measure in the $\hat{Z}_k$ basis for all indices k and then use the sample mean to estimate the expectation value. As can already be seen, we need K distinct circuits to prepare $|\psi(\vec{\theta})\rangle$. Therefore, having a small number of terms to measure is desirable. However, that is not sufficient to have an efficient energy estimation method. When measuring, we incur an error due to the finite number of samples we use. This error is referred to as the shot noise. The shot noise is a statistical property of our estimator (16) and hence it depends on the variance of said estimator. More specifically, the set of estimators $\langle\psi(\vec{\theta})|\hat{P}_k|\psi(\vec{\theta})\rangle$ are independent. Therefore, because Eq (15) is an LCU, the variance of the estimator of the energy Eq (16) is a linear combination of the variances of $\langle|\hat{P}_k|\rangle$

$$var(\langle\psi(\vec{\theta})|\hat{\mathcal{H}}|\psi(\vec{\theta})\rangle) = \sum_{k=0}^{K} c_k^2 var(\langle\psi(\vec{\theta})|\hat{P}_k|\psi(\vec{\theta})\rangle). \tag{17}$$

State-of-the-art methods can be divided to two categories. First, we have Pauli grouping schemes [74, 27, 60]. The second category explores the 8-fold symmetry of the $(pq|rs)$ tensor to reduce the costly part of the Hamiltonian from a four-index sum to a three-index sum. These are called DF methods. First, let us consider the $(pq|rs)$ tensor as a $N^2 \times N^2$ matrix. $(pq|rs)$ is symmetric, hence it can be diagonalized with a unitary matrix. We note by (g_t, V^t) the pair of eigenvalues and their corresponding eigenvectors. We can then write $(pq|rs)$ as

$$(pq|rs) = \sum_{t=0}^{n^2-1} g_t V_{pq}^t V_{rs}^t. \tag{18}$$

Now, let us reshape the eigenvectors V^t into $n \times n$ matrices $\forall t$. In Lemma 1 in [47], we prove that the 8-fold symmetry of the $(pq|rs)$ tensor is equivalent to the symmetry of the matrices V_{pq}^t ($V_{pq}^t = V_{qp}^t \ \forall t$) and that the size of the set $\{g^t \neq 0, t\}$ is in fact at most $T \coloneqq \frac{n(n-1)}{2}$. Therefore, V^t can be diagonalized with $n \times n$ unitary rotations. Let us note the eigenvalue and eigenvector pair by (Λ^t, U^t). We then have

$$V_{pq}^t = \sum_{k=0}^{n-1} \Lambda_k^t U_{pk}^t U_{qk}^t. \tag{19}$$

Plugging Eq (19) into Eq (18), we get

$$(pq|rs) = \sum_{t=0}^{T} \sum_{k,l=0}^{n-1} g^t \Lambda_k^t \Lambda_l^t U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t. \tag{20}$$

We define $Z_{kl}^t \coloneqq g^t \Lambda_k^t \Lambda_l^t$ and end up with the DF expression of the two-electron integrals tensor

$$(pq|rs) = \sum_{t=0}^{T} \sum_{k,l=0}^{n-1} Z_{kl}^t U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t. \tag{21}$$

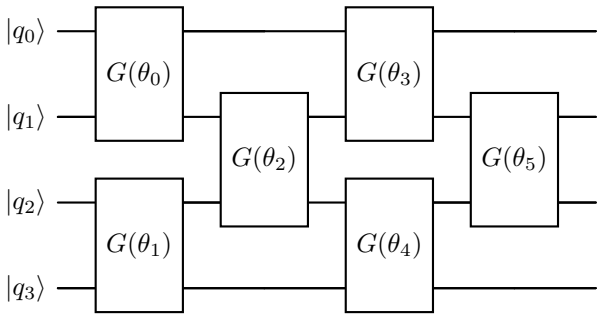


Figure 2: Orbital rotation circuit on 4 qubit example.

Now, let us substitute $(pq|rs)$ in the $(pq|rs)$ summand in Eq (5) with Eq (21)

$$\sum_{p,q,r,s=0}^{n-1}(pq|rs)\hat{E}_{pq}\hat{E}_{rs} = \sum_{p,q,r,s=0}^{n-1}\sum_{t=0}^{T}\sum_{k,l=0}^{n-1} Z_{kl}^t U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t \hat{E}_{pq}\hat{E}_{rs} \tag{22}$$

$$= \sum_{t=0}^{T}\sum_{k,l=0}^{T}\sum_{p,q,r,s=0}^{n-1} Z_{kl}^t U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t \hat{a}_p^\dagger \hat{a}_q \hat{a}_r^\dagger \hat{a}_s \tag{23}$$

$$= \sum_{t=0}^{T}\sum_{k,l=0}^{n-1} Z_{kl}^t \sum_{p,q,r,s=0}^{n-1} U_{pk}^t \hat{a}_p^\dagger U_{qk}^t \hat{a}_q U_{rl}^t \hat{a}_r^\dagger U_{sl}^t \hat{a}_s \tag{24}$$

$$= \sum_{t=0}^{T}\sum_{k,l=0}^{n-1} Z_{kl}^t \left(\sum_{p=0}^{n-1} U_{pk}^t \hat{a}_p^\dagger\right)\left(\sum_{q=0}^{n-1} U_{qk}^t \hat{a}_q\right)\left(\sum_{r=0}^{n-1} U_{rl}^t \hat{a}_r^\dagger\right)\left(\sum_{s=0}^{n-1} U_{sl}^t \hat{a}_s\right). \tag{25}$$

Let us note the rotated fermionic creation annihilation operators by $\hat{c}_i^{t\dagger} = \sum_{k=0}^{n-1} U_{ik}^t \hat{a}_k^\dagger$ and $\hat{c}_i^t = \sum_{k=0}^{n-1} U_{ik}^t \hat{a}_k$. Picking up from Eq (25), we have

$$\sum_{p,q,r,s=0}^{n-1}(pq|rs)\hat{E}_{pq}\hat{E}_{rs} = \sum_{t=0}^{T}\sum_{k,l=0}^{n-1} Z_{kl}^t \hat{c}_k^{t\dagger}\hat{c}_k^t \hat{c}_l^{t\dagger}\hat{c}_l^t \tag{26}$$

$$= \sum_{t=0}^{T}\sum_{k,l=0}^{n-1} Z_{kl}^t \hat{E}_{kk}^t \hat{E}_{ll}^t. \tag{27}$$

Note that the $(pq|rs)$ summand in Eq (5) is transformed into a summand of diagonal operators, and, as we shall see shortly, the estimator can be transformed into $O(n^2)$ commutative terms that can be measured jointly. Before that however, we need a way to map the rotation over the fermions to a qubit operator. This can be achieved using a network of two qubit gates called Givens rotations $G(\theta)$

$$G(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) & 0 \\ 0 & -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{28}$$

In [15] a method for deriving the parameters of the resulting circuit is presented. The circuit is an n deep circuit with $n(n-1)/2$ Givens rotations (See FIG 2).

Similarly, the $\mathcal{F}$ is a symmetric tensor and can be diagonalized with unitary matrices. Let us note by $U^\varnothing$ and $(g_i^\varnothing, i = 0..n-1)$ the unitary rotations and the set of eigenvalues of $\mathcal{F}$. We then have

$$\mathcal{F}_{pq} = \sum_{k=0}^{n-1} g_k^\varnothing U_{pk}^\varnothing U_{qk}^\varnothing \tag{29}$$

and the one-electron summand in Eq (5) can be written as

$$\sum_{p,q=0}^{n-1}\mathcal{F}_{pq}\hat{E}_{pq} = \sum_{k=0}^{n-1} g_k^\varnothing \hat{E}_{kk}^\varnothing. \tag{30}$$

Combining Eq (30) and Eq (27) we have the DF expression of the Hamiltonian

$$\hat{\mathcal{H}} = E_0 + \sum_{k=0}^{n-1} g_k^{\varnothing} U^{\varnothing\dagger} \hat{E}_{kk}^{\varnothing} U^{\varnothing} + \frac{1}{2} \sum_{t=0}^{T} \sum_{k,l=0}^{n-1} Z_{kl}^t U^{t\dagger} \hat{E}_{kk}^t \hat{E}_{ll}^t U^t. \tag{31}$$

Now, noting $\hat{U}^t$ the Givens network circuit unitary and using Eq (12) we have

$$\begin{aligned}\hat{\mathcal{H}} = E_0 &+ \sum_{k=0}^{n-1} g_k^{\varnothing} \hat{U}^{\varnothing\dagger} \left(I - \frac{\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)}}{2} \right) \hat{U}^{\varnothing} \\ &+ \frac{1}{2} \sum_{t=0}^{T} \sum_{k,l=0}^{n-1} Z_{kl}^t \hat{U}^{t\dagger} \left(I - \frac{\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)}}{2} \right) \left(I - \frac{\hat{Z}_{f(l,\alpha)} + \hat{Z}_{f(l,\beta)}}{2} \right) \hat{U}^t. \end{aligned} \tag{32}$$

Hence, we get an XDF LCU form of the Hamiltonian. We will see shortly that we can further minimize the number of terms in XDF but now the first thing to notice is that the orbital rotations $\hat{U}^{\varnothing}$ and $\hat{U}^t$ can be brought out of the one-electron sum and the inner sum of the two-electron sum respectively. These are called t-leafs and we have $n^2 + 1$ t-leafs in Eq (31). The terms inside each t-leafs commute because they are diagonal. Therefore, they can be measured simultaneously. The energy expression in the DF form is

$$\langle\psi(\vec{\theta})|\hat{\mathcal{H}}|\psi(\vec{\theta})\rangle = E_0 + \sum_{k=0}^{N-1} g_k^{\varnothing} + \frac{1}{2} \sum_{t=0}^{N^2-1} \sum_{k,l=0}^{N} Z_{kl}^t - \frac{1}{2} \langle\psi(\vec{\theta})|\hat{U}^{\varnothing\dagger} \sum_{k=0}^{N-1} \left(\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)} \right) \hat{U}^{\varnothing} |\psi(\vec{\theta})\rangle + \tag{33}$$

$$\frac{1}{8} \sum_t \hat{U}^{t\dagger} \left(\sum_{k,l} Z_{kl}^t \left(\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)} \right) \left(\hat{Z}_{f(l,\alpha)} + \hat{Z}_{f(l,\beta)} \right) \right) \hat{U}^t \tag{34}$$

$$= \widetilde{E}_0 + \frac{1}{2} \langle\psi(\vec{\theta})|\hat{U}^{\varnothing\dagger} \sum_{k=0}^{N-1} \hat{Z}_k \hat{U}^{\varnothing} |\psi(\vec{\theta})\rangle + \frac{1}{8} \sum_t \langle\psi(\vec{\theta})|\hat{U}^{t\dagger} \left(-2 \sum_k (\sum_l Z_{kl}^t) \hat{Z}_k + \sum_{k,l} Z_{kl}^t \hat{Z}_k \hat{Z}_l \right) \hat{U}^t |\psi(\vec{\theta})\rangle. \tag{35}$$

Using the following identity, we can reduce the number of single $\hat{Z}_k$ terms by combining them in one leaf

$$\hat{E}_{kk}\hat{E}_{ll} = -\hat{I} + \hat{E}_{kk} + \hat{E}_{ll} + \frac{1}{4}(\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)})(\hat{Z}_{f(l,\alpha)} + \hat{Z}_{f(k,\beta)}) \tag{36}$$

and we can rewrite the Hamiltonian from Eq (31) as

$$\begin{aligned}\hat{\mathcal{H}} =& E_0 - \frac{1}{2} \sum_{t=0}^{T} \sum_{k,l=0}^{n-1} Z_{kl}^t + \sum_{k=0}^{n-1} g_k^{\varnothing} \hat{U}^{\varnothing\dagger} \hat{E}_{kk}^{\varnothing} \hat{U}^{\varnothing} + \sum_{t=0}^{T} \sum_{k=0}^{n-1} \left(\sum_{l=0}^{n-1} Z_{kl}^t \right) \hat{U}^{t\dagger} \hat{E}_{kk}^t \hat{U}^t \\ &+ \frac{1}{8} \sum_{t=0}^{T} \sum_{k,l=0}^{n-1} Z_{kl}^t \hat{U}^{t\dagger} (\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)})(\hat{Z}_{f(l,\alpha)} + \hat{Z}_{f(l,\beta)}) \hat{U}^t. \end{aligned} \tag{37}$$

Now, using Eq (21) we have a couple of interesting identities

$$\sum_{p,q=0}^{n-1} (pp|qq) = \sum_{t=0}^{T} \sum_{k,l=0}^{n-1} Z_{kl}^t \tag{38}$$

$$\sum_{k=0}^{n-1} (pq|kk) = \sum_{t=0}^{T} \sum_{k=0}^{n-1} U_{pk}^t U_{qk}^t \left(\sum_{l=0}^{n-1} Z_{kl}^t \right). \tag{39}$$

Implying that

$$\sum_{t=0}^{T} \sum_{k=0}^{n-1} \left(\sum_{l=0}^{n-1} Z_{kl}^t \right) \hat{U}^{t\dagger} \hat{E}_{kk}^t \hat{U}^t = \sum_{t=0}^{T} \sum_{k=0}^{n-1} \left(\sum_{l=0}^{n-1} Z_{kl}^t \right) \sum_{p,q=0}^{n-1} U_{pk}^t U_{qk}^t \hat{E}_{pq} \tag{40}$$

$$= \sum_{p,q=0}^{n-1} \sum_{t=0}^{T} \sum_{k=0}^{n-1} \left(\sum_{l=0}^{n-1} U_{pk}^t U_{qk}^t Z_{kl}^t \right) \hat{E}_{pq} \tag{41}$$

$$= \sum_{p,q=0}^{n-1} \left(\sum_{k=0}^{n-1} (pq|kk) \right) \hat{E}_{pq}. \tag{42}$$

Therefore, back-transforming the one-electron part and adding Eq (42) we get

$$\sum_{k=0}^{n-1} g_k^{\varnothing} U^{\varnothing} \hat{E}_{kk} U^{\varnothing} + \sum_{t=0}^{T} \sum_{k=0}^{n-1} \left(\sum_{l=0}^{n-1} Z_{kl}^t \right) \hat{U}^{t\dagger} \hat{E}_{kk}^t \hat{U}^t = \sum_{p,q=0}^{n-1} \left(\mathcal{F}_{pq} + \sum_{k=0}^{n-1} (pq|kk) \right) \hat{E}_{pq}. \tag{43}$$

We note $\mathcal{G}_{pq} \coloneqq \mathcal{F}_{pq} + \sum_k (pq|kk)$ and its eigenvalues and diagonalizing unitary pair $(g_k^{\varnothing}, U^{\varnothing})$. As a result, by diagonalizing $\mathcal{G}$ and using Eq (38), we have the following expression of the Hamiltonian

$$\begin{aligned} \hat{\mathcal{H}} = E_0 + \sum_{k=0}^{n-1} g_k^{\varnothing} - \frac{1}{2} \sum_{p,q=0}^{n-1} (pp|qq) - \frac{1}{2} \sum_{k=0}^{n-1} g_k^{\varnothing} \hat{U}^{\varnothing\dagger} (\hat{Z}_k + \hat{Z}_{k+1}) \hat{U}^{\varnothing} \\ + \frac{1}{8} \sum_{t=0}^{T} \sum_{k,l=0}^{n-1} Z_{kl}^t \hat{U}^{t\dagger} \left(\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)} \right) \left(\hat{Z}_{f(l,\alpha)} + \hat{Z}_{f(l,\beta)} \right) \hat{U}^t. \end{aligned} \tag{44}$$

Lastly, we know that $\sum_{k=0}^{n-1} g_k^{\varnothing} = \sum_{k=0}^{n-1} \mathcal{G}_{kk} = \mathrm{Tr}(\mathcal{G})$ and $\frac{1}{2} \left(\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)} \right)^2 = \frac{1}{2} \left(\hat{I} + \hat{Z}_{f(k,\alpha)} \hat{Z}_{f(k,\beta)} \right)$. So, combining these identities in Eq (44) we get

$$\begin{aligned} \hat{\mathcal{H}} &= E_0 + \sum_{k=0}^{n-1} \mathcal{G}_{kk} - \frac{1}{2} \sum_{p,q=0}^{n-1} (pp|qq) - \frac{1}{2} \sum_{k=0}^{n-1} g_k^{\varnothing} \hat{U}^{\varnothing\dagger} (\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)}) \hat{U}^{\varnothing} + \frac{1}{4} \sum_{t=0}^{T} \sum_{k=0}^{n-1} \left(\hat{I} + \hat{Z}_{f(k,\alpha)} \hat{Z}_{f(k,\beta)} \right) + \\ &\frac{1}{8} \sum_{t=0}^{T} \sum_{k \neq l}^{n-1} Z_{kl}^t \hat{U}^{t\dagger} \left(\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)} \right) \left(\hat{Z}_{f(l,\alpha)} + \hat{Z}_{f(l,\beta)} \right) \hat{U}^t \end{aligned} \tag{45}$$

$$\begin{aligned} &= E_0 + \sum_{p=0}^{n-1} \mathcal{F}_{pp} + \sum_{p,q=0}^{n-1} (pp|qq) - \frac{1}{2} \sum_{p,q=0}^{n-1} (pp|qq) + \frac{1}{4} \sum_{t=0}^{T} \sum_{k=0}^{n-1} Z_{kk}^t - \frac{1}{2} \sum_{k=0}^{n-1} g_k^{\varnothing} \hat{U}^{\varnothing\dagger} (\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)}) \hat{U}^{\varnothing} + \\ &\frac{1}{4} \sum_{t=0}^{T} \sum_{k=0}^{n-1} Z_{kk}^t \hat{U}^{t\dagger} \hat{Z}_{f(k,\alpha)} \hat{Z}_{f(k,\beta)} \hat{U}^t + \frac{1}{8} \sum_{t=0}^{T} \sum_{k \neq l}^{n-1} Z_{kl}^t \hat{U}^{t\dagger} \left(\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)} \right) \left(\hat{Z}_{f(l,\alpha)} + \hat{Z}_{f(l,\beta)} \right) \hat{U}^t \end{aligned} \tag{46}$$

$$\begin{aligned} &= E_0 + \sum_{k=0}^{n-1} (p|\hat{h}|p) - \frac{1}{2} \sum_{p,q=0}^{n-1} (pq|pq) + \frac{1}{2} \sum_{p,q=0}^{n-1} (pp|qq) + \frac{1}{4} \sum_{p,q=0}^{n-1} (pq|pq) - \frac{1}{2} \sum_{k=0}^{n-1} g_k^{\varnothing} \hat{U}^{\varnothing\dagger} (\hat{Z}_k + \hat{Z}_{k+1}) \hat{U}^{\varnothing} + \\ &\frac{1}{4} \sum_{t=0}^{T} \sum_{k=0}^{n-1} Z_{kk}^t \hat{U}^{t\dagger} \hat{Z}_{f(k,\alpha)} \hat{Z}_{f(k,\beta)} \hat{U}^t + \frac{1}{4} \sum_{t=0}^{n-1} \sum_{k<l}^{n-1} Z_{kl}^t \hat{U}^{t\dagger} \left(\hat{Z}_{f(k,\alpha)} + \hat{Z}_{f(k,\beta)} \right) \left(\hat{Z}_{f(l,\alpha)} + \hat{Z}_{f(l,\beta)} \right) \hat{U}^t \end{aligned} \tag{47}$$

$$\begin{aligned} &= E_0 + \sum_{k=0}^{n-1} (p|\hat{h}|p) - \frac{1}{4} \sum_{p,q=0}^{n-1} (pq|pq) + \frac{1}{2} \sum_{p,q=0}^{n-1} (pp|qq) - \frac{1}{2} \sum_{k=0}^{n-1} g_k^{\varnothing} \hat{U}^{\varnothing\dagger} (\hat{Z}_k + \hat{Z}_{k+1}) \hat{U}^{\varnothing} \\ &+ \frac{1}{4} \sum_{t=0}^{T} \sum_{k<l}^{n-1} Z_{kl}^t \hat{U}^{t\dagger} \left(\hat{Z}_k \hat{Z}_l \right) \hat{U}^t. \end{aligned} \tag{48}$$

To wrap up, and by noting $\mathcal{E} \coloneqq E_0 + \sum_k (p|\hat{h}|p) - \frac{1}{4} \sum_{pq} (pq|pq) + \frac{1}{2} \sum_{p,q} (pp|qq)$, the XDF expression of the molecular Hamiltonian is

$$\hat{\mathcal{H}} = \mathcal{E} - \frac{1}{2} \hat{U}^{\varnothing\dagger} \left(\sum_{k=0}^{2n-1} g_{\lfloor k/2 \rfloor}^{\varnothing} \hat{Z}_k \right) \hat{U}^{\varnothing} + \frac{1}{4} \sum_{t=0}^{T} \hat{U}^{t\dagger} \left(\sum_{k<l=0}^{2n-1} Z_{\lfloor k/2 \rfloor \lfloor l/2 \rfloor}^t \hat{Z}_k \hat{Z}_l \right) \hat{U}^t. \tag{49}$$

The compressed double-factorization (CDF) [16] has a similar form, however, the orbital rotations U^t and the prefactors $g^{\varnothing}, Z^t$, are obtained using a non-linear optimization of a cost function measuring the distance between the double-factorized form $(pq|rs)$ and the true $(pq|rs)$ tensor. The number of t-leafs in the CDF factorization can be significantly reduced compared to XDF. As a result, in the context of VQE, the number of distinct terms to be measured is similarly reduced.

However, upon further investigations that we have conducted and reported in FIG. 2 and FIG. 3 of 1.4, we discovered that the variance of the CDF energy estimator grows very significantly. Hence, the corresponding shot budget M for a given accuracy ϵ is very costly

$$M = v/\epsilon^2. \tag{50}$$

We specifically address this issue, in our RC-DF method. The RC-DF produces an efficient measurement procedure by compressing the t-leafs and simultaneously minimizing the Eucledean norm of the tensors Z^t. We achieve this by optimizing the 2-norm of the difference between the $(pq|rs)$ and the DF representation with parameters U^t, Z^t and adding a regularization term on the Euclidean norm of Z^t.

1.4 Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double-Factorization

Bibliographic Information

Oumarou, O., Scheurer, M., Parrish, R. M., Hohenstein, E. G., & Gogolin, C. (2024). Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization. Quantum, 8, 1371. CLOCKSS. https://doi.org/10.22331/q-2024-06-13-1371

Summary

Double-factorization emerged as a very effective linear combination of unitaries (LCU) representation of the molecular Hamiltonian that, via exploiting the 8-fold symmetry of the electron repulsion integrals, replaces the four-index sum of fermionic operators by a three-index sum of diagonal fermionic operators in a suitable basis. In this work, we devise a new flavor of double-factorized LCU expansion of molecular Hamiltonians, called regularized compressed double-factorization. We obtain the double-factorization representation parameters using a nested optimization loop that minimizes simultaneously the distance between the original Hamiltonian and the double-factorized one, as well as the one norm of the prefactors of the double-factorization LCU λ by introducing a regularization term in the cost function. The optimization is constrained to fulfill two main criteria: (1)low λ values and (2)compressing the number of the non-commuting terms or basis rotations to a minimum. This benefits both NISQ algorithms such as VQE in that it reduces the number of basis rotations and the total shot budget for an energy evaluation compared to various other double-factorization and Pauli-grouping techniques. Furthermore, the λ factor is an essential quantity for the QPE algorithm. Therefore, by minimizing such quantity, we reduced the depth and total runtime of QPE. For the benchmark results, we consider the molecular Hamiltonians of p-benzyne and naphtalene in 12 spin orbitals and 20 spin orbitals (qubits) respectively. We then evaluate the variance and subsequently the number of shots for a given ground state energy (difference) accuracy. We found that in the p-benzyne case our method produces the lowest shot budget to reach chemical accuracy energy level compared to various other double-factorized and Pauli-grouping methods. The number of shots was 3 to 6 times lower. Moreover, RC-DF yielded, at least, $3\times$ lower number of basis rotations. For the napthalene case, the results were as pronounced where RC-DF was the only method that reached chemical accuracy at already six t-leafs (basis rotations) and a shot budget of 2×10^6. We also benchmarked RC-DF on the p-450 molecular Hamiltonian and found that we produce $2\times$ smaller λ factors compared to THC.

Contribution

O.O. designed the cost function, derived its analytic gradient, determined the expression of the optimal Z tensor, formulated and proved the lemmas presented in the appendices, implemented and tested the optimization procedure, the required circuits for implementing the rotations U^t using Pennylane, generated the data for two of the benchmark cases and co-wrote the manuscript.

Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization

Oumarou Oumarou[1], Maximilian Scheurer[1], Robert M. Parrish[2], Edward G. Hohenstein[2], and Christian Gogolin[1]

[1] Covestro Deutschland AG, Leverkusen 51373, Germany
[2] QC Ware Corporation, Palo Alto, CA 94301, USA
06/04/2024

We propose the regularized compressed double factorization (RC-DF) method to classically compute compressed representations of molecular Hamiltonians that enable efficient simulation with noisy intermediate scale (NISQ) and error corrected quantum algorithms. We find that already for small systems with 12 to 20 qubits, the resulting NISQ measurement scheme reduces the number of measurement bases by roughly a factor of three and the shot count to reach chemical accuracy by a factor of three to six compared to truncated double factorization (DF) and we see order of magnitude improvements over Pauli grouping schemes. We demonstrate the scalability of our approach by performing RC-DF on the Cpd I species of cytochrome P450 with 58 orbitals and find that using the resulting compressed Hamiltonian cuts the run time of qubitization and truncated DF based error corrected algorithms almost in half and even outperforms the lambda parameters achievable with tensor hypercontraction (THC) while at the same time reducing the CCSD(T) energy error heuristic by an order of magnitude.

The run time of algorithms on most of today's noisy intermediate scale (NISQ) hardware platforms is largely independent of the anyway shallow circuit depth. Instead, it is mostly a function of the number of distinct circuits that need to be evaluated and the number of repeated circuit executions because re-programming the quantum device to perform a different circuit and the time for measurement and qubit reset dominate the overall circuit execution time [1].

Robert M. Parrish: rob.parrish@qcware.com
Christian Gogolin: christian.gogolin@covestro.com

Many NISQ algorithms such as the variational quantum eigensolver (VQE) [2] are essentially methods to reduce circuit depth at the expense of requiring many repetitions, also called shots. In a similar fashion, error mitigation techniques [3, 4] create a tolerance for the noise of NISQ devices by further increasing the number of shots needed to obtain a final result. The total number of shots is thus often the limiting factor on the path towards quantum advantage.

This is a particularly pressing issue in in quantum chemistry simulation. Here, molecular Hamiltonians, which in their second-quantized form have $O(n^4)$ terms, where n is the number of spatial orbitals, need to be measured with very high accuracy. Naive measurement schemes require an extremely fast growing number of distinct observables and total number of shots [5] for reaching the required accuracy.

Methods to cope with this problem fall in two broad classes. First, methods [6, 7, 8] which, starting from a decomposition of the observable into Pauli operators, group or otherwise combine these Pauli operators into sets that are jointly measurable with no or only minimal increase in circuit depth. Second, methods which yield a compressed and possibly approximate representation of the original Hamiltonian in the form of a tensor contraction. For fermionic second-quantized Hamiltonians, these are mainly density fitting [9], tensor hypercontraction (THC) [10, 11, 12], and double factorization (DF) [13, 14] (see [14] for a comparison). While the Pauli grouping methods are applicable to general qubit Hamiltonians, the second class of methods typically yields better performance when applicable [5].

These compressed representations also enable drastic resource reductions in leading fault tolerant algorithms for the simulation of chemistry based on linear combinations of unitaries (LCU) and qubitization [15, 12, 16, 17, 11]. Here run time is mainly a function of the so-called lambda parameter. Its precise definition depends on the algorithm and will be discussed later, but it can be thought of as a norm-like quantity that depends on the magnitude of the coefficients of the representation of the Hamiltonian. THC typically yields lower lambda parameters than existing DF schemes. The fact that some tensors in the THC decomposition are non-square and non-unitary causes other overheads and complications [17] which does not make THC a viable option for typical NISQ quantum algorithms.

In contrast, explicit double factorization (X-DF) and compressed double factorization (C-DF) [18, 19, 20, 21, 22] naturally yield a NISQ-friendly measurement scheme that only requires a linear depth orbital/Givens rotation circuit before the final measurements, is compatible with particle number post selection, and also an LCU representation of the Hamiltonian suitable for error corrected algorithms based on qubitization [23].

The X-DF measurement scheme reduces the number of distinct measurement bases to at most $n(n+1)/2$ and drastically decreases the number of shots to reach a target accuracy, when compared to Pauli-based schemes. The number of bases can be further reduced by truncating the X-DF representation of the Hamiltonian, thereby making the representation approximate. This can reduce the required number of shots, but the error resulting from the now approximate representation of the Hamiltonian quickly outweighs this. C-DF is designed to overcome this issue by performing a tighter least-squares numerical tensor fitting of the molecular Hamiltonian to truncated double-factorized form. By lifting a rank constraint in the equation defining the X-DF Hamiltonian and using the resulting additional freedom to improve the representation of the molecular Hamiltonian by means of parameter optimization starting from a truncated X-DF guess, it achieves lower approximation errors than truncated X-DF. However, when attempting practical deployment of C-DF in the context of quantum algorithms, one encounters an additional major barrier: the optimization of the C-DF tensor fitting to minimize least squares error does not consider the variance properties of the resulting representation. In practice, this means that the variance of the resulting energy estimator can erratically fluctuate and can be orders of magnitude higher than the variance of the X-DF energy estimator and the approximation error of both X-DF and C-DF.

In this work we propose the regularized compressed double factorization method (RC-DF) to fix this. RC-DF uses the same functional form of the compressed Hamiltonian as C-DF but it adds a regularization term to the C-DF cost function that is used when optimizing the parameters of the compressed representation [1]. The regularization term stabilizes the optimization and reduces the variance of the resulting NISQ energy estimator as well as the λ parameter determining the resources of fault tolerant quantum algorithms. We find that RC-DF consistently outperforms both previous double factorization schemes in terms of variance, approximation error, and lambda parameter and even yields lambda parameters lower than THC.

1 Comparison of factorization methods

We start from the well known form of the second-quantized electronic structure Hamiltonian

$$\hat{H} = E_c + \sum_{pq}(p|\hat{h}_c|q)\hat{E}_{pq} + \frac{1}{2}\sum_{pqrs}(pq|rs)\left(\hat{E}_{pq}\hat{E}_{rs} - \delta_{qr}\hat{E}_{ps}\right), \quad (1)$$

where

$$(p|\hat{h}_c|q) = \int \phi_p^*(r)(-\frac{1}{2}\nabla^2(r) - \sum_m \frac{Z_m}{r - r_m})\phi_q(r)dr$$

$$(pq|rs) = \iint \phi_p^*(r_1)\phi_q(r_2)\frac{1}{r_{12}}\phi(r_1)_r^*\phi(r_2)_s dr_1 dr_2$$

are the symmetric one-electron integrals and the real and 8-fold symmetric two-electron integrals with Z_m and r_m the charges and positions of the

[1] When working out the implications of RC-DF for fault tolerant quantum algorithms, we became aware that a similar erratic behavior of the λ parameter of THC had been observed in [11] and an L1 regularization has been proposed as a cure there.

nuclei and ϕ the spacial molecular orbitals, and $\hat{E}_{pq} \coloneqq \hat{p}^\dagger \hat{q} + \hat{\bar{p}}^\dagger \hat{\bar{q}}$ is the singlet excitation operator. The exact X-DF representation of the Hamiltonian is determined by diagonalizing the modified one-electron integrals tensor $\mathcal{F}_{pq}$ and doubly diagonalizing the two-electron integrals tensor to obtain

$$\mathcal{F}_{pq} \coloneqq (p|\hat{h}_c|q) - \frac{1}{2}\sum_r (pr|qr) + \sum_r (pq|rr) \quad (2)$$

$$= \sum_k U^{\varnothing}_{pk}\, \mathcal{F}^{\varnothing}_k\, U^{\varnothing}_{qk} \quad (3)$$

and

$$(pq|rs) = \sum_{t=1}^{n_t} V^t_{pq}\, g_t\, V^t_{rs} \quad (4)$$

$$= \sum_{t=1}^{n_t}\sum_{kl=1}^{n} U^t_{pk}\, U^t_{qk}\, Z^t_{kl}\, U^t_{rl}\, U^t_{sl}, \quad (5)$$

where the U^t_{pk} result from diagonalizing the $V^t_{pq} = \sum_k^n U^t_{pk}\Lambda_k U^t_{qk}$ and consequently $Z^t_{kl} = \Lambda_k g^t \Lambda_l$ is, for every t, a symmetric outer product, hence of rank one, and the U^t_{pk} are unitary (in fact without loss of generality special orthogonal). The second factorization is possible whenever $(pq|rs)$ is real and 8-fold symmetric (as is always the case for non-relativistic Coulomb repulsion integrals), as this is enough to ensure that the V^t_{pq} are not only orthogonal but also real and symmetric for every t (see Lemma 1 in Appendix D). With n_t equal to the maximum number $n(n+1)/2$ of non-zero eigenvalues of $(pq|rs)$ the Hamiltonian can then be written exactly (see Appendix F for the full derivation) as

$$\begin{aligned}\hat{H} = \mathcal{E} &- \frac{1}{2}\sum_k \mathcal{F}^{\varnothing}_k U^{\varnothing}\left(\hat{Z}_k + \hat{Z}_{\bar{k}}\right)(U^{\varnothing})^\dagger \\ &+ \frac{1}{8}\sum_{t=1}^{n_t}\sum_{kl=1}^{n} Z^t_{kl} U^t \left(\hat{Z}_k\hat{Z}_l - \delta_{kl} + \hat{Z}_k\hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}}\hat{Z}_l + \hat{Z}_{\bar{k}}\hat{Z}_{\bar{l}} - \delta_{\bar{k}\bar{l}}\right)(U^t)^\dagger,\end{aligned} \quad (6)$$

where

$$\mathcal{E} = E_c + \sum_p (p|\hat{h}_c|p) + \frac{1}{2}\sum_{pq}(pp|qq) - \frac{1}{4}\sum_{pq}(pq|pq), \quad (7)$$

is independent of the state, and $U^{\varnothing}$ and U^t rotate the orbitals for each t according to U^t_{pk} (See Fig. 1), $\hat{Z}_k, \hat{Z}_{\bar{k}}$ are respectively pauli operators on qubit $2k$ and $2k+1$, $k \in [0, n-1]$.

If the sum over t is ordered according to $|g_t|$, the Hamiltonian can be approximated with a truncated X-DF representation with fewer terms (also called leafs).

In (R)C-DF the rank-one constraint on the Z^t_{kl} is lifted and they are allowed to be arbitrary symmetric matrices. The orbital rotations U^t_{pk} and coefficients Z^t_{kl} are then obtained using a two-step gradient based optimization procedure by first exponentially parametrizing the orbital rotations $U^t_{pq} \coloneqq \exp(X^t)_{pq}$ via anti-symmetric generators X^t_{pq} and then minimizing the squared Frobenius norm (in [16] this is called the incoherent error)

$$\frac{1}{2}\Big\| \underbrace{\sum_{t=1}^{n_t}\sum_{kl=1}^{n} U^t_{pk}\, U^t_{qk}\, Z^t_{kl}\, U^t_{rl}\, U^t_{sl} - (pq|rs)}_{\Delta_{pqrs}} \Big\|^2_{\mathcal{F}} \quad (8)$$

of the difference between the left and right hand side of (5) for some pre-set $n_t \leq n\,(n+1)/2$ starting from a truncated X-DF initial guess (for details see [18]).

Irrespective of whether a Hamiltonian representation of the form (6) was found via X-DF or (R)C-DF, the energy can be then be measured by means of quantum circuit with a linear gate depth overhead of the form shown in Fig. 1.

The most favorable resource estimates of quantum algorithms for error corrected simulation of chemistry with qubitization have been obtained with THC [17, 11]. In this context THC approximates the two-body part of the Hamiltonian ac-

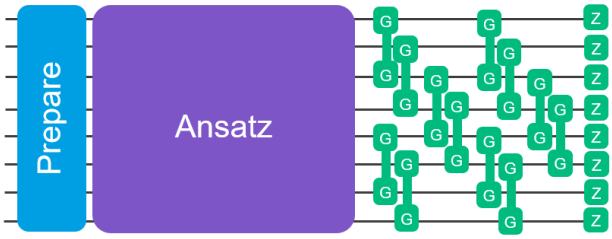


Figure 1: DF measurement scheme according to the LCU decomposition in (6). From each $U^{\varnothing}$ and U^t the parameters of a square shaped fabric of givens gates G can be computed. The results of $\hat{Z}$ and $\hat{Z}\otimes\hat{Z}$ measurements in these n_t+1 distinct bases can then be contracted against the $\mathcal{F}_k^{\varnothing}$ and Z_{kl}^t tensors to obtain an energy estimator.

cording to

$$(pq|rs) \approx \sum_{kl=1}^{M} \chi_p^k \chi_q^k \zeta_{kl} \chi_r^l \chi_s^l, \tag{9}$$

with $\sum_k^M \chi_p^k \chi_q^k \leq 1$ and $M \leq n^2$ the THC rank. Also here the symmetric ζ_{kl} and rectangular χ_p^k are found by means of minimizing the Frobenius norm error. When DF is used with qubitization [12, 16] it is usually presented and performed according to

$$\begin{aligned}(pq|rs) &= \sum_{t=1}^{n^2} L_{pq}^t L_{rs}^t \qquad (10)\\ &= \frac{1}{2}\sum_{t=1}^{n^2} U^t \Big(\sum_k^n \lambda_k^t \,(\hat{Z}_k + \hat{Z}_{\bar{k}})\Big)^2 (U^t)^\dagger, \qquad (11)\end{aligned}$$

where the L_{pq}^t can be found with Cholesky decomposition or eigen decomposition as $L_{pq}^t = \sqrt{g_t} V_{pq}^t$ and the scalar $\lambda_k^t = \sqrt{g_t}\Lambda_k$ are the eigenvectors and the U^t the diagonalizing unitaries of the L_{pq}^t (here g_t, V_{pq}^t, and Λ_k refer to the quantities introduced in the context of X-DF above). For a NISQ measurement scheme such as that in Fig. 1 it makes no sense to partially discard a leaf (because the measurement data is available for all contributions from a leaf), but in qubitization truncation can be done on the level of setting individual λ_k^t equal to zero.

The (R)C-DF form of the Hamiltonian, in which Z_{kl}^t is no longer rank one, can be re-cast as a sum of operator squares similar to (10). By taking the matrix square root $W_{kl}^t := (\sqrt{Z^t})_{kl}$ so that $Z_{kl}^t = \sum_i^n W_{ki}^t W_{li}^t$ (we use the implementation in scipy of the algorithm from [24], which works also for non-positive Z_{kl}^t matrices, in which case the W_{ki}^t come out complex, which is compatible with the scheme from [16]) one can write

$$\begin{aligned}(pq|rs) = \sum_t^{n_t}\sum_i^n \Big(\sum_k U_{pk}^t\, U_{qk}^t\, W_{ki}^t\Big) \\ \times \Big(\sum_l W_{il}^t\, U_{rl}^t\, U_{sl}^t\Big).\end{aligned} \tag{12}$$

This allows one to run the algorithm of [16] with (R)C-DF Hamiltonians as input.

The block encoding of the qubitization method needs the Hamiltonian in the form of an LCU. The number of ancillary qubits and T gates is then determined by the number of terms of the LCU and a sort of normalization factor, called the lambda parameter [16, 17]. The result of (truncated) X-DF, C-DF, and RC-DF is itself an LCU with a lambda factor

$$\lambda_{\mathrm{DF}}^{\mathrm{LCU}} := \sum_k^n |\mathcal{F}_k^{\varnothing}| + \sum_{t=1}^{n_t}\Big(\sum_{k<l}^n |Z_{kl}^t| + \frac{1}{4}\sum_k^n |Z_{kk}^t|\Big). \tag{13}$$

Alternatively, because of (12), one can use the algorithm from [16], which achieves a contribution from the two-body part of the Hamiltonian to lambda of $1/4\,\sum_t \|(L_{pq}^t)_{pq}\|_1^2$ for Hamiltonians of the form (10) (where $\|\cdot\|_1$ is the Schatten 1-norm) and using (12) one can thus obtain

$$\lambda_{\mathrm{DF}}^{\mathrm{Burg}} := \sum_k^n |\mathcal{F}_k^{\varnothing}| + \frac{1}{4}\sum_t^{n_t}\sum_i^n \Big\|(\sum_k^n U_{pk}^t U_{qk}^t W_{ki}^t)_{pq}\Big\|_1^2 \tag{14}$$

$$= \sum_k^n |\mathcal{F}_k^{\varnothing}| + \frac{1}{4}\sum_t^{n_t}\sum_i^n \Big(\sum_k^n |W_{ki}^t|\Big)^2. \tag{15}$$

The difference between $\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$ and $\lambda_{\mathrm{DF}}^{\mathrm{LCU}}$ stems from the different LCU representation of the Hamiltonian. The latter uses the equation in (6) which is evidently an LCU and subsequently its $||.||_1$ is (13). The former however uses the algorithm and LCU from (13) of [16]. Finallyl, for THC, Lee et al. [17] have obtained a lambda of

$$\lambda_{\mathrm{THC}}^{\mathrm{Lee}} := \sum_k^n |\mathcal{F}_k^{\varnothing}| + \frac{1}{2}\sum_{kl}^M |\zeta_{kl}|. \tag{16}$$

The precise run times of the algorithms corresponding to the different lambda values differ

and depend on factorization-specific quantities such as the THC rank M but they all scale like their respective lambda divided by the allowable phase estimation energy error times the sum of run times of certain circuit primitives plus logarithmic overheads. The differences between the lambda values have turned out to outweigh the influence of other factors when comparing algorithm run times for similar overall target accuracies [17].

2 Regularized compressed double factorization

While C-DF allows to reduce the number of leafs needed for good accuracy from close to n^2 for X-DF to roughly linear in n while maintaining an approximate but sufficiently accurate representation of the Hamiltonian, it turns out that the optimization of C-DF often converges to Z^t_{kl} tensors with very large entries. This is problematic since the variance of the NISQ energy estimator and both lambda parameters grow with the number and magnitude of $|Z^t_{kl}|$ values. To solve this issue, we propose to add to the C-DF cost function from (8) a regularization term penalizing large $|Z^t_{kl}|$ via a tensor of weights $\rho_{tkl} \geq 0$

$$\frac{1}{2}\|\Delta_{pqrs}\|^2_{\mathcal{F}} + \sum_{tkl} \rho_{tkl}|Z^t_{kl}|^2, \tag{17}$$

We have tested both weighted L1 and L2 regularizations of the form $(\sum_{tkl} \rho_{tkl}|Z^t_{kl}|^\gamma)^{\frac{2}{\gamma}}$ with $\gamma \in 1, 2$ but concentrate on L2 regularization with uniform regularization strength $\rho_{tkl} = \rho$ in the rest of the main text.

As in C-DF, a joint optimization of the U^t_{pq} and Z^t_{pq} has unfavorable performance also with regularization, but the the two-step optimization of C-DF proposed in [18] can be adopted to the regularized case. Further, for large n, a very expensive 6-index matrix inversion can be circumvented by carrying it out in a matrix-free manner with, e.g., a conjugate gradient algorithm (for details see Appendix A). We have found that this step benefits from the regularization (L2), as it improves the conditioning of the matrix. In RC-DF, initialization can be done either from X-DF truncated to the target number of C-DF leafs or one can start from the full X-DF factorization and put a high penalty on the leafs that are to be truncated in the end. In practice, contrary to the difficulties reported in [11, 17] on converging THC, RC-DF seems to be rather well behaved. Convergence may take thousands of iterations, but we had no difficulty converging RC-DF in large active spaces to much tighter residual Frobenius norm errors (8) and coupled cluster with singles, doubles, and perturbative triples (CCSD(T)) energy errors than those reported for THC [11] (see Appendix A for further details of the optimization procedure).

3 Numerical results

Unless otherwise explicitly stated all results shown in the following were obtained with L2 regularization starting from a truncated X-DF guess and with a uniform regularization factor for all n_t leafs of $\rho_{tkl} =: \rho$.

We first investigate the advantages of RC-DF over other NISQ measurement schemes. The performance of any such scheme is determined by both the systematic error introduced in case the Hamiltonian is approximated and the variance, Var, of the estimator. We quantify the overall performance by the mean squared error and take the square root to obtain a quantity that has units of energy $\sqrt{\text{MSE}} := \sqrt{\langle \hat{H} - \hat{H}' \rangle^2 + \text{Var}}$, where $\hat{H}'$ is the compressed Hamiltonian obtained with the respective flavor of double factorization and for the Pauli grouping based schemes $\hat{H}' = \hat{H}$.

The variance further depends on the state, the overall shot budget, and the shot distribution. In the main text we show data for the state being the complete active space configuration interaction (CASCI) ground state, but the plots look very similar for representative states along a VQE optimization trajectory. We consider two shot distribution schemes: A "uniform" distribution which divides the total number of shots uniformly among all bases in which measurements need to be preformed, and an "according to weights" distribution which distributes the shots according to the L2 norm of the coefficients of each group of jointly measurable Pauli operators. We chose the overall shot budget so that the best method is able to achieve chemical accuracy of 10^{-3} Hartree.

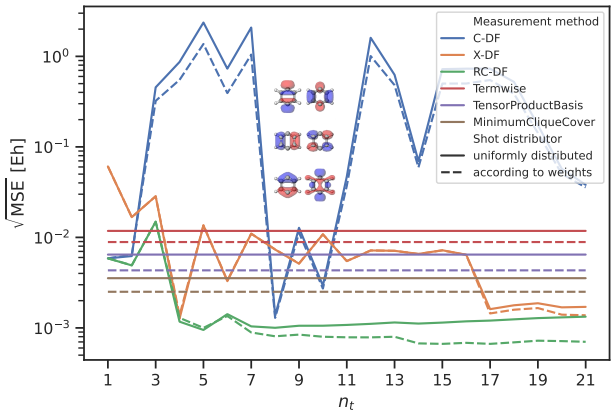


Figure 2: Performance of measurement schemes based on C-DF, X-DF, and RC-DF (with L2 regularization $\rho = 10^{-6}$) in comparison with the naive termwise Pauli scheme as well as the tensor product basis [28] and a minimum clique cover [6] based Pauli grouping methods as implemented in [29] for 3×10^5 shots each in computing a single point energy in the (6e, 6o) CASCI ground state of para-benzyne on 12 qubits. To reach a MSE on par with that of RC-DF at $n_t = 7$ would require over 1.1×10^6 shots with X-DF and $n_t = 17$ and about 3.8×10^6 shots with the best Pauli grouping scheme. We also compare shot distribution schemes (see main text) and find that "according to weights" improves performance.

3.1 NISQ measurement of the para-benzyne ground state

As a first test case we consider the CASCI ground state in a (6e, 6o) active space of FON-HF/cc-pVDZ (with fractional occupations determined with temperature 0.1[1/Eh] and Gaussian broadening within the active space) [25, 26, 27] orbitals of para-benzyne (see Fig. 2). The weighted shot distribution yields lower variance than uniform shot distribution in all cases by directing more shots to the more important first leafs. RC-DF beats the second best method (X-DF) by approximately a factor of five and while the maximum number $21 = 6\,(6+1)/2$ of leafs yields the lowest MSE, chemical accuracy is consistently achievable with RC-DF from $n_t = 7$ on. The MSE of C-DF fluctuates widely and while it by chance achieves a good variance and approximation error for 8 leafs, this is of limited use for practical applications. The results are virtually unchanged over a broad range of ρ values and good values for ρ can be found in a systematic way even when running on quantum hardware (see Appendix B).

3.2 NISQ measurement of the singlet-triplet gap of naphthalene

As a second test case, we investigate the singlet-triplet energy gap in a (10e, 10o) active space consisting of the π system of naphthalene constructed with AVAS [30] as implemented in PySCF [31, 32]. The reference state was computed at the HF/def2-SVP[33] level of theory. We only show data for double factorization based measurement schemes since the Pauli grouping based schemes become increasingly less competitive for larger systems. Since the "according to weights" method is significantly better than uniform shot distribution, we use it exclusively in this case. We compute the factorization once per n_t and then evaluate the energetically lowest singlet and triplet energies from the same decomposition. The MSE of the singlet-triplet energy gap is given by the square of the difference between the noiseless energy gaps Δ computed with the exact and the gap Δ' from the compressed Hamiltonian plus the sum of the variances of the singlet $\mathrm{Var_S}$ and triplet $\mathrm{Var_T}$ energies $\mathrm{MSE} = (\Delta - \Delta')^2 + \mathrm{Var_S} + \mathrm{Var_T}$. We find the variances $\mathrm{Var_S}$ and $\mathrm{Var_T}$ of both states to be very similar, and for (R)C-DF, the MSE is dominated by the variance contributions for $n_t \geq 8$ whereas for X-DF, the systematic energy error remains high until $n_t \approx 25$. Surprisingly, RC-DF reaches chemical accuracy already at $n_t = 6$ with the assigned shot budget and is consistently more accurate than the other two factorization methods for all $n_t > 2$. By contrast, the variance of C-DF singlet-triplet gap estimations is quite erratic, as explained previously. While the accuracy in predicting the energy gap with X-DF is well controllable, an approximately 12 times larger shot budget and more leafs would be needed to reach chemical accuracy. This test case shows the reliable performance of RC-DF for medium-sized active spaces and indicates that the method could be employed for other chemical properties such as activation energies.

3.3 Combination with fluid fermionic fragments

We further explore how X-DF and RC-DF can be combined with the fluid fermionic fragments (FFF) [21] technique to further reduce shot budgets. The FFF technique exploits the fact

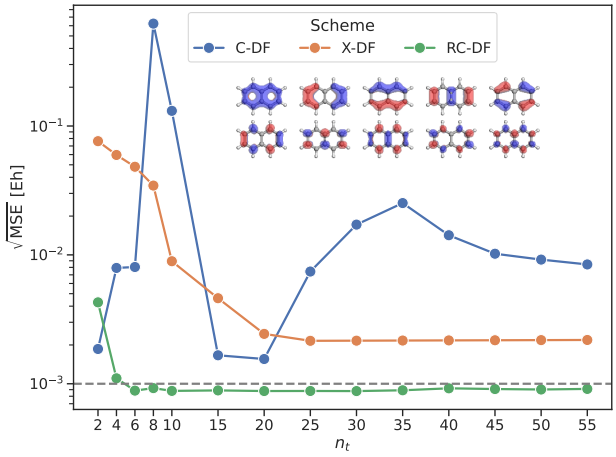


Figure 3: Square-root of the MSE for the singlet-triplet energy gap of naphthalene on 20 qubits (i.e., (10e, 10o) active space, shown in the inset). For each factorization method, 2×10^6 shots were used in total for the singlet and triplet state, distributed according to weights. A regularization factor of $\rho = 10^{-3}$ was used for RC-DF. RC-DF reaches chemical accuracy with only $n_t = 6$ leafs. X-DF would require at least $n_t = 25$ leafs and at least 12×10^6 shots to achieve the same accuracy as RC-DF with only $n_t = 6$ leafs.

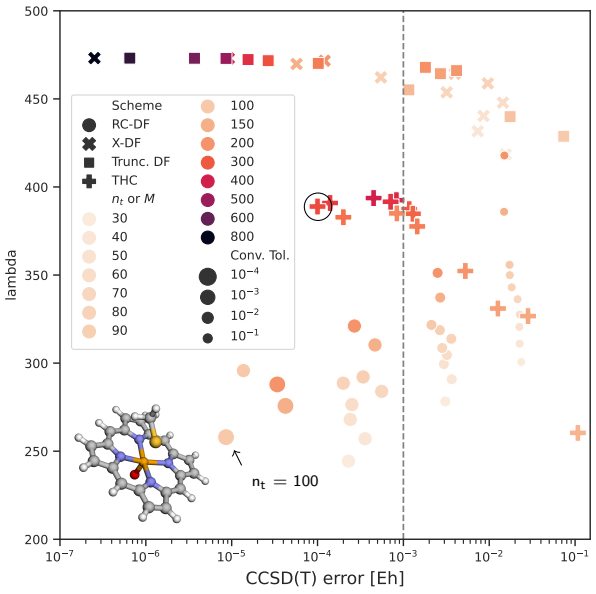


Figure 4: Comparison of the achievable CCSD(T) error heuristic and lambda values $\lambda_{\text{DF}}^{\text{Burg}}$ for the truncated DF method based on (10), X-DF and RC-DF, as well as $\lambda_{\text{THC}}^{\text{Lee}}$ for THC. The color scheme represents the number of leafs n_t for double factorization schemes, or the THC rank M. The active space Hamiltonian of the Cpd I model of cytochrome P450 and the data for THC and truncated DF were taken from [11]. The encircled THC data point was used for the resource estimates there. To compare different levels of convergence we vary the squared Frobenius norm error (8) at which we abort the RC-DF optimization (Conv. Tol.) and use $\rho = 10^{-3}$. The data is tabulated in Appendix E.

that certain quadratic terms of the Hamiltonian, called fluid fermionic fragments, can be taken care of in different parts of the energy estimator. FFF minimizes the variance by optimizing how these terms are spread over the different possible locations (for more details see Appendix G). The variance can thereby either be approximated with that of a mock state whose variance can be classically efficiently computed or it can for example be estimated with part of the shot budget or from a classical shadow [34, 35], which can then also be used for the energy estimation. In any case, the final form of the FFF optimized energy estimator is then state dependent and optimized to have low variance for certain states. We find (see Figure 8 in Appendix G) that for the cases considered RC-DF and FFF nicely complement each other. Using RC-DF as initial point for FFF yields the lowest shot budgets and that the FFF optimization converges faster when started form RC-DF than from X-DF and that the state independent distribution of the fluid fermionic fragments corresponding to the Hamiltonian as written in (6) is a good initial guess for the FFF coefficients.

3.4 RC-DF for error corrected quantum computing

We now turn to exploring the usefulness of RC-DF for error corrected algorithms based on qubitization. As example we take the $(34\alpha{+}29\beta\text{e}, 58\text{o})$ active space of the Cpd I species of cytochrome P450 proposed in [11]. As this system is beyond the regime accessible with CASCI, we use, as in [17, 11], the CCSD(T) energy error as a heuristic to assess the quality of the compressed representation. Computational details can be found in Appendix E.

The aim is then to find compressed representation of the Hamiltonian that achieves both a low CCSD(T) error and a low lambda value. We find $\lambda_{\text{DF}}^{\text{Burg}} < \lambda_{\text{DF}}^{\text{LCU}}$ in all cases and thus only display and compare the former. As can be seen in Fig. 4, RC-DF outperforms both previous DF methods and THC by a substantial amount. Compared to

truncated DF or X-DF we can almost cut $\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$ in half, and thereby also the run time of the quantum computer, while at the same time achieving a CCSD(T) error smaller than 5×10^{-5}, which was not reached with THC [11]. Further comparing with THC, we can reduce the CCSD(T) error by an order of magnitude and the lambda value by roughly one third or, alternatively, RC-DF can achieve a $\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$ that is only 60% of λ_{THC} at comparable CCSD(T) error. This improvement becomes even more noteworthy when recognizing that when qubitizing the RC-DF Hamiltonian one need not worry about the complications caused by the non-orthogonal nature of THC that had to be worked around in [17] (see Section III of [17]). Converging the best RC-DF data point with $n_t = 100$ took 2944 L-BFGS iterations and approximately 11 hours on a single GPU and our not highly optimized JAX [36] based implementation.

To investigate the scaling of the achievable lambda values with system size we considered the hydrogen chain benchmark in the STO-6G basis previously proposed in [12, 17]. Also there we find that at $n = 100$ orbitals (corresponding to 100 hydrogen atoms) RC-DF achieves values of $\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$ that are lower than previously reported values for λ_{THC} and we find, as for THC, an approximately linear scaling of $\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$ with n if n_t is chosen such that the CCSD(T) error per particle is constant (see Appendix C for details).

It should be clear that the lambda values, while highly significant for determining the quantum resources of fault tolerant algorithms, are not the only relevant quantity and we do not claim that our analysis constitutes a comprehensive comparison of the quantum resources required for different methods. We also leave to future work the possibility of regularizing methods such as RC-DF or THC with quantities more directly related to the required quantum resources than the norm-like term in (17), which could more directly steer the optimization towards factorizations with low qubit or T gate count, as desired.

4 Conclusions

We proposed the regularized compressed double factorization (RC-DF) method which, from a unified framework yields both a NISQ compatible measurement scheme with only linear circuit overhead and can be used in conjunction with qubitization in error corrected quantum algorithms for the simulation of chemistry. We found that in both of these scenarios RC-DF leads to lower quantum run times when compared to previous double factorization (DF) and tensor hypercontraction (THC) schemes. Contrary to THC, the Hamiltonian in DF form can also be used to construct trotter schemes that need to alternate between a very small number of non-commuting operators. It will be interesting to compare the quantum resources (e.g. number of Toffoli gates) required for phase estimation based on THC and RC-DF via qubitization with RC-DF and trotterization.

In the NISQ setting, this advantage is a consequence of the fact that the regularization guides the optimization towards compressed representations of the Hamiltonian with smaller coefficients, which reduces the variance of the resulting energy estimator. In qubitization schemes, the smaller coefficients reduce the norm-like lambda parameter of the Hamiltonian on which the T gate count depends in a multiplicative fashion.

Avoiding a six-index intermediate quantity during the RC-DF optimization and adopting a two step gradient based scheme previously developed by some of us for non-regularized compressed DF, we were able to make RC-DF scale well into the regime where quantum computers may provide an advantage over classical methods. More work is needed to understand the precise scaling of RC-DF with active space and basis set size and to explore other options for regularization.

Data Availability

The data supporting the findings of this manuscript have been uploaded to Zenodo with the DOI `10.5281/zenodo.7866658` [37], including instructions on how to load the data using Python.

Acknowledgements

We thank Gian-Luca Anselmetti, Fotios Gkritsis, and Pauline Ollitrault, Artur Izmaylov and Seonghoon Choi for valuable discussions. RMP thanks Mario Motta and Jeff Cohen for many

discussions regarding C-DF. QC Ware Corp. acknowledges generous funding from Covestro for the undertaking of 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).

Conflict of Interest: EGH and RMP own stock/options in QC Ware Corp.

References

[1] 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" (2022).

[2] 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**, 1–7 (2014).

[3] 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**, 020317 (2021).

[4] 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" (2022).

[5] 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 Quantum Information **7**, 1–9 (2021).

[6] 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**, 124114 (2020).

[7] Laurin E Fischer, Daniel Miller, Francesco Tacchino, Panagiotis Kl Barkoutsos, Daniel J Egger, and Ivano Tavernelli. "Ancilla-free implementation of generalized

measurements for qubits embedded in a qudit space". Physical Review Research **4**, 033027 (2022).

[8] Daniel Miller, Laurin E. Fischer, Igor O. Sokolov, Panagiotis Kl. Barkoutsos, and Ivano Tavernelli. "Hardware-tailored diagonalization circuits" (2022). arXiv:2203.03646.

[9] J. L. Whitten. "Coulombic potential energy integrals and approximations". The Journal of Chemical Physics **58**, 4496–4501 (1973).

[10] 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**, 044103 (2012).

[11] Joshua J. Goings, Alec White, Joonho Lee, Christofer S. Tautermann, Matthias Degroote, Craig Gidney, Toru Shiozaki, Ryan Babbush, and Nicholas C. Rubin. "Reliably assessing the electronic structure of cytochrome p450 on today's classical computers and tomorrow's quantum computers". Proceedings of the National Academy of Sciences **119**, e2203533119 (2022).

[12] 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 (2019).

[13] Edward G. Hohenstein, Oumarou Oumarou, Rachael Al-Saadon, Gian-Luca R. Anselmetti, Maximilian Scheurer, Christian Gogolin, and Robert M. Parrish. "Efficient quantum analytic nuclear gradients with double factorization". The Journal of Chemical Physics **158**, 114119 (2023).

[14] Jeffrey Cohn, Mario Motta, and Robert M. Parrish. "Quantum Filter Diagonalization with Compressed Double-Factorized Hamiltonians". PRX Quantum **2**, 040352 (2021).

[15] Ryan Babbush, Craig Gidney, Dominic W. Berry, Nathan Wiebe, Jarrod McClean, Alexandru Paler, Austin Fowler, and Hartmut Neven. "Encoding electronic spectra in quantum circuits with linear t complexity". Phys. Rev. X **8**, 041015 (2018).

[16] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. "Quantum computing enhanced computational catalysis". Phys. Rev. Res. **3**, 033055 (2021).

[17] 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**, 030305 (2021).

[18] Robert M. Parrish and Peter L. McMahon. "Quantum filter diagonalization: Quantum eigendecomposition without full quantum phase estimation" (2019). arXiv:1909.08925.

[19] Ignacio Loaiza, Alireza Marefat Khah, Nathan Wiebe, and Artur F Izmaylov. "Reducing molecular electronic hamiltonian simulation cost for linear combination of unitaries approaches". Quantum Science and Technology (2022).

[20] Tzu-Ching Yen and Artur F Izmaylov. "Cartan subalgebra approach to efficient measurements of quantum observables". PRX Quantum **2**, 040320 (2021).

[21] Seonghoon Choi, Ignacio Loaiza, and Artur F. Izmaylov. "Fluid fermionic fragments for optimizing quantum measurements of electronic Hamiltonians in the variational quantum eigensolver". Quantum **7**, 889 (2023).

[22] Artur F Izmaylov, Tzu-Ching Yen, Robert A Lang, and Vladyslav Verteletskyi. "Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method". Journal of chemical theory and computation **16**, 190–195 (2019).

[23] Mario Motta, Erika Ye, Jarrod R McClean, Zhendong Li, Austin J Minnich, Ryan Babbush, and Garnet Kin Chan. "Low rank representations for quantum simulation of electronic structure". npj Quantum Information **7**, 1–7 (2021).

[24] Edvin Deadman, Nicholas J Higham, and Rui Ralha. "Blocked schur algorithms for computing the matrix square root". In International Workshop on Ap-

plied Parallel Computing. Pages 171–182. Springer (2012).

[25] Mario Barbatti, Adélia JA Aquino, and Hans Lischka. "Ultrafast two-step process in the non-adiabatic relaxation of the ch2 molecule". Molecular Physics **104**, 1053–1060 (2006).

[26] Bernhard Sellner, Mario Barbatti, Thomas Müller, Wolfgang Domcke, and Hans Lischka. "Ultrafast non-adiabatic dynamics of ethylene including rydberg states". Molecular physics **111**, 2439–2450 (2013).

[27] Thom H. Dunning. "Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen". J. Chem. Phys. **90**, 1007–1023 (1989).

[28] 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).

[29] 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" (2018).

[30] Elvira R. Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia. "Automated construction of molecular active spaces from atomic valence orbitals". Journal of Chemical Theory and Computation **13**, 4063–4078 (2017).

[31] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. "Pyscf: the python-based simulations of chemistry framework". WIREs Computational Molecular Science **8**, e1340 (2018).

[32] Qiming Sun, Xing Zhang, Samragni Banerjee, Peng Bao, Marc Barbry, Nick S. Blunt, Nikolay A. Bogdanov, George H. Booth, Jia Chen, Zhi-Hao Cui, Janus J. Eriksen, Yang Gao, Sheng Guo, Jan Hermann, Matthew R. Hermes, Kevin Koh, Peter Koval, Susi Lehtola, Zhendong Li, Junzi Liu, Narbe Mardirossian, James D. McClain, Mario Motta, Bastien Mussard, Hung Q. Pham, Artem Pulkin, Wirawan Purwanto, Paul J. Robinson, Enrico Ronca, Elvira R. Sayfutyarova, Maximilian Scheurer, Henry F. Schurkus, James E. T. Smith, Chong Sun, Shi-Ning Sun, Shiv Upadhyay, Lucas K. Wagner, Xiao Wang, Alec White, James Daniel Whitfield, Mark J. Williamson, Sebastian Wouters, Jun Yang, Jason M. Yu, Tianyu Zhu, Timothy C. Berkelbach, Sandeep Sharma, Alexander Yu. Sokolov, and Garnet Kin-Lic Chan. "Recent developments in the pyscf program package". The Journal of Chemical Physics **153**, 024109 (2020).

[33] Florian Weigend and Reinhart Ahlrichs. "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy". Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[34] Kianna Wan, William J. Huggins, Joonho Lee, and Ryan Babbush. "Matchgate shadows for fermionic quantum simulation".

Communications in Mathematical Physics **404**, 629–700 (2023).

[35] Guang Hao Low. "Classical shadows of fermions with particle number symmetry" (2022). arXiv:2208.08964 [quant-ph].

[36] Roy Frostig, Matthew Johnson, and Chris Leary. "Compiling machine learning programs via high-level tracing". SysML (2018). url: https://mlsys.org/Conferences/doc/2018/146.pdf.

[37] Oumarou Oumarou, Maximilian Scheurer, Robert M. Parrish, Edward G. Hohenstein, and Christian Gogolin. "Data for Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization". Zenodo (2023).

[38] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. "Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization". ACM Transactions on mathematical software (TOMS) **23**, 550–560 (1997).

[39] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". Nature Methods **17**, 261–272 (2020).

[40] Joshua J. Goings, Alec White, Christofer S. Tautermann, Matthias Degroote, Craig Gidney, Toru Shiozaki, Ryan Babbush, and Nicholas C. Rubin. "Data for reliably assessing the electronic structure of cytochrome p450 on today's classical computers and tomorrow's quantum computers". Zenodo (2022).

[41] Dougal Maclaurin, David Duvenaud, and Ryan P Adams. "Autograd: Effortless gradients in numpy". In ICML 2015 AutoML workshop. Volume 238. (2015). url: https://indico.ijclab.in2p3.fr/event/2914/contributions/6483/subcontributions/180/attachments/6060/7185/automl-short.pdf.

Appendix A: RC-DF optimization procedure

The RC-DF cost function is

$$C(X,Z) := \frac{1}{2}\|\Delta_{pqrs}\|_{\mathcal{F}}^2 + \sum_{tkl}\rho_{tkl}|Z_k^t l|^\gamma, \tag{18}$$

with regularization tensor ρ_{tkl} and $\gamma = 1$ in the L1 case and $\gamma = 2$ in the L2 case. Just like C-DF as outlined in [18], also for RC-DF it is advisable to use a nested two-step optimization process to minimize C with respect to X and Z following these steps:

1. Update the X_{pq}^t for fixed Z_{kl}^t. This can be done with a gradient based optimizer using the gradient

$$\frac{\partial C}{\partial X_{pq}^t} = \frac{\partial\|\Delta_{pqrs}\|_{\mathcal{F}}^2}{\partial X_{pq}^t} = \sum_{mn}\frac{\partial C}{\partial U_{mn}^t}\frac{\partial U_{mn}^t}{\partial X_{pq}^t}. \tag{19}$$

2. Determine the optimal Z_{kl}^t given the updated X_{pq}^t by solving

$$\frac{\partial C}{\partial Z_{kl}^t} = 0. \tag{20}$$

The gradient with respect to the orbital rotations' generators X_{pq}^t for a given Z_{pq}^t is independent of the regularization and stays unchanged compared to the original C-DF as the U_{pq}^t do not appear in the regularization contribution to the cost function. Therefore

$$\frac{\partial C}{\partial U_{mn}^t} = -4\sum_{qrsl}\Delta_{mqrs}U_{qn}^t Z_{nl}^t U_{rl}^t U_{sl}^t. \tag{21}$$

Let us now look at the second step. In the L2 case (20) yields

$$\frac{\partial C}{\partial Z_{kl}^t} = -\sum_{pqrs}\Delta_{pqrs}U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t + \rho_{kl}^t Z_{kl}^t = 0, \tag{22}$$

by replacing Δ_{pqrs} by its expression in 8, we have

$$\sum_{pqrs}(pq|rs)U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t = \rho_{kl}^t Z_{kl}^t + \sum_{omn}\left[\sum_p U_{pk}^t U_{pm}^o\right]\left[\sum_q U_{qk}^t U_{qm}^o\right] Z_{mn}^o \left[\sum_r U_{rl}^t U_{rn}^o\right]\left[\sum_s U_{sl}^t U_{sn}^o\right]. \tag{23}$$

Note that the left-hand side of the equation is independent of Z_{mn}^o i.e constant, and the right-hand side can be written as a linear combination of the unknowns Z_{mn}^o. Therefore, we have a system of linear equations of the form

$$b_{kl}^t((pq|rs), U_{pq}^t) = \sum_{omn} A_{omn}^{tkl}(U_{pq}^t, \rho_{kl}^t)Z_{mn}^o \tag{24}$$

with

$$b((pq|rs), U_{pq}^t)_{tkl} = \sum_{pqrs}(pq|rs)U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t \tag{25}$$

and

$$A(U_{pq}^t, \rho_{kl}^t)_{tkl,omn} = \sum_{omn}\left(\left[\sum_p U_{pk}^t U_{pm}^o\right]\left[\sum_q U_{qk}^t U_{qm}^o\right]\left[\sum_r U_{rl}^t U_{rn}^o\right]\left[\sum_s U_{sl}^t U_{sn}^o\right] + \delta_{(omn,tkl)}\rho_{tkl}\right). \tag{26}$$

In the L1 case, (20) yields

$$\frac{\partial C}{\partial Z^t_{kl}} = -\sum_{pqrs} \Delta_{pqrs} U^t_{pk} U^t_{qk} U^t_{rl} U^t_{sl} + \rho^t_{kl} \mathrm{sign}(Z^t_{kl}) = 0, \tag{27}$$

which implies

$$\sum_{pqrs} (pq|rs) U^t_{pk} U^t_{qk} U^t_{rl} U^t_{sl} - \rho^t_{kl} \mathrm{sign}(Z^t_{kl}) = \sum_{omn} \left[\sum_p U^t_{pk} U^o_{pm}\right] \left[\sum_q U^t_{qk} U^o_{qm}\right] Z^o_{mn} \left[\sum_r U^t_{rl} U^o_{rn}\right] \left[\sum_s U^t_{sl} U^o_{sn}\right] \tag{28}$$

which is again a system of equations of the form

$$b^t_{kl}\left((pq|rs), U^t_{pq}, \rho^t_{kl}, Z^t_{kl}\right) = \sum_{omn} A^{tkl}_{omn}(U^t_{pq}) Z^o_{mn}. \tag{29}$$

with

$$b^t_{kl}((pq|rs), U^t_{pq}, \rho^t_{kl}, Z^t_{kl}) = \sum_{pqrs} (pq|rs) U^t_{pk} U^t_{qk} U^t_{rl} U^t_{sl} - \rho^t_{kl} \mathrm{sign}(Z^t_{kl}) \tag{30}$$

$$A(U^t_{pq})_{tkl,omn} = \sum_{omn} \left[\sum_p U^t_{pk} U^o_{pm}\right] \left[\sum_q U^t_{qk} U^o_{qm}\right] \left[\sum_r U^t_{rl} U^o_{rn}\right] \left[\sum_s U^t_{sl} U^o_{sn}\right]. \tag{31}$$

So in both cases one can find the optimal Z^t_{kl} by pseudo-inverting $A_{tkl,omn}$. While pseudo-inverting the six-index tensor $A_{tkl,omn}$ to determine the Z^t_{kl} can be done for medium size systems, it is intractable for large systems. To circumvent this problem, the inversion can be carried out in a matrix-free manner with, e.g., a conjugate gradient algorithm. This procedure only requires the matrix-vector product and the matrix diagonal rather than the dense matrix. Lastly, any gradient-descent based optimization algorithm can be used. For the result presented in this work, we have used the L-BFGS [38] algorithm as implemented in SciPy [39]. All time critical routines for the steps above were just-in-time compiled and ran on an NVIDIA Tesla V100 with the help of JAX [36].

Appendix B: Tuning of the regularization

The regularization tensor ρ_{tkl} is a hyper-parameter that controls how much effort is put on achieving small $|Z^t_{pq}|$ versus reducing the Frobenius norm error. A good balance must be found to obtain both a small systematic energy error and a small variance and lambda value. We concentrate only the case that after truncating some number of n_t leafs the regularization is chosen uniformly $\rho_{tkl} =: \rho$.

In the main text we have quantified the overall performance of the measurement scheme by means of the MSE, which combines the systematic error in the energy because of the approximation of the Hamiltonian and the statistical error due to variance and shot noise. Here let us look at those two contributions separately. We consider the same data underlying Figure 2 from the main text.

In Figure 5 we show only the systematic error that results from RC-DF approximating the two body part of the Hamiltonian and represents what is achievable in the limit of infinitely many shots. The energy approximation error grows roughly linearly with $\rho/10$ and when the number of leafs is increased it seems to become easier for the optimizer to find Z^t_{kl} that are not too much distorted by the regularization and thus represent the $(pq|rs)$ tensor well leading to a smaller energy error.

In Figure 6 we show the standard deviation $\sqrt{\mathrm{Var}}$ of the ground state energy estimator, quantifying how far from the infinite shot budget limit a measured energy value is likely to lie. Larger regularization factors ρ manifestly reduce the standard deviation. If the shot distribution takes into account the weight of the leafs this effect is greatly enhanced. Very strong regularization does not seem to help. Increasing or decreasing the total shot budget would simply move the data point up and down according

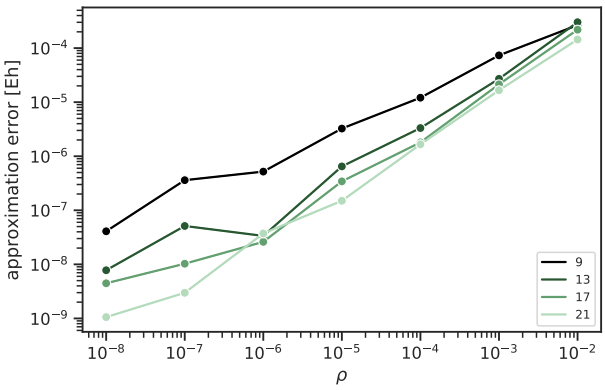


Figure 5: RC-DF systematic ground state energy approximation error of para-benzyne as a function of the regularization factor ρ with different numbers of leafs n_t.

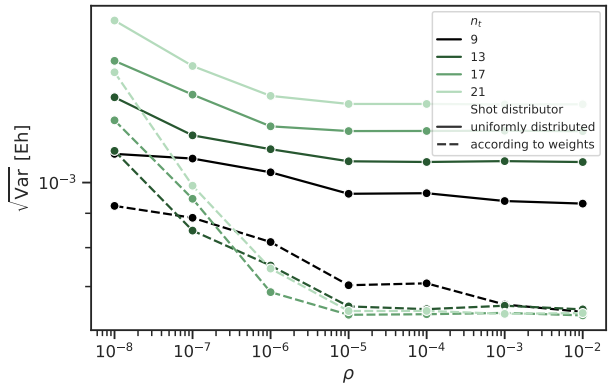


Figure 6: Standard deviation of the ground state energy estimator of para-benzyne for a budget of 300.000 shots as a function of ρ for different numbers of leafs n_t.

to the well known one over square root raw. As can be nicely seen there is a rather large window between $\rho \in [10^{-6}, 10^{-2}]$ where both errors (and thus their sum) is well below $10^{-}3$. In practice one find a good ρ by estimating the standard deviation on a quantum device starting from a comparably large regularization and then reduce the regularization to reduce the systematically error until the variance becomes too large to be compensated for by the affordable shot budget.

Appendix C: Lambda scaling with system size

To investigate the scaling of the lambda parameters achievable with double factorization in the large system size limit we use the hydrogen chain benchmark previously employed in the context of qubitization combined with low rank factorization in [12] and with THC in [17]. In this benchmark system neutral hydrogen atoms are placed with 1.4 Bohr distance on a line and represented in the STO-6G basis, which has one spatial orbital per hydrogen so that the size of the complete active space n is equal to the number of hydrogen atoms.

The lambda factor of the qubitization scheme from [12] is the sum of a one-body contribution λ_T and a two-body contribution λ_W and is strongly dominated by the latter. Between $n = 10$ and $n = 100$ the authors find that λ_W grows roughly proportional to $n^{2.5}$ to values up to about 2×10^5 (see Fig 11(a) in [12]).

In [17] a THC based qubitization scheme is proposed that improves over what the authors call the "naïve" approach. In Fig 20 of that work the two-body contribution λ_2 of these two schemes is

compared, again for $10 \leq n \leq 100$. The "naïve" λ_2 is found to grow approximately proportional to $n^{3.16}$ to values well beyond 5×10^6. The λ_2 achievable with the non-orthogonal basis THC based qubitization proposed in [17] (which corresponds to the two-body part of $\lambda_{\mathrm{THC}}^{\mathrm{Lee}}$) is found to only grow approximately proportionally to $n^{1.16}$ and reaches approximately 4×10^2 at $n = 100$. Fig 9 of the same work compares the full lambda values (including the one-body contribution) for different factorization methods including THC (the main focus of that work and the most competitve of the factorization methods in this plot) and the Cholesky based DF described around (10). The parameters of these methods are chosen to yield a CCSD(T) correlation energy error per atom of at most 50 micro Hartree, amounting to 0.5×10^{-3} Hartree at $n = 100$. The authors find a scaling of roughly $n^{1.88}$ for DF and of $n^{1.11}$ for THC and at $n = 100$ values of $\lambda > 2 \times 10^3$ for DF and $\lambda \approx 5 \times 10^2$ for THC.

With a regularization of $\rho = 5 \times 10^{-5}$ and even just a linear number of leafs $n_t = \lfloor (n+1)/2 \rfloor$ (red line in Fig. 7) we find that RC-DF (with two norm regularization $\gamma = 2$) is able to yield a constant Frobenius norm error and an absolute CCSD(T) error per atom $|\Delta_{\mathrm{CCSD(T)}}|/n$ in line with the 50 micro Hartree per atom used in Fig 9 of [17] (see the green line in Fig. 7b). At these parameters $\lambda_{\mathrm{RC-DF}}^{\mathrm{Burg}}$ is found to scale approximately like $n^{1.08 \pm 0.10}$ (fit through the values for $70 \leq n \leq 100$) and reaches approximately 2.5×10^2 at $n = 100$, roughly a factor of two better than the THC results from [17]. Compared to X-DF (which requires n_t to grow faster than linearly to obtain acceptable accuracy (see the blue and yellow lines in Fig. 7b), RC-DF yields roughly one order of magnitude lower lambda values at $n = 100$, mostly owing to the fact that we find that the X-DF lambda values scale roughly quadratic with n. λ^{Burg} and λ^{LCU} seem to have a similar scaling for all double factorization schemes considered here.

Appendix D: Necessary and sufficient condition for the symmetry of the V_{pq}^t

In this section we present a proof that shows that 8-fold symmetry of the $(pq|rs)$ tensor is a sufficient condition for the V_{pq}^t to be symmetric for every t where the eigenvalue $g_t \neq 0$. Hence, the Z_{pq}^t are real, which is required for the X-DF procedure to work, and the U_{pq}^t are orthogonal (and thud can be chosen to be special orthogonal without loss of generality), which is essential for their implementation on a quantum computer by means of a fabric of givens rotations. Slightly abusing notation, in the following lemma, we use $(pq|rs)$ for a four index tensor that does not necessarily arise from electron overlap integrals, but has the stated properties.

Lemma 1. *Let $(pq|rs)$ be any real, symmetric, i.e., $(pq|rs) = (rs|pq)$ tensor of shape $n \times n \times n \times n$. By grouping the indices pq and rs, let $(g_t)_t$ be its n^2 eigenvalues, V_{pq}^t its diagonalizing unitary and let $T_+ = \{t : g_t > 0\}$, $T_- = \{t : g_t < 0\}$, and $T \coloneqq T_+ \cup T_-$. Then $(pq|rs)$ can be written in the form*

$$(pq|rs) = \sum_{t \in T} V_{pq}^t \, g^t \, V_{rs}^t. \tag{32}$$

Further, if and only if $(pq|rs)$ is in addition 8-fold symmetric, i.e., $(qp|rs) = (qp|sr) = (pq|sr) = (pq|rs) = (rs|pq) = (sr|pq) = (sr|qp) = (rs|qp)$ the matrices $(V_{pq}^t)_{pq}$ are symmetric, i.e., $V_{pq}^t = V_{qp}^t \ \forall t \in T$, and $|T| \leq n(n+1)/2$.

Proof. That symmetry of the V_{pq}^t implies 8-fold symmetry of $(qp|rs)$ can be verified directly from (32) and whenever the V_{pq}^t are symmetric, since they are also orthogonal for different values of t, there can be at most $n\,(n+1)/2$ of them and thus the bond on the size $|T|$ of T holds.

To show that 8-fold symmetry implies symmetry of all V_{pq}^t with $t \in T$ let us first define

$$(pq|rs)_{\pm} \coloneqq \sum_{t \in T_{\pm}} V_{pq}^t \, g^t \, V_{rs}^t \tag{33}$$

$$|(pq|rs)| \coloneqq \sum_{t \in T} V_{pq}^t \, |g^t| \, V_{rs}^t. \tag{34}$$

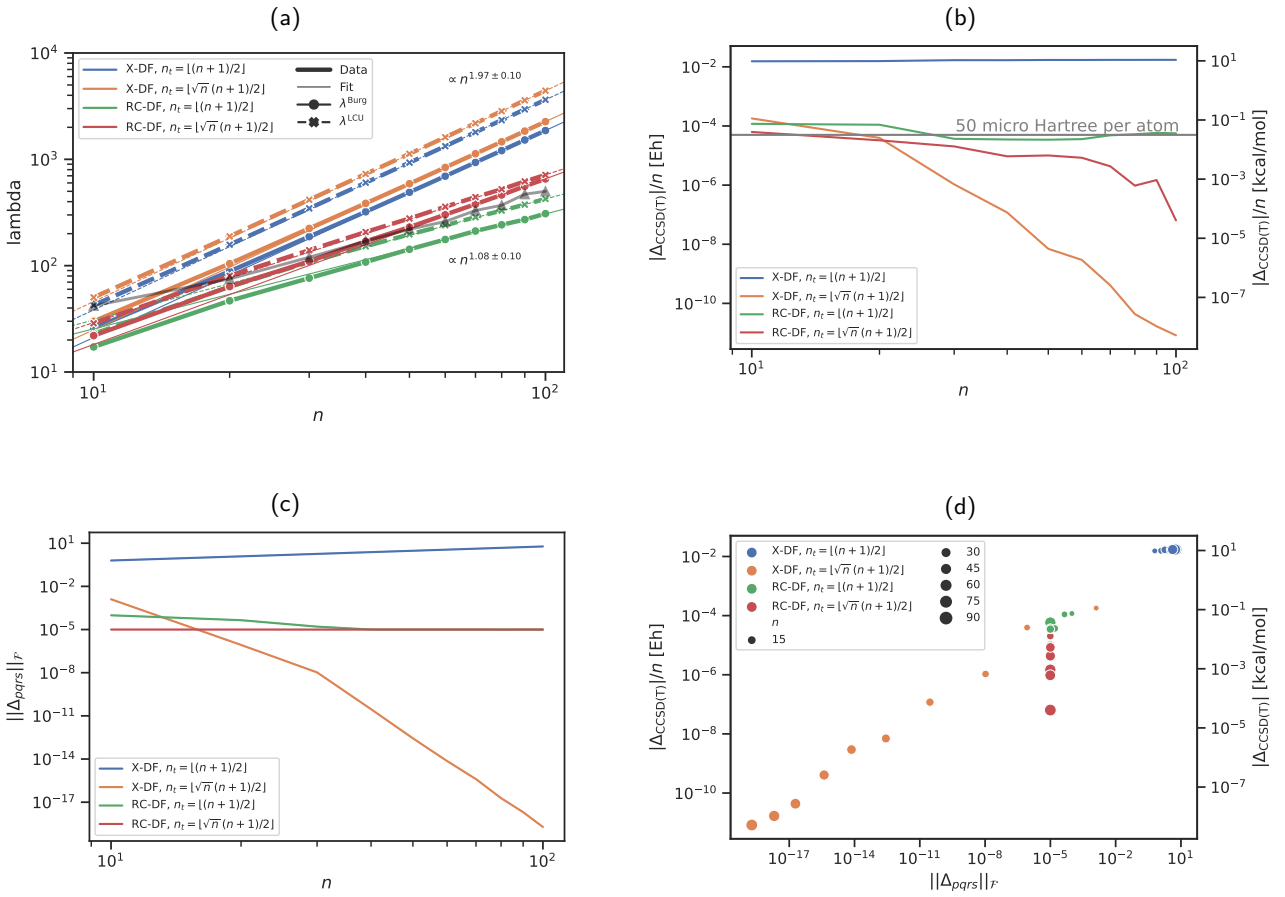


Figure 7: Comparison of lambda values achievable with X-DF and two-norm ($\gamma = 2$) RC-DF with regularization $\rho = 5 \times 10^{-5}$ and convergence tolerance $||\Delta_{pqrs}||_{\mathcal{F}} \leq 10^{-5}$ for the hydrogen chain benchmark in the STO-6G basis previously use in [12, 17] (panel a). The black triangles are the lambda values eyeballed from Fig. 9 of [17]. The thin lines are power law fits to the date for $70 \leq n \leq 100$ used to estimate the exponents stated in the text and displayed in the figure. Two different scalings of n_t with n were chosen such that the CCSD(T) error per particle $|\Delta_{\mathrm{CCSD(T)}}|/n$ (panel b) is approximately below 50 micro Hartree per atom for all cases execpt X-DF with linear number of leafs. The Frobenius norm error of the two-body tensor $\|\Delta_{pqrs}\|_{\mathcal{F}}$ (panel c) seems to be well correlated with $|\Delta_{\mathrm{CCSD(T)}}|$ for X-DF (panel d) but their magnitues can differ by several orders for other methods, so care needs to be taken when drawing conlusions base on $\|\Delta_{pqrs}\|_{\mathcal{F}}$ (The RC-DF data points are on a vertical line because $\|\Delta_{pqrs}\|_{\mathcal{F}} \leq 10^{-5}$ was used as an abort criterion for the RC-DF optimization).

so that we have $(pq|rs) \pm |(pq|rs)| = 2\,(pq|rs)_\pm$. Since we also have $|(pq|rs)| = (pq|rs)\,R$ with

$$R \coloneqq \sum_{t\in T} V^t_{pq}\,\mathrm{sign}(g^t)\,V^t_{rs}. \tag{35}$$

symmetric and invertible, 8-fold symmetry of $(pq|rs)$ implies 8-fold symmetry of $|(pq|rs)|$ and thereby 8-fold symmetry of both $(pq|rs)_+$ and $(pq|rs)_-$ individually. The symmetry $(pq|rs)_\pm = (qp|rs)_\pm$ then implies that

$$\sum_{t\in T_\pm} (V^t_{pq}V^t_{rs} - V^t_{qp}V^t_{rs})\,g^t = 0 \quad \forall p,q,r,s. \tag{36}$$

For $p = r$ and $q = s$ we can specialized this to

$$\sum_{t\in T_\pm} g^t V^t_{pq}V^t_{qp} = \sum_{t\in T_\pm} g^t (V^t_{pq})^2 \quad \forall p,q. \tag{37}$$

The symmetry $(pq|pq)_\pm = (qp|qp)_\pm$ together with (32) implies

$$\sum_{t\in T_\pm} g_t (V^t_{pq})^2 = \sum_{t\in T_\pm} g_t (V^t_{qp})^2 \tag{38}$$

and thus the left and right hand side of (37) can be identified to be the left and right hand side of the Cauchy-Schwarz inequalities of the inner product of the vectors $(\sqrt{g_t}^* V^t_{pq})_{t\in T_\pm}$ and $(\sqrt{g_t} V^t_{qp})_{t\in T_\pm}$:

$$\Big(\sum_{t\in T_\pm} (\sqrt{g_t}^* V^t_{pq})^* \sqrt{g_t} V^t_{qp}\Big)^2 = \tag{39}$$

$$\Big(\sum_{t\in T_\pm} g_t\, V^t_{pq}V^t_{qp}\Big)^2 \le \sum_{t\in T_\pm} |g_t| (V^t_{pq})^2 \sum_{t\in T_\pm} |g_t| (V^t_{qp})^2 \tag{40}$$

$$= \sum_{t\in T_\pm} \pm|g_t| (V^t_{pq})^2 \sum_{t\in T_\pm} \pm|g_t| (V^t_{qp})^2 \tag{41}$$

$$= \sum_{t\in T_\pm} g_t\,(V^t_{pq})^2 \sum_{t\in T_\pm} g_t\,(V^t_{qp})^2 \tag{42}$$

$$= \Big(\sum_{t\in T_\pm} g_t\,(V^t_{qp})^2\Big)^2 \quad \forall p,q, \tag{43}$$

where in the last step we have used (38). Thus (37) implies that these Cauchy-Schwarz inequalities are indeed fulfilled with equality, which is the case if and only if the two vectors in the Cauchy-Schwarz inequality are identical up to a prefactor, i.e., for each p, q there must exists a scalar α such that

$$\sqrt{g_t}^* V^t_{pq} = \alpha\,\sqrt{g_t} V^t_{qp} \quad \forall t \in T_\pm \tag{44}$$

$$\Leftrightarrow V^t_{pq} = \pm\alpha\, V^t_{qp} \quad \forall t \in T_\pm\ . \tag{45}$$

Since $(pq|rs) = (qp|rs)$, we further have

$$(pq|rs) = \sum_{t\in T_\pm} V^t_{pq} g^t V^t_{rs} = \pm\alpha \sum_{t\in T_p m} V^t_{qp} g^t V^t_{rs} \tag{46}$$

$$= \sum_{t\in T\pm} V^t_{qp} g^t V^t_{rs} = (qp|rs), \tag{47}$$

which is possible only if $\pm\alpha = 1$. Thus we have as claimed $V^t_{pq} = V^t_{qp}\ \forall t \in T$.

□

Appendix E: Computational methodology for Cpd I test case

The active space integrals for Cpd I were obtained from the deposited data [40] of Ref. [11]. The system under consideration here is labelled "X" in Ref. [11], and consists of a $(34\alpha{+}29\beta, 58o)$ active space. We factorized the two-body integrals using RC-DF with $n_t \in \{30, 40, 50, 60, 70, 80, 90, 100, 150, 200\}$ and additional $n_t = 400,\ 800$ with X-DF. For RC-DF, we employed $\rho = 10^{-3}$ and varied the convergence tolerance from 10^{-1} to 10^{-4}. From the factorized two-body integrals, a CCSD(T) energy was computed as described in Ref. [11] using the `chemftr` Python library (`https://github.com/ncrubin/chemftr`) interfaced with PySCF [31, 32]. The CCSD(T) energy error is the energy difference of the CCSD(T) energies with exact two-body integrals and the two-body integrals reconstructed from the factorization. For both factorization schemes, we evaluated $\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$ and $\lambda_{\mathrm{DF}}^{\mathrm{LCU}}$. Since the data for the truncated DF scheme in Ref. [11] are not shown in the paper, we recomputed the factorization using `chemftr` and checked that for exact factorizations, the lambda parameters agree exactly with our X-DF implementation. The RC-DF and X-DF results are shown in Tables 1 and 2, respectively. In addition, Table 3 contains the recomputed results for truncated DF from Ref. [11]. The factorized Hamiltonians, together with the resulting energy errors and lambda factors were deposited on Zenodo [37].

Appendix F: Derivation of the double factorized Hamiltonian in terms of Pauli operators

Inserting (5) into (1) we have

$$H = E_c + \sum_{pq}(p|\hat{\kappa}|q)E_{pq}^{+} + \frac{1}{2}\sum_{tkl} Z_{kl}^{t} U^{t} E_{kk}^{+} E_{ll}^{+} U^{\dagger t} \tag{48}$$

with

$$(p|\hat{\kappa}|q) = (p|\hat{h}_c|q) - \frac{1}{2}\sum_{r}(pr|qr). \tag{49}$$

Using the Jordan Wigner mapping we can write

$$E_{kk}^{+} = I - \frac{\hat{Z}_k + \hat{Z}_{\bar{k}}}{2} \tag{50}$$

and using the following identity

$$E_{kk}^{+}E_{ll}^{+} = -I + E_{kk}^{+} + E_{ll}^{+} + \frac{1}{4}(\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}}) \tag{51}$$

yields

$$H = E_c - \frac{1}{2}\sum_{tkl} Z_{kl}^{t} + \sum_{pq}(p|\hat{\kappa}|q)E_{pq}^{+} + \sum_{tk}\sum_{l} Z_{kl}^{t} U^{t} E_{ll}^{+} U^{t\dagger} + \frac{1}{8}\sum_{tkl} Z_{kl}^{t} U^{t}(\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})U^{t\dagger}. \tag{52}$$

Our aim is now to sort terms according to whether they contain an even or odd number of $\hat{Z}$ operators to arrive at expression (6). Using again (5) we can rewrite $\sum_{tk}\sum_{l} Z_{kl}^{t} U^{t} E_{ll} U^{t\dagger} = \sum_{pq}\sum_{tk}\sum_{l} U_{pl}^{t} U_{ql}^{t} Z_{kl}^{t} E_{pq}^{+}$ and hence we have

$$(pq|rr) = \sum_{tkl} U_{pk}^{t} U_{qk}^{t} Z_{kl}^{t} {U_{rl}^{t}}^{2} \tag{53}$$

$$\implies \sum_{r}(pq|rr) = \sum_{tkl} U_{pk}^{t} U_{qk}^{t} Z_{kl}^{t}. \tag{54}$$

Using these expressions and the shorthand $\mathcal{F}_{pq}$ defined in (3) we can rewrite (52) to read

$$H = E_c - \frac{1}{2}\sum_{tkl} Z^t_{kl} + \sum_{pq}((p|\hat{\kappa}|q) + \sum_{tk}\sum_{l} U^t_{pl}U^t_{ql}Z^t_{kl})E^+_{pq} + \frac{1}{8}\sum_{tkl} Z^t_{kl}U^t(\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})U^{t\dagger} \tag{55}$$

$$= E_c - \frac{1}{2}\sum_{tkl} Z^t_{kl} + \sum_{pq}((p|\hat{\kappa}|q) + \sum_{r}(pq|rr))E^+_{pq} + \frac{1}{8}\sum_{tkl} Z^t_{kl}U^t(\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})U^{t\dagger} \tag{56}$$

$$= E_c - \frac{1}{2}\sum_{tkl} Z^t_{kl} + \sum_{pq}\mathcal{F}_{pq}E^+_{pq} + \frac{1}{8}\sum_{tkl} Z^t_{kl}U^t(\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})U^{t\dagger} \tag{57}$$

$$= E_c - \frac{1}{2}\sum_{tkl} Z^t_{kl} + \sum_{k}\mathcal{F}^{\varnothing}_{k}U^{\varnothing\dagger}E^+_k U^{\varnothing} + \frac{1}{8}\sum_{tkl} Z^t_{kl}U^t(\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})U^{t\dagger} \tag{58}$$

Replacing E^+_k according to (50) and pulling out the $k = l$ terms from the last sum we arrive at

$$H = E_c - \frac{1}{2}\sum_{tkl} Z^t_{kl} + \sum_{p}\mathcal{F}^{\varnothing}_{p} + \frac{1}{8}\sum_{tk} Z^t_{kk} + \sum_{k}\mathcal{F}^{\varnothing}_{k}U^{\varnothing\dagger}(Z_k + Z_{\bar{k}})U^{\varnothing} + \tag{59}$$

$$\frac{1}{8}\sum_{tkl} Z^t_{kl}U^t\left(\hat{Z}_k\hat{Z}_l - \delta_{kl} + \hat{Z}_k\hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}}\hat{Z}_l + \hat{Z}_{\bar{k}}\hat{Z}_{\bar{l}} - \delta_{\bar{k}\bar{l}}\right)U^{t\dagger}. \tag{60}$$

Using 69 and 71, we write the total offset in the above equation as:

$$\mathcal{E} = E_c - \frac{1}{2}\sum_{tkl} Z^t_{kl} + \sum_{p}\mathcal{F}^{\varnothing}_{p} + \frac{1}{8}\sum_{tk} Z^t_{kk} \tag{61}$$

$$= E_c - \frac{1}{2}\sum_{pq}(pp|qq) + \sum_{p}\mathcal{F}^{\varnothing}_{p} + \frac{1}{8}\sum_{pq}(pq|pq) \tag{62}$$

Since $trace(\mathcal{F}_{pq}) = \sum_p \mathcal{F}^{\varnothing}_p$, we have:

$$\sum_{p}\mathcal{F}^{\varnothing}_{p} = trace(\mathcal{F}_{pq}) \tag{63}$$

$$= \sum_{p}\mathcal{F}_{pp} = \sum_{p}(p|\hat{\kappa}|p) + \sum_{r}(pp|rr) \tag{64}$$

$$= \sum_{p}((p|\hat{h}_c|p) - \frac{1}{2}\sum_{r}(pr|pr) + \sum_{r}(pp|rr)) \tag{65}$$

As a result we have:

$$\mathcal{E} = E_c - \frac{1}{2}\sum_{pq}(pp|qq) + \sum_{p}((p|\hat{h}_c|p) - \frac{1}{2}\sum_{r}(pr|pr) + \sum_{r}(pp|rr)) + \frac{1}{8}\sum_{pq}(pq|pq) \tag{66}$$

$$= E_c - \frac{1}{2}\sum_{pq}(pp|qq) + \sum_{p}(p|\hat{h}_c|p) - \frac{1}{2}\sum_{pr}(pr|pr) + \sum_{pr}(pp|rr)) + \frac{1}{8}\sum_{pq}(pq|pq) \tag{67}$$

$$= E_c + \frac{1}{2}\sum_{pq}(pp|qq) + \sum_{p}(p|\hat{h}_c|p) - \frac{1}{4}\sum_{pr}(pr|pr) \tag{68}$$

Using (5) again we can identify the the second term in the last expression to be

$$(pp|qq) = \sum_{tkl} {U^t_{pk}}^2 Z^t_{kl} {U^t_{ql}}^2 \tag{69}$$

$$\implies \sum_{pq}(pp|qq) = \sum_{tkl} Z^t_{kl}. \tag{70}$$

Similarly

Moreover, since for $k = l \implies \hat{Z}_k \hat{Z} l = I$, we have an extra offset of $\sum_{tk} Z^t_{kl}$:

$$(pq|pq) = \sum_{tkl} U^t_{pk} U^t_{qk} Z^t_{kl} U^t_{pl} U^t_{ql} \tag{71}$$

$$\implies \sum_{pq} (pq|pq) = \sum_{tkl} \sum_{p} U^t_{pk} U^t_{pl} Z^t_{kl} \sum_{q} U^t_{qk} U^t_{ql} \tag{72}$$

$$= \sum_{tkl} \delta_{kl} Z^t_{kl} \delta_{kl} \tag{73}$$

$$= \sum_{tk} Z^t_{kk} \tag{74}$$

Appendix G: Comparison of RC-DF and FFF

The Fluid Fermionic Fragments (FFF) method is based on the fact that some contributions to the Hamiltonian can be moved back and fourth freely between the second and third term of the electronic structure Hamiltonian as written in (1). In the fermionic picture, these "fluid" parts of the Hamiltonian correspond to terms that are quadratic in the creation and annihilation operators (see (9) an (10) in Ref [21]) and which, after diagonalization of the quadratic part contribute to the terms proportional to particle number operators, and which under Jordan Wigner yield Pauli $\hat{Z}$ operators.

Here we present the FFF method in the qubit picture. To that end, starting from (1), we first factorize the $(pq|rs)$ part of the Hamiltonian only, which yields:

$$\hat{H} = E_0 + \sum_{pq} (p|\hat{\kappa}|q) E^+_{pq} + \frac{1}{2} \sum_{pqrs} (pq|rs) E^+_{pq} E^+_{rs} \tag{75}$$

$$= E_0 + \sum_{pq} (p|\hat{\kappa}|q) E^+_{pq} + \frac{1}{2} \sum_{t} U^t (\sum_{kl} Z^t_{kl} E^+_{kk} E^+_{ll}) U^{t\dagger} \tag{76}$$

$$= E_0 + \sum_{pq} (p|\hat{\kappa}|q) E^+_{pq} + \frac{1}{2} \sum_{t} U^t (\sum_{kl} Z^t_{kl} (-I + E^+_{kk} + E^+_{ll} + \frac{1}{4} (\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})) U^{t\dagger} \tag{77}$$

$$= E_0 - \frac{1}{2} \sum_{tkl} Z^t_{kl} + \sum_{pq} (p|\hat{\kappa}|q) E^+_{pq} + \sum_{t} U^t (\sum_{k} (\sum_{l} Z^t_{kl}) E^+_{kk}) U^{t\dagger} \tag{78}$$

$$+ \frac{1}{8} \sum_{t} U^t (\sum_{kl} Z^t_{kl} ((\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})) U^{t\dagger} \tag{79}$$

Now we can add and subtract terms of the form $U^t(\sum_k c^t_k E^+_{kk}) U^{t\dagger} = \sum_k c^t_k \sum_{pq} u^t_{pk} u^t_{qk} E^+_{pq}$ with c^t_k the FFF coefficients to obtain

$$\hat{H} = E_0 - \frac{1}{2} \sum_{tkl} Z^t_{kl} + \sum_{pq} (p|\hat{\kappa}|q) E^+_{pq} - \sum_{t} U^t (\sum_{k} c^t_k E^+_{kk}) U^{t\dagger} + \sum_{t} U^t (\sum_{k} (\sum_{l} Z^t_{kl} + c^t_k) E^+_{kk}) U^{t\dagger} \tag{80}$$

$$+ \frac{1}{8} \sum_{t} U^t (\sum_{kl} Z^t_{kl} ((\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})) U^{t\dagger} \tag{81}$$

$$= E_0 - \frac{1}{2} \sum_{tkl} Z^t_{kl} + \sum_{pq} ((p|\hat{\kappa}|q) - \sum_{t} \sum_{k} u^t_{pk} u^t_{qk} c^t_k) E^+_{pq} + \sum_{t} U^t (\sum_{k} (\sum_{l} Z^t_{kl} + c^t_k) E^+_{kk}) U^{t\dagger} \tag{82}$$

$$+ \frac{1}{8} \sum_{t} U^t (\sum_{kl} Z^t_{kl} ((\hat{Z}_k + \hat{Z}_{\bar{k}})(\hat{Z}_l + \hat{Z}_{\bar{l}})) U^{t\dagger} \tag{83}$$

$$= \mathcal{E}' + U'^{\varnothing\dagger} \sum_{k} \mathcal{C}^{\varnothing}_k \hat{Z}_k U'^{\varnothing} - \frac{1}{2} \sum_{t} U^t (\sum_{k} (\sum_{l} Z^t_{kl} + c^t_k) \hat{Z}_k) U^{t\dagger} \tag{84}$$

$$+ \frac{1}{8} \sum_{t} U^t (\sum_{kl} Z^t_{kl} (\hat{Z}_k \hat{Z}_l - \delta_{kl} + \hat{Z}_k \hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}} \hat{Z}_l + \hat{Z}_{\bar{k}} \hat{Z}_{\bar{l}} - \delta_{\bar{k}\bar{l}}) U^{t\dagger} \tag{85}$$

with

$$\mathcal{E}' = E_0 - \frac{1}{2}\sum_{tkl} Z^t_{kl} + \sum_k \mathcal{C}_{kk} + \sum_{tk}(\sum_l Z^t_{kl} + c^t_k) + \frac{1}{4}\sum_{tk} Z^t_{kk} \tag{86}$$

$$\mathcal{C}_{pq} = (p|\hat{\kappa}|q) - \sum_t \sum_k u^t_{pk} u^t_{qk} c^t_k \tag{87}$$

and $U'^{\varnothing}$ and $\mathcal{C}^{\varnothing}_k$ are the diagonalizing unitaries and eigenvalues of $\mathcal{C}_{pq}$.

The case all $c^t_k = 0$ corresponds to how X-DF was introduced in [18] and this was taken as the prior art benchmark in [21]. The case

$$c^t_k = -\sum_l Z^t_{kl} \tag{88}$$

corresponds to the way we wrote the Hamiltonian in (6), with no single qubit $\hat{Z}$ contributions in the two body leafs. It turns out that neither of these choices is optimal with respect to variance and thus shot count and optimizing the c^t_k can further yield improvements,

Optimization can be done with a gradient based optimizer and in [21] as well as here we used LBFGSB as implemented in scipy. The number of shots is then optimized, via a proxy state, using a nested loop where in each iteration the coefficients c^t_k are updated using the partial derivative at fixed shot distribution then the shots are optimally distributed according to the variances computed with the new c^t_k in the proxy state. For simplicity and better comparability with the results form [21] we use the exact ground state as the proxy state. This is not efficient but [21] found little difference between using the real ground state and an approximate proxy state for which variances can be computed efficiently. We compute the needed derivatives by means of a fully auto-differentiable code that computes the variances as a function of the c^t_k [29, 36, 41].

For some cases for which we have performed simulations (see Figure 8) we find that all $c^t_k = 0$ and/or X-DF with the c^t_k corresponding to (6) is a local minimum and hence gradient based optimization of the c^t_k does not work. We consistently found good final shot counts from random uniformly distributed within $[0, 1[$ initializations. Alternatively one can initialize from coefficients according to (88) which lead to faster convergence but seems to yield the same final or very similar shot budgets. Overall we find that combining RC-Df with FFF yields the lowest shot budgets. The term mapping used in (6) is significantly better than choosing all $c^t_k = 0$ and RC-DF (with and without FFF) consistently outperforms X-DF.

Table 1: RC-DF performance summary for Cpd I

Conv. Tol.	n_t	$\|\Delta_{pqrs}\|_{\mathcal{F}}$	CCSD(T) error [mEh]	$\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$	$\lambda_{\mathrm{DF}}^{\mathrm{LCU}}$
10^{-1}	30	0.4472	−23.5907	300.6	456.5
	40	0.4412	−22.6134	311.0	468.6
	50	0.4382	−22.4729	320.6	478.0
	60	0.4415	−22.7342	327.5	478.1
	70	0.4452	−21.4039	336.3	483.7
	80	0.4352	−18.2827	343.0	490.5
	90	0.4310	−17.3526	349.9	496.1
	100	0.4278	−17.2627	355.8	498.1
	150	0.4232	−14.8979	385.9	521.0
	200	0.4258	−15.0266	417.9	545.2
10^{-2}	30	0.1411	−3.0769	278.2	430.4
	40	0.1411	−3.6537	290.8	448.5
	50	0.1408	−2.9621	299.5	462.7
	60	0.1414	−3.2143	304.6	470.0
	70	0.1406	−2.8366	308.6	475.4
	80	0.1407	−3.6169	313.8	482.2
	90	0.1407	−2.6580	318.7	489.2
	100	0.1409	−2.1238	321.7	492.1
	150	0.1410	−2.6875	337.2	508.8
	200	0.1413	−2.5034	351.2	524.2
10^{-3}	30	0.0447	−0.2292	244.4	386.6
	40	0.0447	−0.3581	257.1	394.1
	50	0.0447	−0.2405	268.0	409.6
	60	0.0447	−0.2515	276.5	422.8
	70	0.0447	−0.5585	284.0	434.3
	80	0.0447	−0.1995	288.6	442.6
	90	0.0447	−0.3395	292.2	450.4
	100	0.0447	−0.0137	295.8	456.6
	150	0.0446	−0.4667	310.3	481.2
	200	0.0446	−0.2690	321.1	499.3
10^{-4}	100	0.0141	0.0086	258.0	390.8
	150	0.0141	−0.0423	275.7	414.4
	200	0.0141	−0.0340	287.9	435.3

Table 2: X-DF performance summary for Cpd I

n_t	$\|\Delta_{pqrs}\|_{\mathcal{F}}$	CCSD(T) error [mEh]	$\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$	$\lambda_{\mathrm{DF}}^{\mathrm{LCU}}$
30	1.01455	15.5688	418.5	759.2
40	0.72606	−7.3444	431.6	784.9
50	0.58478	−8.5971	440.3	802.0
60	0.47298	−14.4322	447.9	816.8
70	0.38068	−3.1890	453.7	828.2
80	0.29495	−9.6247	458.8	838.3
90	0.21062	0.5473	462.2	844.9
100	0.15487	−3.9773	464.3	848.9
150	0.04288	−0.0570	469.8	859.8
200	0.01609	0.1208	471.7	863.4
400	0.00126	0.0092	473.0	866.0
800	0.00001	0.0003	473.2	866.3

Table 3: Truncated DF performance summary for Cpd I[a)]

n_t[b)]	threshold[c)]	$\|\Delta_{pqrs}\|_{\mathcal{F}}$	CCSD(T) error [mEh]	$\lambda_{\mathrm{DF}}^{\mathrm{Burg}}$
99	0.07500	0.4909	−73.8021	428.8
114	0.05000	0.3631	−17.4890	439.9
131	0.02500	0.1892	1.1578	455.1
178	0.01000	0.0836	2.6993	464.3
194	0.00750	0.0644	4.1638	466.1
207	0.00500	0.0457	1.8040	467.9
241	0.00250	0.0245	−0.1019	470.1
309	0.00100	0.0106	0.0266	471.7
364	0.00050	0.0054	0.0155	472.3
505	0.00010	0.0012	−0.0086	473.0
568	0.00005	0.0006	0.0037	473.1
702	0.00001	0.0001	0.0007	473.1

[a)] Recomputed with `chemftr`.
[b)] Referred to as L in Refs. [12, 17].
[c)] Eigenvector screening threshold with which the accuracy of the factorization is tuned, see Ref. [12].

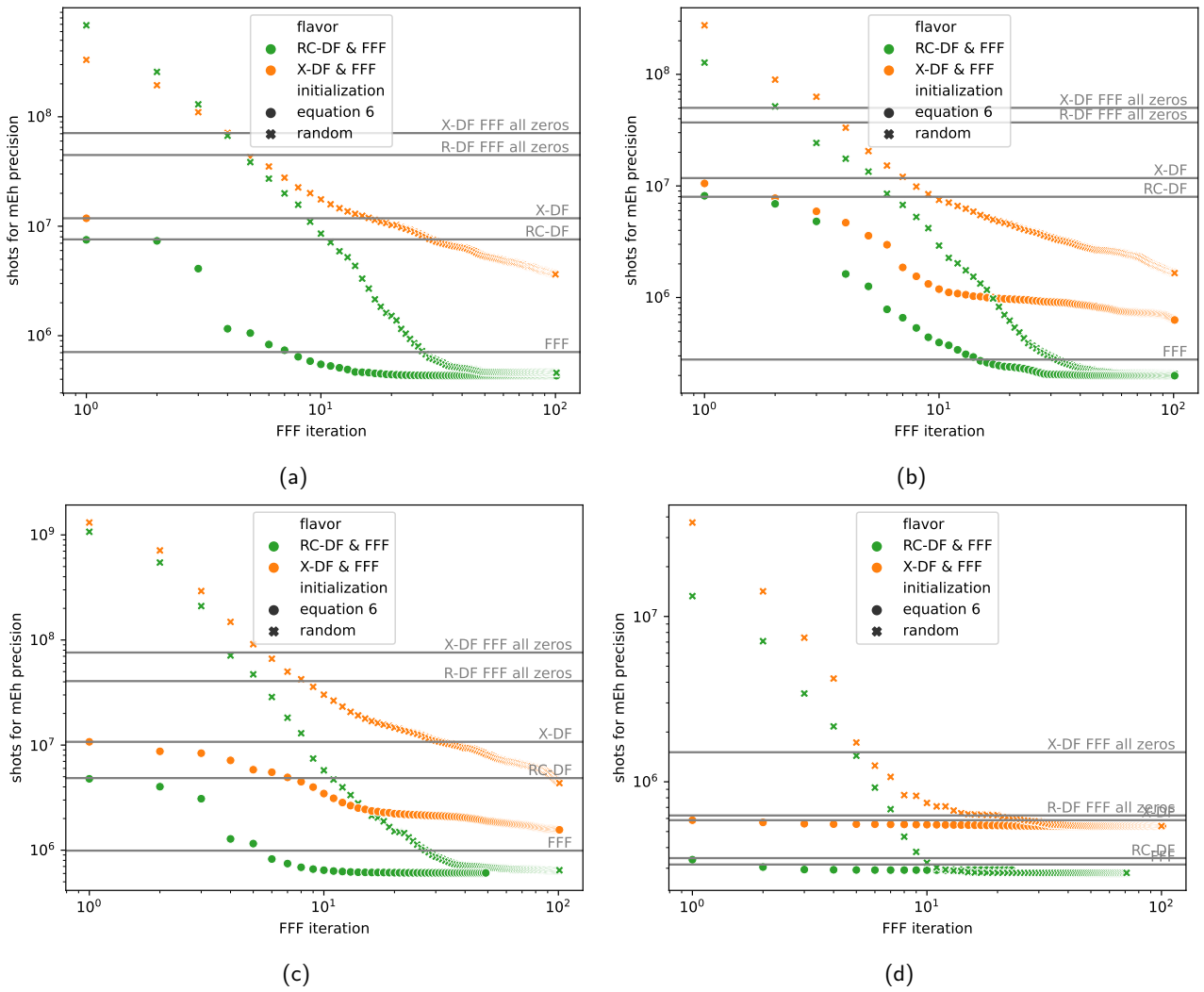


Figure 8: Number of shots to reach mili Hartree precision with optimal shot distribution for combinations of X-DF and RC-DF with the fluid fermionic fragments (FFF) method for (a) H_2O with $n_t = 24$ and (10e, 7o) active space, (b) HF with $n_t = 16$ and (10e, 6o) active space, (c) NH_3 with $n_t = 28$ and (10e, 8o) active space and (d) H_4 with $n_t = 10$ and (4e, 4o) active space (all with RHF orbitals in the STO-3G basis). The gray horizontal lines are the number of shots when the Hamiltonians are measured as written in (85) with all FFF coefficients c_k^t equal to zero (X-DF/RC-DF all zero) or as written in (6) (X-DF/RC-DF) and the best original results of FFF from [21]. The examples (a) to (c) were specifically picked because the gap between the plain RC-DF shot budget and the FFF shot budget from [21] were large. For other cases, plain RC-DF already yields similar results to FFF. The crosses/dots show how the number of shots decreases during optimization of the c_k^t FFF coefficients (with the true ground state taken as proxy state for simplicity) in (85) after initializing them randomly according to a uniform distribution within [0,1[/such that the initial Hamiltonian coincides with (6).

2 Relaxed Reduced Density Matrices and Energy Gradient within Double-Factorization Setup

To extend on the utility of double-factorization methods, we need to provide on top of the energy computation a method that computes the gradient of the energy as well. This is primarily motivated by the importance of energy gradients as explained in Section . In the next sections, we shall present our quantum algorithm [24] to calculate the analytical gradient of energies obtained with DF methods. We first start, in Section 2.1 by some preliminaries that set the problem and introduce the necessary functions and terms. Then, we present our result on efficient calculation of double-factorized energy gradient.

2.1 Double-Factorized Energy Gradient

Without loss of generality and only for the purpose of clarifying the XDF gradient, we consider a state with no response term coming from the ansatz that prepares it. This assumption applies to the standard cases of interest where we want to calculate the gradient of the optimal energy produced by a converged VQE (See [52]). To calculate the relaxed RDMs, we use the Lagrange formalism. We construct the Lagrangian with a complete set of constraints that the XDF parameters, namely the orbital rotations and the prefactors, must meet. The first class of constraints arises from the unitarity property of the orbital rotations U^t. The second set of constraints results from the first diagonalization and, more specifically, the relation between the unitaries V^t and the eigenvalues g_t. The last group of constraints originate from the second diagonalization, that is, the relation between the orbital rotations U^t and the eigenvalues Λ^t. We use the following notation to denote the expectation values in the t-leafs

$$\omega_k^{\varnothing} \coloneqq \langle\psi|\hat{U}^{\varnothing}\hat{Z}_k\hat{U}^{\varnothing}|\psi\rangle, k = 0..2n-1 \tag{51}$$

$$\omega_{kl}^{t} \coloneqq \langle\psi|\hat{U}^{\varnothing}\hat{Z}_k\hat{Z}_l\hat{U}^{\varnothing}|\psi\rangle, l > k = 0..2n-1 \tag{52}$$

and use the matrices T defined in [15] to decompose a given rotation. The expression of the XDF Lagrangian is

$$\begin{aligned}\mathcal{L}_{XDF} \coloneqq &\mathcal{E} + \sum_{k=0}^{2n-1} g^{\varnothing}_{\lfloor k/2\rfloor}\omega_k^{\varnothing} + \sum_{k>l=0}^{2n-1} Z^t_{\lfloor k/2\rfloor\lfloor l/2\rfloor}\omega^t_{kl} + \sum_{p>k}\eta^{\varnothing}_{pk}\left[\left(\prod_g T_g(\theta_g)\right)_{pk} - U^{\varnothing}_{pk}\right] + \\ &\sum_t\sum_{p>k}\eta^t_{pk}\left[\left(\prod_g T^t_g(\theta_g)\right)_{pk} - U^t_{pk}\right] + \sum_k \mu^{\varnothing}_{k,k'}g^{\varnothing}_{k,k'} + \sum_t\sum_{k>k'}\mu^t_{kk'}\Lambda^t_{kk'} + \sum_{t>t'}\nu^t_{tt'}g^{tt'}.\end{aligned} \tag{53}$$

We also note that the cost of calculating the energy gradient, in terms of number of observables, is asymptotically equal to that of calculating a single energy estimation. Then by solving the system of equations obtained through the Lagrangian stationarity, we obtain the Lagrangian coefficients. Thence, we deduce the relaxed RDM as the partial derivative of the stationary Lagrangian. We test the correctness of our method by comparing the resulting relaxed RDM to those of the finite-difference under different settings (full/truncated XDF, tightly/loosely converged VQE) method where the results matched to a very high accuracy.

For the CDF and RC-DF cases, we employ the Lagrangian formalism as well. However, the set of constraints differ between XDF and (R)C-DF. In addition to the constraints stemming from the unitarity condition of the orbital rotations U^t, we have an additional set of constraints arising from the convergence of the nested optimization procedure in the (R)C-DF method. That is, the gradient of the cost function with respect to the unitary generators must be null. Moreover, we have another constraint originating from the anti-symmetry of X^t_{mn} which are the generators of $U^t \coloneqq e^{X^t}$. One can bypass the addition of another constraint that captures the dependence of Z^t on X^t by simply considering the tensor Z^t as a function of X^t. The full expression of the (R)C-DF Lagrangian is

$$\begin{aligned}\mathcal{L}_{(R)C-DF} \coloneqq \mathcal{L} = &\mathcal{E} + \sum_k g^{\varnothing}_{\lfloor k/2\rfloor}\omega_k^{\varnothing} + \sum_t^T\sum_{kl} Z^t_{kl}(X^l_{mn})\omega^t_{kl} + \sum_{p>k}\eta^{\varnothing}_{pk}\left[\left(\prod_g T_g(\theta_g^{\varnothing})\right)_{pk} - U^{\varnothing}_{pk}\right] \\ &+ \sum_t^T\sum_{p>k}\eta^t_{pk}\left[\left(\prod_g T_g(\theta^t_g)\right)_{pk} - U^t_{pk}(X^l_{mn})\right] \\ &+ \sum_{k>k'}\mu^{\varnothing}_{kk'}g^{\varnothing}_{kk'} + \sum_t^T\sum_{p>k}\mu^t_{pk}\frac{\partial C}{\partial X^t_{pk}} + \sum_t\sum_{k\le k''}\nu^t_{kk''}(X^t_{kk''} + X^t_{k''k})\end{aligned} \tag{54}$$

where C is the cost function that is optimized in (R)C-DF [47, 16].

2.2 Efficient quantum analytic nuclear gradients with double factorization

Bibliographic Information

Hohenstein, E. G., **Oumarou, O.**, Al-Saadon, R., Anselmetti, G.-L. R., Scheurer, M., Gogolin, C., & Parrish, R. M. (2023). Efficient quantum analytic nuclear gradients with double factorization. The Journal of Chemical Physics, 158(11). https://doi.org/10.1063/5.0137167

Summary

The double-factorization method, which exploits the 8-fold symmetry in the electron repulsion integrals to reduce the four-index sum of fermioninc operators in the Hamiltonian to a three-index sum of diagonal fermionic operators in the suitable basis, is a very efficient method as it reduces the number of observables to be evaluated for a single point energy evaluation. Seeing the importance of first energy derivatives with respect to the one- and two-electron integrals otherwise known as relaxed reduced density matrices, we work out an analytical and efficient method for calculating said relaxed RDMs. The Lagrangian formalism we employ takes into account all the constraints that the exact and truncated double-factorization must fulfill. In the numerical results, we proof-test our Lagrangian formalism by comparing the nuclear gradient obtained with our method versus the finite difference one for truncated and exact Hamiltonains as well as converged and approximate VQEs. We found a high level of accuracy $10^{-}6$ to 10^{-7}. We perform two other benchmarks namely, QM/MM simulations of 327 quantum and 18470 total atoms and a geometry optimization of reactant, product and transition state geometries for the propylene oxide to allyl alcohol isomerization reaction.

Contribution

O.O co-derived the Lagrangian constraints and relaxed RDMs expressions, implemented and tested the double-factorized measurement module that builds and simulates the circuits to compute ω^t and their gradients χ^t as well as the relaxed RDMs equations in the in-house software stack of Covestro, and reviewed and edited the manuscript.

Efficient Quantum Analytic Nuclear Gradients with Double Factorization

Edward G. Hohenstein,[1] Oumarou Oumarou,[2] Rachael Al-Saadon,[1] Gian-Luca R. Anselmetti,[2] Maximilian Scheurer,[2] Christian Gogolin,[2,*] and Robert M. Parrish[1,†]

[1] *QC Ware Corporation, Palo Alto, CA 94301, USA*
[2] *Covestro Deutschland AG, Leverkusen 51373, Germany*

(Dated: July 28, 2022)

Efficient representations of the Hamiltonian such as double factorization drastically reduce circuit depth or number of repetitions in error corrected and noisy intermediate scale quantum (NISQ) algorithms for chemistry. We report a Lagrangian-based approach for evaluating relaxed one- and two-particle reduced density matrices from double factorized Hamiltonians, unlocking efficiency improvements in computing the nuclear gradient and related derivative properties. We demonstrate the accuracy and feasibility of our Lagrangian-based approach to recover all off-diagonal density matrix elements in classically-simulated examples with up to 327 quantum and 18470 total atoms in QM/MM simulations, with modest-sized quantum active spaces. We show this in the context of the variational quantum eigensolver (VQE) in case studies such as transition state optimization, ab initio molecular dynamics simulation and energy minimization of large molecular systems.

I. Introduction

Analytic derivatives of the energy are essential for efficient evaluation of chemical observables [1–4]. Geometry optimization and reaction pathway simulation are routine tasks in computational chemistry which require the repeated evaluation of nuclear derivatives. Other molecular properties such as the dipole and polarizability can be formulated as derivatives of the energy with respect to electric field perturbations. Mixed derivatives of nuclear and electric (or magnetic) perturbations define common spectroscopic observables [5]. While numerical techniques for the nuclear derivatives are often straightforward, they are inefficient and inexact – the compute time scales linearly with the number of perturbations, which would be three times the total number of atoms in the case of the nuclear gradient. Only the availability of reasonably well scaling analytic derivatives makes realistic chemical simulations across large length and/or time scales practical and thus they also need to be developed for quantum algorithms for chemistry.

The key to making analytic derivative methods practical is to reduce the complexity of the derivative to be within a multiplicative prefactor of the energy computation [6]. For first derivatives, this appears to be possible for all classical electronic structure methods where the derivative is a well-defined quantity. The challenge in formulating the gradient is in dealing with the nested series of procedures that defines the final energy. As this pertains to noisy intermediate-scale quantum (NISQ) era quantum computing, the standard implementation of the variational quantum eigensolver (VQE) [7–9] involves the use of an atom-centered Gaussian orbital basis set, the construction of a set of molecular orbitals, and finally the optimization of the VQE wavefunction. To evaluate the derivative of the VQE energy, the response of the VQE wavefunction parameters, the molecular orbitals and the underlying basis functions to the perturbation must be considered [10]. This can be handled elegantly and with a minimal amount of computational effort within a Lagrangian formulation [11]. A Lagrangian is constructed that contains undetermined multipliers that are used to make the energy stationary with respect to the non-variational wavefunction parameters. In most electronic structure methods, the linear equations defining the Lagrange multipliers need to be solved only once and an arbitrary number of nuclear or electric field perturbations can then be evaluated. This removes the dependence of the scaling of the derivative on the number of perturbations, which constitutes the main computational advantage of analytic derivatives over numerical derivatives.

A major challenge for NISQ and FTQC quantum algorithms for chemistry is the complexity of the electron repulsion integral (ERI) contributions to the Hamiltonian, which manifests itself in large circuit depths and/or large numbers of distinct measurement bases and experimental repetitions (shots). Efficient algorithms for quantum chemistry [12–15] work around this problem by means of factorization techniques, which however add additional layers of complexity to the formulation of derivative properties. The case we consider here is the explicit double factorization (X-DF) of the Hamiltonian [12–16]. This factorization offers several advantages for quantum algorithms in the NISQ era and beyond. Perhaps most importantly, the diagonal density matrices necessary to compute the expectation value of the energy can be evaluated with a series of simultaneous measurements in each of these bases [14]. By reducing the number of measurements, the number of times the wavefunction must be constructed is immediately reduced. Additionally, in a finite and limited shot budget setup, shot budgetization is just as important as reducing the number of terms. X-DF proves to be very reliable in this respect providing a small variance in the distribution of the expectation values [14].

* christian.gogolin@covestro.com
† rob.parrish@qcware.com

Similar low-rank factorizations are ubiquitous in classical quantum chemistry. The most closely related to X-DF are the density fitting (or resolution of the identity) [17, 18] and Cholesky decomposition approximations [19–21]. It is the second transformation in X-DF that is omitted in standard algorithms because it is not clear that this would offer any practical advantage on classical hardware. Other factorizations such as the pseudospectral method [22–24], chain-of-spheres [25] and tensor hypercontraction (THC) [26–28] provide a more flexible decomposition of the ERI tensor, but bear fewer similarities to X-DF. It should be noted that superficially the THC factorization appears the most similar to X-DF, however, X-DF offers less algebraic flexibility and, thus, cannot reduce the scaling of exchange-like contractions. Of these integral factorizations, density fitting is probably the most widely applied in chemical simulations. In large part, this is because potential energy surfaces generated by density-fitted methods are continuous and it is straightforward to develop analytic expressions for their derivatives [29–31]. Other methods, such as Cholesky decompositions do not yield continuous potential surfaces in general. Additional restrictions must be placed on the formation of the Cholesky decomposition to obtain continuous potential surfaces and analytic derivatives thereof [32–34]. An advantage of X-DF is that it both defines continuous potential surfaces and its formulation makes it possible to define analytic derivatives.

The focus of our present work is on the development of analytic derivatives of the energy for hybrid quantum classical methods that leverage X-DF Hamiltonians. There are two particular challenges to the formulation of these derivatives. First, the factorization and approximation of the ERI tensor introduces new non-variational parameters whose response to perturbations must be considered. Second, the reduction in the number of measurements to compute the energy afforded by X-DF comes with a catch. Namely, the off-diagonal elements of the density matrices are not needed to compute the energy, but their effective inclusion is required to evaluate the gradient. To measure these off-diagonal matrix elements (e.g., by measurement of Pauli strings in the Jordan-Wigner representation) would negate some of the computational advantage offered by X-DF. Moreover, if X-DF is to be deployed on fermionic quantum simulator hardware, it may be formally impossible to directly measure these off-diagonal matrix elements (e.g., if Pauli gates to directly measure off-diagonal matrix elements in are inadmissible). Clearly, an alternative must be developed. To circumvent this limitation, we take a Lagrangian-based approach to the analytic derivative and introduce undetermined multipliers to account for all of the non-variational parameters that appear in the energy expressions of hybrid quantum classical methods that leverage X-DF Hamiltonians. Our formulation allows the exact derivative of the energy to be evaluated while retaining the advantages of the X-DF Hamiltonian. Importantly, the Lagrangian formulation allows us to avoid solving for the response of the energy to each perturbation individually.

II. Theory

The active space electronic Hamiltonian of a molecular system acting on the fermionic Fock space over some spin adapted orbital basis can be written as

$$\hat{H} = E_{\mathrm{c}} + \sum_{pq} (p|\hat{h}_{\mathrm{c}}|q)\hat{E}_{pq} + \frac{1}{2}\sum_{pqrs}(pq|rs)\left(\hat{E}_{pq}\hat{E}_{rs} - \delta_{qr}\hat{E}_{ps}\right), \quad (1)$$

where $\hat{E}_{pq} = \hat{p}^\dagger\hat{q} + \hat{\bar{p}}^\dagger\hat{\bar{q}}$ are the singlet excitation operators and $\hat{p}^\dagger$ ($\hat{\bar{p}}^\dagger$) is the alpha/spin up (beta/spin down) fermionic creation operator in orbital p, with the annihilation operator counterpart $\hat{p}$. E_{c} is the frozen core energy, combining the coulomb energy of the nuclei and the frozen core electrons, $(p|\hat{h}_{\mathrm{c}}|q)$ are the active space one-body integrals including contribution from the interaction with the frozen core and $(pq|rs)$ are the two-body integrals.

Double factorization yields a compressed representation [12, 14–16] of such Hamiltonian from which energy expectation values for any many-body state $|\Psi\rangle$ can be computed (or estimated from finite shot estimated expectation values) by means of

$$E \equiv \langle\Psi|\hat{H}|\Psi\rangle \equiv \mathcal{E} + \sum_k \mathcal{F}_k^{\varnothing}\omega_k^{\varnothing} + \sum_{t=1}^{n_t}\sum_{kl} Z_{kl}^t\omega_{kl}^t, \quad (2)$$

where we refer to the terms in the last equation as t-leaves for $t \in \{\varnothing, 1, \ldots, n_t\}$ and call

$$\omega_k^{\varnothing} \equiv \frac{1}{2}\langle\Psi(\varnothing)|\hat{Z}_k + \hat{Z}_{\bar{k}}|\Psi(\varnothing)\rangle \quad (3)$$

$$\omega_{kl}^t \equiv \frac{1}{8}\langle\Psi(t)|\hat{Z}_k\hat{Z}_l - \hat{I}\delta_{kl} + \hat{Z}_k\hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}}\hat{Z}_l + \hat{Z}_{\bar{k}}\hat{Z}_{\bar{l}} - \hat{I}\delta_{\bar{k}\bar{l}}|\Psi(t)\rangle \quad (4)$$

the double factorized eigenbasis one- and two-body reduced density matrices and the states $|\Psi(t)\rangle$ are obtained by Jordan Wigner mapping $|\Psi\rangle$ to qubits and rotating the state with a unitary U^t that implements an orthogonal rotation of the spatial orbitals. Furthermore, $\hat{Z}$ is a Pauli-Z operator whose subscript k ($\bar{k}$) indicates that it acts on the qubit corresponding to the alpha (beta) spin orbital corresponding to spatial orbtial k. $\mathcal{E}$ is a constant energy offset containing the core energy E_c.

The U^t as well as the vector $\mathcal{F}_k^{\varnothing}$ and the tensor Z_{kl}^t depend on the integrals $(p|\hat{h}_{\mathrm{c}}|q)$ and $(pq|rs)$, and in X-DF, they are determined by means of (nested) diagonalizations of these quantities. The resulting representation is exact for the case $n_t = N(N+1)/2$ where N is the number of active spatial orbitals. The repesentation can

be arbitrarily truncated, which can be meaningful if the t-leaves of the factorization are ordered appropriately. If quantities are to be measured by computing expectation values from finitely many shots a better approach can be to importance-sample the leafs according to their contribution to the overall variance of the estimator. Compressed double factorization (C-DF) produces representations of the same form but with U^t and Z^t_{kl} determined via a a more elaborate optimization procedure and typically yields a better approximation for the same number of t-leaves [16]. For more details see supplemental material Section I.A.

The U^t can be efficiently implemented on a quantum computer in linear circuit depth even with only linear nearest neighbor connectivity [12, 35–37] by means of a fabric of Givens rotations $\hat{G}_g$ with angles θ^t_g chosen such that

$$\prod_g \hat{G}_g(\theta^t_g) = U^t, \tag{5}$$

and the angles θ^t_g can be efficiently computed from U^t.

The double factorized eigenbasis one- and two-body reduced density matrices $\omega^{\varnothing}_k$ and ω^t_{kl} are sufficient to compute the energy but they are insufficient to reconstruct the standard full one and two-body fermionic reduced density matrices (RDMs),

$$\gamma_{pq} \equiv \frac{\mathrm{d}E}{\mathrm{d}(p|\hat{h}_c|q)} = \langle\Psi|\hat{E}_{pq}|\Psi\rangle \tag{6}$$

$$\Gamma_{pqrs} \equiv \frac{\mathrm{d}E}{\mathrm{d}(pq|rs)} = \frac{1}{2}\langle\Psi|\hat{E}_{pq}\hat{E}_{rs} - \delta_{qr}\hat{E}_{ps}|\Psi\rangle. \tag{7}$$

The first advantage of using equation (2) to compute or estimate the energy is that the number of distinct bases in which measurements need to be performed is bounded by $n_t + 1$ where $n_t \leq n(n+1)/2$. The circuit depth only increases linearly to accommodate the necessary transformation to the orbital basis. At the same time X-DF reduces variance and thereby shows improved error rates with greatly reduced shot counts [14].

Key to the efficient evaluation of the analytic gradient is avoiding the explicit computation of perturbation dependent response. One approach to obtaining the derivative of the energy is to simply apply the chain rule and evaluate all the partial derivatives that appear. While this will certainly lead to a correct result, it also leads to an amount of unnecessary work that scales linearly with the number of perturbations. For example, one would need to solve for $\frac{\partial V^t_{pq}}{\partial \xi}$ and $\frac{\partial U^t_{pk}}{\partial \xi}$ in order to account for the derivatives of the implicit parameters introduced in the X-DF two-electron operator. This is a well-known problem in electronic structure theory and can be avoided by using a Lagrangian formulation of the derivative [11].

These issues are avoided by use of the Lagrangian formalism. The idea behind the Lagrangian approach to analytic derivatives is to construct a Lagrangian whose value is equal to the energy (the quantity we wish to differentiate). To that, an undetermined multiplier is added corresponding to every non-variational, implicitly-defined quantity that contributes to the energy expression. These multiply a constraint chosen from the implicit definition of the parameters. We can then solve a linear system to make the Lagrangian stationary with respect to variation in the implicitly-defined parameters.

The X-DF Lagrangian takes the form

$$\begin{aligned}
\mathcal{L}_{\text{X-DF}} \equiv \mathcal{E} &+ \sum_k \mathcal{F}^{\varnothing}_{kk}\omega^{\varnothing}_k + \sum_t^T \sum_{kl} Z^t_{kl}\omega^t_{kl} \\
&+ \sum_{p>k} \eta^{\varnothing}_{pk} \left[\left(\prod_g \hat{G}_g(\theta^{\varnothing}_g) \right)_{pk} - U^{\varnothing}_{pk} \right] \\
&+ \sum_t^T \sum_{p>k} \eta^t_{pk} \left[\left(\prod_g \hat{G}_g(\theta^t_g) \right)_{pk} - U^t_{pk} \right] \\
&+ \sum_{k>k'} \mu^{\varnothing}_{kk'}\mathcal{F}^{\varnothing}_{kk'} \\
&+ \sum_t^T \sum_{k>k'} \mu^t_{kk'}\lambda^t_{kk'} + \sum_{t>t'} \nu^{tt'} g^{tt'},
\end{aligned} \tag{8}$$

where $\mathcal{F}^{\varnothing}_{kk'} = \sum_{pq} U^{\varnothing}_{pk}\mathcal{F}^{\varnothing}_{pq}U^{\varnothing}_{qk'}$ with

$$\mathcal{F}^{\varnothing}_{pq} = (p|\hat{h}_c|q) + \sum_r (pq|rr) - \frac{1}{2}\sum_r (pr|qr), \tag{9}$$

and Z^t_{kl}, $\lambda^t_{kk'}$, and $g^{tt'}$ are defined in a similar way in terms of nested orthogonal transformations of $(pq|rs)$. η^t_{pk}, μ^t_{pk}, and $\nu^{tt'}$ are Lagrange multipliers. The Lagrangian is designed such that it is stationary with respect to the U^t_{pk}, θ^t_g, $\lambda^t_{kk'}$, and $g^{tt'}$ resulting from the X-DF procedure and the construction of the Givens fabrics and for those U^t_{pk}, e.g., the $\mathcal{F}^{\varnothing}_{kk'}$ are diagonal.

The η^t_{pk} multipliers ensure that the Givens fabrics actually implement the corresponding orbital rotation U^t. The μ^t_{pk} make the Lagrangian stationary with respect to variations in the U^t_{pk}. The $\nu^{tt'}$ multipliers depend on derivatives of the two-body contribution to the energy as well as the μ^t_{pk} multipliers. These multipliers couple the t-leaves together and, in the case of truncated X-DF they account for rotations between t-leaves included in the definition of the Hamiltonian and those that were excluded. These multipliers make the Lagrangian stationary with respect to variation in the eigenvectors of the two-electron integrals

The full Lagrangian of any orbital based active space method will contain further contributions from orbital response, but those will depend on details of and can be taken care of on the level of the method that was used to construct the orbitals. A complete description of each term appearing in the Lagrangian can be found in the Supporting Information in addition to a detailed

derivation of the equations that must be solved to reach stationarity in all of the nonvariational parameters.

There is an interesting interplay between the η multipliers used to make the Lagrangian stationary with respect to variations in the parameterization of the Givens operators and the μ and ν multipliers corresponding to the X-DF matrix elements. The quantities that appear in the η constraint terms carry no explicit dependence on the perturbation, therefore they have no direct contribution to the analytic derivative. However, the η multipliers form the right-hand side of the linear system that defines the μ multipliers. In the case of $\mu^{\varnothing}$, this term does carry explicit dependence on the perturbation through the one-body Hamiltonian matrix elements. In the case of μ^t, it again carries no explicit dependence on the perturbation, but forms a portion of the right-hand side of the linear system defining ν. The ν terms carries explicit dependence on the perturbation through the ERIs. The nested $\eta^{\varnothing}$ and $\mu^{\varnothing}$ multipliers recover the off-diagonal one-body density matrix elements, while the series of η^t, μ^t and ν allow the full two-body density matrix to be recovered. In this way, we are able to recover the information content of the full density matrices through a Lagrangian-based approach without ever needing to measure the full density matrices directly.

We demonstrate how this can be done for the one body density matrix γ_{pq} and leave further details to the SI. We will define the one-body density matrix as the derivative of the Lagrangian (which has been made stationary with respect to all the X-DF non-variational parameters: $\theta_g^{\varnothing}$, θ_g^t, $U_{pk}^{\varnothing}$, U_{pk}^t, V_{pq}^t) with respect to the frozen core operator, $(p|h_c|q)$. We denote this Lagrangian (which omits the z_{pq} and x_{pq} multipliers) as $\mathcal{L}_{\text{X-DF}}$. Then with the Kronecker delta δ_{pq}

$$\frac{\partial \mathcal{L}_{\text{X-DF}}}{\partial (p|h_c|q)} = \gamma_{pq} = \delta_{pq} + \sum_k U_{pk}^{\varnothing} \omega_k^{\varnothing} U_{qk}^{\varnothing} + \sum_{kk'} U_{pk}^{\varnothing} \mu_{kk'}^{\varnothing} U_{qk'}^{\varnothing}. \tag{10}$$

The $\omega_k^{\varnothing}$ were already measured to determine the energy, the $U_{pk}^{\varnothing}$ are a result of the factorization, and it remains to determine the $\mu_{kk'}$. To this end, one starts from $\frac{\mathrm{d}\mathcal{L}_{\text{X-DF}}}{\mathrm{d}U_{pk}^{\varnothing}} = 0$ and arrives at

$$\mu_{ll'}^{\varnothing} = \frac{\eta_{ll'}^{\varnothing} - \eta_{l'l}^{\varnothing}}{\mathcal{F}_{l'l'}^{\varnothing} - \mathcal{F}_{ll}^{\varnothing}}. \tag{11}$$

The $\eta_{ll'}^{\varnothing}$ multipliers are in turn defined via the linear system of equations

$$\sum_{p>k} \eta_{pk}^{\varnothing} \underbrace{\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}\theta_{g'}} \left[\prod_g \hat{G}_g(\theta_g^{\varnothing}) \right]_{pk}}_{A_{g',pk}^{\varnothing}} = -\frac{\mathrm{d}E}{\mathrm{d}\theta_{g'}^t}, \tag{12}$$

which can be solved by inverting the square $N(N-1)/2$ matrix $-A_{g',pk}^{\varnothing}$ and contracting the result against $\frac{\mathrm{d}E}{\mathrm{d}\theta_{g'}^t}$, which is just the derivative of the energy with respect to the Givens angles and can be measured by means of the parameter shift rule.

An important consideration is the stability of this procedure and the potential presence of singularities in the linear equations defining the Lagrange multipliers. Singularities in Lagrange multipliers defined by (11) (or similar equations) can appear if eigenvalues of the underlying operator are degenerate due to symmetry. In this case, the numerator of the Lagrange multiplier should be zero by symmetry as well and, therefore, these blocks of the Lagrange multipliers should be zero by L'Hôpital's rule. Removal of these singularities can be accomplished through adaptation of the Lagrangian for molecular symmetry. The matrix $A_{g,pk}^{\varnothing}$ (and related matrices) may also be singular. To avoid these issues, we solve the linear system in equation (12) by pseudoinversion of this matrix; negligible eigenvalues are set to zero. In the practical application of the analytic derivatives that follows, we did not encounter any numerical issues related to singularities in the linear equations defining the Lagrange multipliers. Should instabilities issues arise they can be worked around in the following way: the energy (Equation (2)) is a linear function of $(p|\hat{h}_c|q)$ and $(pq|rs)$, so it does not matter at what position the derivatives in (6) and (7) are evaluated. That is, the reduced density matrices are a function of the state only and do not depend on $(p|\hat{h}_c|q)$ or $(pq|rs)$. For the purpose of computation of the reduced density matrices one is free to substitute other tensors that minimize the variance of the final quantity one is interested in computing (i.e., by contraction with the density matrices). For example, Equation (11) suggests to choose a substitute for $(p|\hat{h}_c|q)$ with in equally spaced eigenvalues (these eigenvalues replace $\mathcal{F}_{ll'}^{\varnothing}$) and an eigenbasis that makes $A_{g',pk}^{\varnothing}$ well conditioned. Investigating to which extent this allows to further reduce the influence of shot noise will be left to a future investigation.

III. Computational Details

We have implemented the analytic derivatives of VQE energies with X-DF Hamiltonians within an in-house C++/Python quantum circuit simulator code. Hamiltonian matrix elements, molecular orbitals and molecular orbital response are handled by the `Lightspeed` + `TeraChem` [38] code stack. Molecular mechanics computations are performed with `OpenMM` [39] interfaced through `Lightspeed`. VQE wave functions were constructed with quantum-number-preserving (QNP) gate fabrics [40] to create VQE wavefunctions similar to k-UpCCGSD [41, 42]. Molecular orbitals are obtained with Hartree-Fock (HF) and fractional occupation number Hartree-Fock (FON-HF) [43, 44]. Further details on

the systems we simulate are provided as needed in the remainder of the manuscript and Supporting Information.

IV. Results

To validate the correctness of our formulation and implementation of the derivatives of the X-DF Hamiltonian, we compare the analytic derivatives to numerical derivatives obtained from finite difference computations. For this test, we use the X-DF Hamiltonian to construct VQE wavefunctions for butadiene with a 4 electron in 4 orbital $(4e, 4o)$ active space and a 6-31G** basis set. There are four cases we will consider to determine the correctness of the analytic derivative. First, there is the case where the X-DF is exact (i.e., the eigenbasis of the two-electron operator is not truncated) and the VQE parameterization is sufficiently flexible to recover the exact ground state energy. In the second case, the X-DF of the Hamiltonian is approximate, but the VQE is still converged to the exact ground state of that approximate Hamiltonian. In the third case, the X-DF is exact, but the VQE solution is approximate. Finally, we consider the case where both the X-DF and the VQE wavefunctions are approximate. Our Lagrangian formulation of the analytic derivative accounts for truncation of the t-leaves thus accounting for both exact and approximate X-DF Hamiltonians. The analytic derivatives of exact and approximate VQE energies do not require additional effort due to the variational nature of the method.

The complete results of the comparison between our analytic and numerical derivatives can be found in the Supporting Information. In all of the cases described above, we obtain agreement within 10^{-6} E_h/a_0 indicating the correctness of our formulation and implementation of the analytic derivatives. Importantly, we "stress test" our response equations by using a severe truncation of the X-DF Hamiltonian. In the two cases where the X-DF Hamiltonian is approximate, we neglect all eigenvalues below 10^{-2} E_h resulting in only 4 of 10 t-leaves being retained. The agreement between the analytic and numerical derivatives is comparable with both the truncated and untruncated X-DF Hamiltonians, indicating that our formulation of the analytic derivative correctly handles these additional response contributions. While these numerical tests give us confidence in the correctness of derivation and implementation, they do not speak directly to the robustness of the approach, i.e., how stable the implementation is across larger regions of the potential energy surface.

One important application of analytic derivatives is in force evaluation during *ab initio* molecular dynamics (AIMD) simulations. These simulations also provide a numerical check for the consistency of the energy and its derivative. In order to perform conservative dynamics simulations within the microcanonical ensemble (i.e., the NVE ensemble), the forces must be obtained from the exact derivative of the potential: $F = -\frac{\mathrm{d}E}{\mathrm{d}R}$. If this is the case, it is possible to perform energy-conserving AIMD simulations.

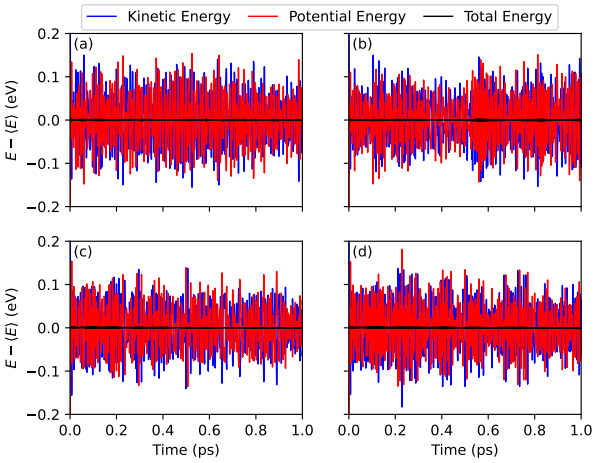


FIG. 1: *Ab initio* molecular dynamics simulation of butadiene using a $(4e, 4o)$ active space with VQE energies obtained using an X-DF Hamiltonian and a 6-31G** basis set. Integration of the classical equations of motion was performed using velocity Verlet with a 0.5 fs time step. (a) The VQE energy was converged to the exact ground state energy and the X-DF Hamiltonian was not truncated. (b) The VQE energy was converged to the exact ground state energy and the eigenvalues of the ERI tensor were truncated at 10^{-1} au. (c) The VQE energy was not converged to the exact ground state energy and the X-DF Hamiltonian was not truncated. (d) The VQE energy was not converged to the exact ground state energy and the eigenvalues of the ERI tensor were truncated at 10^{-3} au.

We again use butadiene as a test of our analytic derivatives (using the same level of theory as described previously). This time, we perform a 1 ps AIMD simulation using velocity Verlet integration [45] with a 0.5 fs time step. In this case, we use the fluctuations (and possible drift) of the total energy as a check for the correctness of the analytic derivatives of the energy. In addition, this tests the stability of the implementation across the regions of the potential energy surface explored by the AIMD trajectory. Discontinuities on the potential surface or instabilities in the response equations will lead to a loss of energy conservation. As before, we test the four cases of exactness and approximation in X-DF and VQE. The results of these simulations are shown in Figure 1. In each case, the AIMD simulations conserve energy indicating that the response equations needed to evaluate the analytic derivatives remain stable throughout. One should note that when the X-DF Hamiltonian is approximate, the potential energy surface only remains continuous if the two-electron operator eigenvalues at the truncation boundary do not become degenerate. That is, the "diabatic" character of the t-leaves must not change

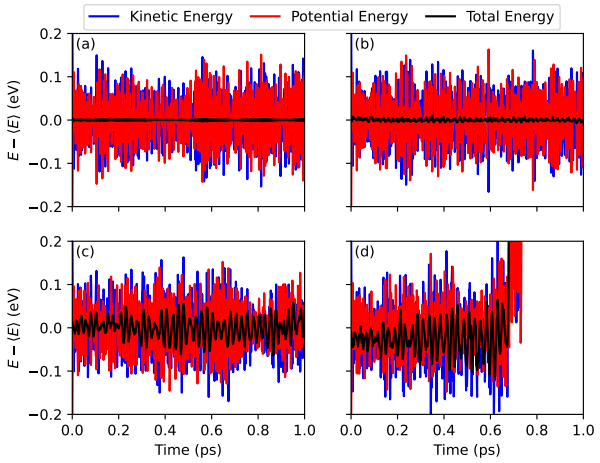


FIG. 2: *Ab initio* molecular dynamics simulation of butadiene using a $(4e, 4o)$ active space with VQE energies obtained using an X-DF Hamiltonian and a 6-31G** basis set. Integration of the classical equations of motion was performed using velocity Verlet with a 0.5 fs time step. In all cases, the eigenvalues of the ERI tensor were truncated at 10^{-1} au in the X-DF Hamiltonian and the VQE energy was converged to the exact ground state energy. (a) No approximations were made to the analytic gradient expressions, this is equivalent to Figure 1(b). (b) The $\eta^{\varnothing}$ and $\mu^{\varnothing}$ multipliers were set to zero during the evaluation of the analytic gradient. (c) The η^{t} and μ^{t} multipliers were set to zero during the evaluation of the analytic gradient. (d) The ν^{t} multipliers were set to zero during the evaluation of the analytic gradient.

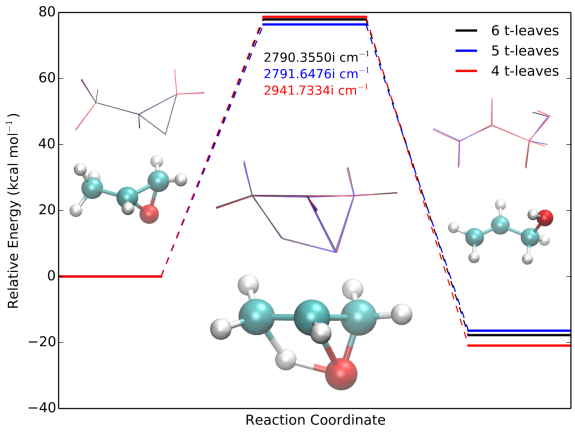


FIG. 3: Optimization of reactant, product and transition state geometries for the propylene oxide to allyl alcohol isomerization reaction. The potential energy surface is obtained from VQE using a $(4e, 3o)$ active space using an X-DF Hamiltonian and a 6-31G** basis set. The full X-DF Hamiltonian contains 6 t-leaves; results are presented with truncated X-DF expansions as well. The relative energies of these geometries are given in kcal mol^{-1}. At the transition state, the frequency of the imaginary mode (i.e., the negative Hessian eigenvalue) is given in cm^{-1}. The optimized structures are shown overlaid to compare the accuracy of the geometries obtained from different truncations of the X-DF Hamiltonian.

discontinuously with geometry. These eigenvalues define surfaces (in analogy to the Born-Oppenheimer potential energy surface) and those eigensurfaces may have conical intersections between them. In the event that such intersections occur, it must be the case that either both intersecting eigenvalues are retained or omitted in the X-DF Hamiltonian to maintain continuity of the potential energy surface.

To test the importance of the response contributions to the analytic derivative, we intentionally introduce errors in to the derivative expression by artificially setting some of the Lagrange multipliers to zero. Using these approximate derivatives, we perform AIMD simulations as before and test the energy conservation. The results of these simulations are shown in Figure 2. The response of the one-electron terms in the X-DF Hamiltonian can be removed by setting either the $\eta^{\varnothing}$ or the $\mu^{\varnothing}$ multipliers to zero. Remarkably, the neglect of these response contributions to the gradient does not have an outsized impact on energy conservation (although the AIMD simulation with these errors introduced clearly does not conserve energy as well as those using the exact analytic derivatives). What is remarkable about this is that setting these multipliers to zero effectively neglects the off-diagonal matrix elements of the one-body density matrix; one might have expected this to have a more deleterious effect on the dynamics of the system. In the third case, the η^{t} or, equivalently, the μ^{t} multipliers are set to zero. This removes significant contributions to the two-body density matrix and greatly affects the energy conservation. In this case, fluctuations in the total energy are nearly as large as the fluctuations in kinetic and potential energy. Despite this, the butadiene molecule remains intact throughout the 1 ps simulation. In the final case, the $\nu^{tt'}$ multipliers are set to zero (this also removes contributions from η^{t} and μ^{t}). With these errrors introduced to the derivative, it is no longer possible to perform the AIMD simulation for the full 1 ps. This indicates the importance of the response contributions to the analytic derivative – particularly the responses originating in the factorization of the two-electron operator.

Another important application of analytic derivatives of the energy is in the location and characterization of stationary points on the potential energy surface. As a representative example of this, we examine the isomerization reaction of propylene oxide to allyl alcohol (Figure 3) [46]. The potential energy surface is obtained by VQE with a $(4e, 3o)$ active space and a 6-31G** basis set –

the VQE parameterization is sufficient to obtain the exact ground state energy. At this level of theory, we optimize the geometries of the reactant (propylene oxide), product (allyl alcohol) and the transition state connecting them. The optimizations were converged tightly with the maximum absolute value of any component of the derivative being less than $10^{-6}\ E_\mathrm{h}/a_0$. To characterize the transition state, the nuclear Hessian was evaluated by finite differences of the analytic derivatives; from that Hessian, the harmonic vibrational frequencies could be determined. These optimizations were performed using exact and approximate X-DF Hamiltonians. Since the active space is relatively small, the exact X-DF Hamiltonian contains only 6 t-leaves. As a result, truncation of this expansion has a significant impact on the potential energy surface. We were only able to obtain optimized geometries when at least 4 t-leaves were retained in the factorization.

It has been well established in other methods leveraging approximate factorizations of the two-electron operator that relative energies can be obtained more accurately than absolute energies. This is particularly important to applications in chemistry where the absolute energy of a molecule is rarely – if ever – of practical importance. Case in point, the reaction energy and reaction barrier we obtain for the propylene oxide to allyl alcohol isomerization are relative energies that are chemically relevant. We find that the reaction barrier is slightly underestimated (by 1.5 kcal mol^{-1}) when 5 t-leaves are retained and slightly overestimated (by 0.8 kcal mol^{-1}) when 4 t-leaves are retained. Similar results are obtained for the reaction energy. The magnitude of the reaction energy is underestimated by 1.4 kcal mol^{-1} when 5 t-leaves are retained and overestimated by 3.1 kcal mol^{-1} when 4 t-leaves are retained. The imaginary frequency of the transition state is well reproduced with 5 t-leaves; an error on the order of 1 cm^{-1} is introduced in that case. With 4 t-leaves, the error increases to 150 cm^{-1}. In all cases, the geometries of these stationary points are in good agreement. This indicates that not only can X-DF be used in the characterization of chemical reactions, but also that truncated X-DF can be applied.

For an implementation of analytic gradients to be practical, it should be able to be applied to molecular systems with many nuclear degrees of freedom without a significant increase in the computational cost relative to the underlying energy evaluation. An example where this is particularly important is in hybrid quantum mechanics molecular mechanics (QM/MM) computations [48]. Here, the number of total nuclear degrees of freedom may be many orders of magnitude larger than what is contained in the quantum mechanical region alone. A representative example is shown in Figure 4. Here a QM region containing 50 atoms is coupled to an MM region containing 18470 atoms. The electronic Hamiltonian of the QM system contains the point charges associated with the nuclei in the QM region as well as all of the MM atoms. In the context of analytic derivatives, this means that

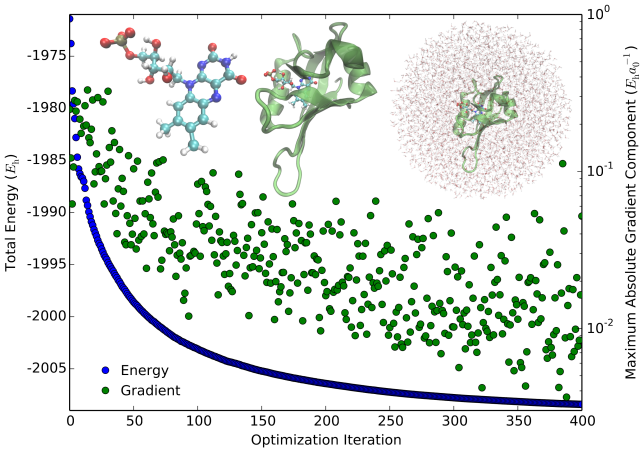


FIG. 4: QM/MM optimization of a flavin-binding light, oxygen or voltage (LOV) domain from the phototropin module of *Adiantum* phy3 in its dark state [47]. The QM region contains the flavin mononucleotide chromophore (50 atoms), while the remaining atoms are treated at the MM level. This chromophore is noncovalently bound to the protein. Including solvent molecules, this system contain 18470 atoms. The QM region is described by VQE using a $(4e, 4o)$ active space, an X-DF Hamiltonian without truncation and a 6-31G basis set; molecular orbitals were obtained from HF computation. The total energy of the system as well as the maximum absolute component of the gradient are reported for each iteration of an L-BFGS optimization performed in Cartesian coordinates.

there are 55410 nuclear perturbations that must be considered. A formulation of the gradient requiring explicit evaluation of perturbation-dependent response would be intractable. Our Lagrangian-based approach, however, allows QM/MM computations to be performed at the expense of evaluating the additional one-electron integrals (and their derivatives) and their subsequent contraction with the appropriate density matrices and Lagrange multipliers.

To demonstrate the ability to treat extended molecular systems, we optimized the geometry of a flavin mononucleotide-containing protein [47] in the presence of explicit solvent (Figure 4). We report the first 400 steps of an limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) optimization performed in Cartesian coordinates. The chromophore (containing 50 atoms) comprised the QM region and all 55410 nuclear degrees of freedom were relaxed during the optimization. Consistent progress was made in minimizing both the energy and gradient. This indicates the suitability of our approach for simulations of large molecular systems.

Another important consideration for implementations of analytic gradients is whether or not they scale ap-

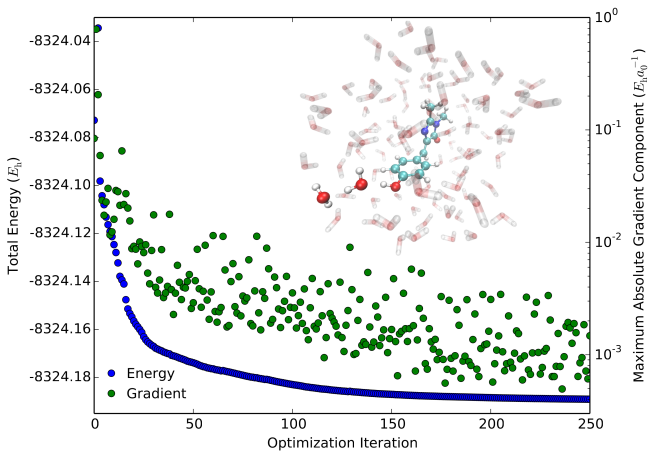


FIG. 5: Optimization of the anionic HBDI chromophore solvated by 100 water molecules including an extra proton. The entire is system is treated with VQE using a $(4e, 3o)$ active space, an X-DF Hamiltonian without truncation and a 6-31G** basis set; orbitals were generated with FON-HF using Gaussian broadening of the orbital energy levels and temperature parameter of 0.2 a.u. The optimization is started from a minimum energy conical intersection between the ground and first excited state of the chromophore. The total energy of the system as well as the maximum absolute component of the gradient are reported for each iteration of an L-BFGS optimization performed in Cartesian coordinates.

propriately with the size of the molecular system. It has been established that the matrix elements for active space methods such as VQE can be generated using the same algorithms used for Fock matrix construction in Hartree-Fock (HF) or Kohn-Sham density functional theory (KS-DFT) [49]. This means that – provided the size of the active space is constant – the VQE method should scale like HF or KS-DFT as the system size is increased. This also applies to analytic derivatives of active space methods; the analytic derivative can be cast in terms of derivatives of Fock-like quantities [49, 50]. All of this is to say that an implementation of VQE gradients should be able to be performed on large molecular systems. An example of such a system is shown in Figure 5. Here, the HBDI (4-hydroxybenzylidene-1,2-dimethylimidazolinone) chromophore of the green fluorescent protein (GFP) is solvated by 100 water molecules [51]. The starting structures were obtained from minimum energy conical intersection optimization using floating occupation molecular orbital complete active space configuration interaction (FOMO-CASCI) [52]. Since these optimizations were possible with classical electronic structure methods, it is essential that they are also possible using VQE with an X-DF Hamiltonian.

The optimizations shown in Figure 5 highlight a few important aspects of our analytic gradient formulation. First, and perhaps most obvious, is that it was possible to perform many optimization cycles for a molecular system containing over 300 atoms and 2800 basis functions. This is because our implementation scales appropriately with the size of the overall system as well as the size of the active space. This optimization intentionally used a starting structure near a conical intersection to emphasize the ability of methods like VQE for describing electronic structure in the presence of strong correlation. The optimization proceeded smoothly despite starting at such a structure. Finally, the starting structure we chose was for an anionic HBDI chromophore that had previously dissociated a proton (bonded to two neighboring water molecules). Over the course of the first 50 optimization cycles, the proton transfers back to the HBDI chromophore that is now relaxing on its ground electronic state. By using explicit solvent molecules treated quantum mechanically, it is possible to observe such proton transfer reactions during optimization or molecular dynamics simulations. During the final 200 optimization cycles, the HBDI chromophore returns from the twisted geometry that minimizes the energy of the conical intersection to a planar geometry that minimizes the energy of the ground state.

V. Conclusions

We have presented a framework for evaluating analytic derivatives from X-DF Hamiltonians. The increased efficiency in quantum computation offered by X-DF is preserved in our derivative formulation. One of the principle advantages of X-DF is that is reduces the number of measurements required by evaluating the energy with density matrices expressed in the bases (t-leaves) determined by X-DF. While the information contained in these density matrices is sufficient to determine the energy, it is not sufficient to determine derivatives of the energy. Our Lagrangian formulation of the derivative recovers the missing information without the need to explicitly measure the missing off-diagonal contributions to the density matrices (and thereby limiting the computational advantage offered by X-DF). Importantly, our formulation of the analytic derivative does not require perturbation-dependent response, i.e., the quantum computing effort scales with the number of orbitals treated on the quantum computer only and not with the total number of atoms. This is an essential characteristic of a method that is intended to have practical applications in chemistry.

Our initial tests of the analytic derivatives used VQE wavefunctions, however, our Lagrangian framework can be extended to other quantum algorithms. For example, it would be straightforward to extend this to derivatives of multistate contracted VQE (MC-VQE) energies [53] where the VQE parameters are not variationally optimized for a single state. One would simply need to include additional Lagrange multipliers to make the

state-specific MC-VQE energies stationary with respect to their wavefunction parameters [54]. Further, this general approach can be extended to other factorizations of the Hamiltonian such as the compressed double factorization (C-DF) [16]. Finally, our work also opens the door to exploiting the advantages of compressed representations of the Hamiltonian when computing quantities beyond energy in an error corrected setting. This is a necessary step towards making quantum computation more broadly applicable in chemical simulation.

Acknowledgements: We thank Fotios Gkritsis for valuable discussions. QC Ware Corp. acknowledges generous funding from Covestro for the undertaking of 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).

Conflict of Interest: A portion of the quantum Lagrangian gradient methodology pertains to a provisional patent application filed jointly by Covestro and QC Ware Corp. EGH and RMP own stock/options in QC Ware Corp.

[1] D. M. Bishop and M. Randič, Ab initio calculation of harmonic force constants, J. Chem. Phys. **44**, 2480 (1966).
[2] P. Pulay, Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules, Mol. Phys. **17**, 197 (1969).
[3] Y. Yamaguchi, Y. Osamura, J. D. Goddard, and H. F. Schaefer, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory* (Oxford University Press, Oxford, 1994).
[4] P. Pulay, Analytical derivatives, forces, force constants, molecular geometries, and related response properties in electronic structure theory, WIREs Comput. Mol. Sci. **4**, 169 (2013).
[5] F. Jensen, *Introduction to computational chemistry* (John Wiley & sons, 2017).
[6] N. C. Handy and H. F. Schaefer, On the evaluation of analytic energy derivatives for correlated wave functions, J. Chem. Phys. **81**, 5031 (1984).
[7] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 1 (2014).
[8] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).
[9] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, Quantum Sci. Technol. **4**, 014008 (2018).
[10] J. E. Rice, R. D. Amos, N. C. Handy, T. J. Lee, and H. F. Schaefer, The analytic configuration interaction gradient method: Application to the cyclic and open isomers of the S_3 molecule, J. Chem. Phys. **85**, 963 (1986).
[11] T. U. Helgaker, Simple derivation of the potential energy gradient for an arbitrary electronic wave function, Int. J. Quantum Chem. **21**, 939 (1982).
[12] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Quantum simulation of electronic structure with linear depth and connectivity, Phys. Rev. Lett. **120**, 110501 (2018).
[13] D. W. Berry, C. Gidney, M. Motta, J. R. McClean, and R. Babbush, Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization, Quantum **3**, 208 (2019).
[14] W. J. Huggins, J. R. McClean, N. C. 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 Inf. **7**, 10.1038/s41534-020-00341-7 (2021).
[15] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. K.-L. Chan, Low rank representations for quantum simulation of electronic structure, npj Quantum Inf. **7**, 10.1038/s41534-021-00416-z (2021).
[16] J. Cohn, M. Motta, and R. M. Parrish, Quantum filter diagonalization with compressed double-factorized hamiltonians, PRX Quantum **2**, 10.1103/prxquantum.2.040352 (2021).
[17] O. Vahtras, J. Almlöf, and M. Feyereisen, Integral approximations for LCAO}-{SCF calculations, Chem. Phys. Lett. **213**, 514 (1993).
[18] M. Feyereisen, G. Fitzgerald, and A. Komornicki, Use of approximate integrals in ab initio theory. an application in MP2 energy calculations, Chem. Phys. Lett. **208**, 359 (1993).
[19] N. H. F. Beebe and J. Linderberg, Simplifications in the generation and transformation of two-electron integrals in molecular calculations, Int. J. Quantum Chem. **12**, 683 (1977).
[20] I. Røeggen and E. Wisløff-Nilssen, On the {Beebe}-linderberg two-electron integral approximation, Chem. Phys. Lett. **132**, 154 (1986).
[21] H. Koch, A. Sánchez de Merás, and T. B. Pedersen, Reduced scaling in electronic structure calculations using cholesky decompositions, J. Chem. Phys. **118**, 9481 (2003).
[22] R. A. Friesner, Solution of the Hartree–Fock equations by a pseudospectral method: Application to diatomic molecules, J. Chem. Phys. **85**, 1462 (1986).
[23] R. A. Friesner, Solution of the Hartree–Fock equations for polyatomic molecules by a pseudospectral method, J. Chem. Phys. **86**, 3522 (1987).
[24] R. A. Friesner, New methods for electronic structure calculations on large molecules, Annu. Rev. Phys. Chem. **42**, 341 (1991).
[25] R. Izsák and F. Neese, An overlap fitted chain of spheres exchange method, J. Chem. Phys. **135**, 144105 (2011).
[26] E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, Tensor hypercontraction density fitting. i. Quartic scaling second- and third-order Møller-Plesset perturbation theory, J. Chem. Phys. **137**, 044103 (2012).
[27] R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, Tensor hypercontraction. II. Least-squares

renormalization, J. Chem. Phys. **137**, 224106 (2012).
[28] R. M. Parrish, E. G. Hohenstein, N. F. Schunck, C. D. Sherrill, and T. J. Martínez, Exact tensor hypercontraction: A universal technique for the resolution of matrix elements of local finite-range N-body potentials in many-body quantum problems, Phys. Rev. Lett. **111**, 132505 (2013).
[29] R. A. Distasio, R. P. Steele, Y. M. Rhee, Y. Shao, and M. Head-Gordon, An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order møller-plesset perturbation theory: Application to alanine tetrapeptide conformational analysis, J. Comput. Chem. **28**, 839 (2007).
[30] J. Boström, F. Aquilante, T. B. Pedersen, and R. Lindh, Analytical gradients of Hartree–Fock exchange with density fitting approximations, J. Chem. Theory Comput. **9**, 204 (2012).
[31] M. G. Delcey, T. B. Pedersen, F. Aquilante, and R. Lindh, Analytical gradients of the state-average complete active space self-consistent field method with density fitting, J. Chem. Phys. **143**, 044110 (2015).
[32] F. Aquilante, R. Lindh, and T. B. Pedersen, Analytic derivatives for the cholesky representation of the two-electron integrals, J. Chem. Phys. **129**, 034106 (2008).
[33] F. Aquilante, L. Gagliardi, T. B. Pedersen, and R. Lindh, Atomic cholesky decompositions: A route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency, J. Chem. Phys. **130**, 154107 (2009).
[34] J. Boström, V. Veryazov, F. Aquilante, T. Bondo Pedersen, and R. Lindh, Analytical gradients of the second-order møller-plesset energy using cholesky decompositions, Int. J. Quantum Chem. **114**, 321 (2013).
[35] D. Thouless, Stability conditions and nuclear rotations in the hartree-fock theory, Nucl. Phys. **21**, 225 (1960).
[36] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, Solving strongly correlated electron models on a quantum computer, Phys. Rev. A **92**, 062318 (2015).
[37] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Experimental realization of any discrete unitary operator, Phys. Rev. Lett. **73**, 58 (1994).
[38] S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. L. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder, C. Song, A. V. Titov, I. S. Ufimtsev, L.-P. Wang, and T. J. Martínez, TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics, WIREs Comput. Mol. Sci. **11**, e1494 (2020).
[39] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande, OpenMM 7: Rapid development of high performance algorithms for molecular dynamics, PLOS Comput. Biol. **13**, e1005659 (2017).
[40] G.-L. R. Anselmetti, D. Wierichs, C. Gogolin, and R. M. Parrish, Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems, New J. Phys. **23**, 113010 (2021).
[41] B. O'Gorman, W. J. Huggins, E. G. Rieffel, and K. B. Whaley, Generalized swap networks for near-term quantum computing, arXiv preprint arXiv:1905.05118 10.48550/arxiv.1905.05118 (2019).
[42] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, J. Chem. Theory Comput. **15**, 311 (2018).
[43] N. Gidopoulos and A. Theophilou, Hartree-fock equations determining the optimum set of spin orbitals for the expansion of excited states, Philos. Mag. B **69**, 1067 (1994).
[44] N. Gidopoulos, P. Papaconstantinou, and E. Gross, Ensemble-{Hartree–Fock} scheme for excited states. the optimized effective potential method, Physica B **318**, 328 (2002).
[45] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters, J. Chem. Phys. **76**, 637 (1982).
[46] F. Dubnikova and A. Lifshitz, Isomerization of propylene oxide. quantum chemical calculations and kinetic modeling, J. Phys. Chem. A **104**, 4489 (2000).
[47] S. Crosson and K. Moffat, Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction, Proc. Natl. Acad. Sci. **98**, 2995 (2001).
[48] A. Warshel and M. Levitt, Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, J. Mol. Biol. **103**, 227 (1976).
[49] E. G. Hohenstein, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, An atomic orbital-based formulation of the complete active space self-consistent field method on graphical processing units, J. Chem. Phys. **142**, 224103 (2015).
[50] E. G. Hohenstein, M. E. F. Bouduban, C. Song, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, Analytic first derivatives of floating occupation molecular orbital-complete active space configuration interaction on graphical processing units, J. Chem. Phys. **143**, 014111 (2015).
[51] D. Mandal, T. Tahara, and S. R. Meech, Excited-state dynamics in the green fluorescent protein chromophore, The Journal of Physical Chemistry B **108**, 1102 (2003).
[52] E. G. Hohenstein, Analytic formulation of derivative coupling vectors for complete active space configuration interaction wavefunctions with floating occupation molecular orbitals, J. Chem. Phys. **145**, 174110 (2016).
[53] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, Quantum computation of electronic transitions using a variational quantum eigensolver, Phys. Rev. Lett. **122**, 230401 (2019).
[54] R. M. Parrish, G.-L. R. Anselmetti, and C. Gogolin, Analytical ground- and excited-state gradients for molecular electronic structure theory from hybrid Quantum/Classical methods 10.48550/arxiv.2110.05040 (2021).
[55] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, Optimal design for universal multiport interferometers, Optica **3**, 1460 (2016).
[56] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, Phys. Rev. Lett. **118**, 150503 (2017).
[57] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).
[58] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, *et al.*, Pennylane: Automatic differen-

tiation of hybrid quantum-classical computations, arXiv preprint arXiv:1811.04968 10.48550/arxiv.1811.04968 (2018).

[59] A. Mari, T. R. Bromley, and N. Killoran, Estimating the gradient and higher-order derivatives on quantum hardware, Phys. Rev. A **103**, 10.1103/physreva.103.012405 (2021).

[60] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, General parameter-shift rules for quantum gradients (2021), arXiv:2107.12390 [quant-ph].

Supporting Information

I. Problem Statement

In this section, we describe the scope of the work presented in this paper. We focus on the case of analytic nuclear gradients of variational energies determined from the explicit double factorization of the Hamiltonian but it will be apparent how the methods outlined here can be used to compute full relaxed active space one and two body reduced density matrices that can serve as input parameters for the computation of other quantities in the same way as is can be done in other active space methods.

A. Hamiltonian Factorization

The electronic Hamiltonian can be normal-ordered with respect to some closed-shell Slater determinant representing the frozen core electrons and written as:

$$\hat{H} = E_{\mathrm{c}} + \sum_{pq}(p|\hat{h}_{\mathrm{c}}|q)\hat{E}_{pq} + \frac{1}{2}\sum_{pqrs}(pq|rs)\left(\hat{E}_{pq}\hat{E}_{rs} - \delta_{qr}\hat{E}_{ps}\right), \quad (13)$$

where $\hat{E}_{pq}$ are the usual singlet excitation operators, E_{c} is the energy of the frozen core determinant, and $(p|\hat{h}_{\mathrm{c}}|q)$ are the matrix elements of the so-called frozen core operator,

$$(p|\hat{h}_{\mathrm{c}}|q) = (p|\hat{h}|q) + \sum_{i}^{\mathrm{core}}\left[2(pq|ii) - (pi|qi)\right]. \quad (14)$$

Explicit double factorization of the Hamiltonian yields an exact or approximate and compressed representation of such Hamiltonian by means of diagonalization of the integral tensors. First, consider the modified one-electron integrals

$$(p|\hat{\kappa}|q) = (p|\hat{h}_{\mathrm{c}}|q) - \frac{1}{2}\sum_{r}(pr|qr). \quad (15)$$

These can be trivially eigendecomposed as

$$(p|\hat{\kappa}|q) = \sum_{k} U^{\varnothing}_{pk} f^{\varnothing}_{k} U^{\varnothing}_{qk}. \quad (16)$$

The two-electron integrals can be factorized first by eigendecomposing the integral tensor

$$(pq|rs) = \sum_{t} V^{t}_{pq} g^{t} V^{t}_{rs}. \quad (17)$$

Subsequently, the eigenvectors can be decomposed for each "eigenleaf" (term in the above sum), which will be referred to as a t-leaf throughout the remainder of this work,

$$V^{t}_{pq} = \sum_{k} U^{t}_{pk}\lambda^{t}_{k}U^{t}_{qk}. \quad (18)$$

Combining these nested decompositions, we arrive at the double factorized form of the two-electron integrals.

$$(pq|rs) = \sum_{tkl} U^{t}_{pk}U^{t}_{qk}\underbrace{\lambda^{t}_{k}g^{t}\lambda^{t}_{l}}_{Z^{t}_{kl}}U^{t}_{rl}U^{t}_{sl} \quad (19)$$

The approach outlined above defines the explicit double factorization (X-DF) method. Other approaches of obtaining double factorized two-electron integral (19) have been developed, but we will concentrate on X-DF in the present work. X-DF can be either exact or approximate; this depends on whether or not all the t-leaves are retained or those with small eigenvalues (17) have been removed. Now, we can write the Hamiltonian in terms of these factorized one- and two-electron integrals as

$$\hat{H} = E_{\mathrm{c}} + \sum_{k} f^{\varnothing}_{k}\hat{E}_{kk}(\varnothing) + \frac{1}{2}\sum_{tkl} Z^{t}_{kl}\hat{E}_{kk}(t)\hat{E}_{ll}(t). \quad (20)$$

Notice that the dependence of $\hat{E}_{kk}(t)$ on the t-leaf specifies that these operators are applied in a rotated orbital basis defined by the eigendecompositions shown above.

In the present work, we make a few additional modifications to the Hamiltonian. If we consider the form of the $\hat{E}_{kk}$ operators within the Jordan-Wigner mapping,

$$\hat{E}_{kk} = \hat{I} - \frac{1}{2}\left(\hat{Z}_{k} + \hat{Z}_{\bar{k}}\right), \quad (21)$$

then we can define new zero- and one-body terms that include effective contributions from the one- and two-body terms in (20),

$$\mathcal{E} = E_{\mathrm{c}} + \sum_{p}(p|\hat{h}_{\mathrm{c}}|p) + \frac{1}{2}\sum_{pq}(pp|qq) - \frac{1}{4}\sum_{pq}(pq|pq), \quad (22)$$

$$\mathcal{F}_{pq} = (p|\hat{h}_{\mathrm{c}}|q) + \sum_{r}(pq|rr) - \frac{1}{2}\sum_{r}(pr|qr). \quad (23)$$

As before, the effective one-electron operator can be eigendecomposed as

$$\mathcal{F}^{\varnothing}_{pq} = \sum_{k} U^{\varnothing}_{pk}\mathcal{F}^{\varnothing}_{k}U^{\varnothing}_{qk}. \quad (24)$$

Now, the Hamiltonian can be written as

$$\hat{H} = \mathcal{E} - \frac{1}{2}\sum_{k}\mathcal{F}^{\varnothing}_{k}\left(\hat{Z}_{k} + \hat{Z}_{\bar{k}}\right) + \frac{1}{8}\sum_{tkl} Z^{t}_{kl} \times \left(\hat{Z}_{k}\hat{Z}_{l} - \hat{I}\delta_{kl} + \hat{Z}_{k}\hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}}\hat{Z}_{l} + \hat{Z}_{\bar{k}}\hat{Z}_{\bar{l}} - \hat{I}\delta_{\bar{k}\bar{l}}\right). \quad (25)$$

Here, we suppress the dependence of the Z_k operators on the orbital basis of the particular t-leaf on which they are acting in the presentation of the equation, but it remains as in (20). Throughout this work, we will use the Hamiltonian in the form presented in (25).

B. Electronic Wavefunction

In the present work, we will consider only wavefunctions that variationally minimize the energy such as those obtained from the variational quantum eigensolver (VQE). In this approach, the wavefunction is defined by a parameterized unitary operator acting on some reference state

$$|\Psi_{\mathrm{VQE}}\rangle = \hat{U}(\vec{\phi})|\Psi_0\rangle. \tag{26}$$

As the name of the method implies, the parameters, $\vec{\phi}$ are obtained by variationally minimizing the expectation value of the energy

$$E = \min_{\vec{\phi}} \langle\Psi_0|\hat{U}(\vec{\phi})^\dagger \hat{H}\hat{U}(\vec{\phi})|\Psi_0\rangle. \tag{27}$$

The key to our subsequent development of analytic derivatives of the energy is that the VQE energy is stationary with respect to its parameters,

$$\frac{\partial E}{\partial \vec{\phi}} = 0. \tag{28}$$

Therefore, we do not need to consider the response of the VQE wavefunction to the perturbation under consideration. We note that – although we do not consider such cases in the present work – the analytic derivatives we formulate could apply to other, non-variational methods provided that Lagrange multipliers can be added to make the energy stationary with respect to changes in its parameters, as was done in [54] for the case of multi state contracted VQE.

C. X-DF/VQE

The use of the double factorized Hamiltonian imparts a particular structure on the VQE density matrices. First, consider the usual molecular orbitals,

$$\phi_p(\mathbf{r}) = \sum_\mu C_{\mu p}\phi_\mu(\mathbf{r}). \tag{29}$$

Here, some some set of atomic orbitals ($\{|\phi_\mu\rangle\}$) are transformed by the molecular orbital coefficient matrix, $C_{\mu p}$. The double factorized Hamiltonian introduces additional orbital bases – one for each t-leaf,

$$\phi^t_k(\mathbf{r}) = \sum_p U^t_{pk}\phi_p(\mathbf{r}). \tag{30}$$

The application of the Hamiltonian requires a series of transformations between these new orbital bases. These transformations are performed by a parameterized "Givens" operator, $\hat{G}(\{\theta^t_g\})$. Any unitary matrix (of rank N) can be represented as a product of $N(N-1)/2$ Givens rotations. This can be done in a number of different ways, resulting in different circuit layouts and depths. This fact is used to determine the set of angles, $\{\theta^t_g\}$, that define each Givens operator. The angles are chosen such that there is a unique Givens operator for each t-leaf,

$$\left[\prod_g \hat{G}_g(\theta^t_g)\right]^t_{pk} = U^t_{pk}. \tag{31}$$

The classical matrix elements of the X-DF Hamiltonian can be obtained from transformations using the U^t_{pk} matrices, while transformations of the state vector on the quantum computer require the Givens operator to be applied as a series of Givens circuits. Now we can write the Hamiltonian in its complete form,

$$\begin{aligned}\hat{H} = \mathcal{E} &- \frac{1}{2}\sum_k \mathcal{F}^\varnothing_k \hat{G}^\dagger(\theta^\varnothing_g)\left(\hat{Z}_k + \hat{Z}_{\bar{k}}\right)\hat{G}(\theta^\varnothing_g) \\ &+ \frac{1}{8}\sum_{tkl} Z^t_{kl} \times \hat{G}^\dagger(\theta^t_g)\big(\hat{Z}_k\hat{Z}_l - \hat{I}\delta_{kl} + \hat{Z}_k\hat{Z}_{\bar{l}} \\ &+ \hat{Z}_{\bar{k}}\hat{Z}_l + \hat{Z}_{\bar{k}}\hat{Z}_{\bar{l}} - \hat{I}\delta_{\bar{k}\bar{l}}\big)\hat{G}(\theta^t_g),\end{aligned} \tag{32}$$

where the transformations to the orbital eigenbases of each t-leaf have been made explicit.

In methods based on the usual electronic Hamiltonian, the energy can be obtained from the one- and two-body density matrices in the molecular orbital basis,

$$E = E_{\mathrm{c}} + \sum_{pq}(p|\hat{h}_{\mathrm{c}}|q)\gamma_{pq} + \sum_{pqrs}(pq|rs)\Gamma_{pqrs}. \tag{33}$$

Here, the molecular orbital-based density matrices have their usual definition,

$$\gamma_{pq} = \langle\Psi|\hat{E}_{pq}|\Psi\rangle \tag{34}$$

$$\Gamma_{pqrs} = \frac{1}{2}\langle\Psi|\hat{E}_{pq}\hat{E}_{rs} - \delta_{qr}\hat{E}_{ps}|\Psi\rangle. \tag{35}$$

While it is certainly possible to form the molecular orbital density matrices from the X-DF Hamiltonian, it is more advantageous to work in the natural eigenbasis representation of each t-leaf,

$$E = \mathcal{E} + \sum_k \mathcal{F}^\varnothing_k \omega^\varnothing_k + \sum_{tkl} Z^t_{kl}\omega^t_{kl}. \tag{36}$$

Here, the density matrices are defined as

$$\omega^\varnothing_k = -\frac{1}{2}\langle\Psi|\hat{G}^\dagger(\{\theta^\varnothing_g\})\left(\hat{Z}_k + \hat{Z}_{\bar{k}}\right)\hat{G}(\{\theta^\varnothing_g\})|\Psi\rangle, \tag{37}$$

and

$$\omega_{kl}^t = \frac{1}{8}\langle\Psi|\hat{G}^\dagger(\{\theta_g^t\})\big(\hat{Z}_k\hat{Z}_l - \hat{I}\delta_{kl} + \hat{Z}_k\hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}}\hat{Z}_l + \hat{Z}_{\bar{k}}\hat{Z}_{\bar{l}} - \hat{I}\delta_{\bar{k}\bar{l}}\big)\hat{G}(\{\theta_g^t\})|\Psi\rangle. \quad (38)$$

In the remainder of this work, we will work exclusively in terms of these eigenbasis density matrices.

D. Nuclear Gradients

The main focus of this work is the evaluation of analytic derivatives of the energy with respect to the Cartesian positions of the nuclei. The derivation is specialized for the case of atom-centered Gaussian basis functions. This means that the position of the basis functions is tied to the position of the nuclei. As a result, derivatives of integrals over operators with no dependence on the nuclei (electronic kinetic or overlap operators, for example) will still have dependence on the nuclear positions. Derivatives of the overlap integrals in the atomic orbital basis provide a representative example,

$$\frac{\mathrm{d}}{\mathrm{d}R}(\mu|\nu) = (\frac{\mathrm{d}}{\mathrm{d}R}\mu|\nu) + (\mu|\frac{\mathrm{d}}{\mathrm{d}R}\nu). \quad (39)$$

As a consequence of these derivatives surviving, straightforward application of the Hellmann-Feynman theorem is impossible within Gaussian orbital formulations of electronic structure theory unless a complete (i.e., infinite) basis is applied.

E. Fractional Occupation Number Hartree-Fock

In this work, we obtain the molecular orbitals that underly the X-DF/VQE energies from fractional occupation number Hartree-Fock (FON-HF) [43, 44], which includes standard HF as a special case. This method introduces temperature dependent fractional occupations of the orbitals. Commonly, a Fermi-Dirac distribution is used to obtain the fractional occupations,

$$n_i = \frac{2}{1 + \exp[(\epsilon_i - \epsilon_f)/kT]}, \quad (40)$$

where ϵ_i and n_i are the orbital energy and fraction occupation of orbital i, respectively. The Fermi energy, ϵ_f, is chosen such that the fractional occupations conserve the total number of electrons,

$$N = \sum_i n_i. \quad (41)$$

It is also possible to obtain the fractional occupations from a Gaussian broadening of the orbital energy levels or other distributions.

The motivation for using FON-HF orbitals is that they often provide a good approximation to variationally optimized orbitals while being nearly identical in computational cost to Hartree-Fock (HF) orbitals. Canonical HF orbitals are a poor choice for strongly correlated systems, because they suffer from bias towards the single determinant for which they are optimized. In extreme cases, HF orbitals can artificially break degeneracies that exist in the system (for example, exact degeneracies between the highest occupied and lowest unoccupied orbital cannot be described). FON-HF, by contrast, does not suffer from these issues; in the case of exact degeneracies, FON-HF will simply place equal fractional occupations in each of the degenerate orbitals. A final note is that although we focus on the choice of FON-HF orbitals in this work, the analytic derivative expressions we provide can be trivially extended to most other definitions of molecular orbitals, provided they are fully decoupled from the VQE energy.

II. Inapplicability of the Hellmann-Feynman Theorem

At first glance, analytic derivatives of the energy appear trivial to evaluate in light of the Hellmann-Feynman theorem, which states that the derivative of the energy is simply the expectation value of the derivative Hamiltonian,

$$\frac{\mathrm{d}E}{\mathrm{d}R} = \langle\Psi|\frac{\mathrm{d}\hat{H}}{\mathrm{d}R}|\Psi\rangle. \quad (42)$$

However, upon closer examination, it becomes clear that the Hellmann-Feynman theorem rarely has practical relevance. First, the Hellmann-Feynman derivative does not include all contributions from atom-centered Gaussian functions, in particular, the derivatives of the basis function overlap that are necessary to maintain orthonormality in the molecular orbitals. Second, the Hellmann-Feynman theorem assumes a fully variational wavefunction (i.e., one in which the energy is minimized with respect to all parameters). In the present work, we consider fully variational VQE ansätze, but do not minimize the VQE energy with respect to the molecular orbitals. As a result it is essential to include the response of the molecular orbitals to the perturbation in the analytic derivative. Here, we obtain the molecular orbitals from FON-HF, which minimize the Mermin free energy – not the VQE energy. As a result, orbital response contributions to the analytic gradient beyond the Hellmann-Feynman theorem must be considered. We should note that these considerations are common to many electronic structure methods and are in no way unique to X-DF/VQE.

A more interesting point is that the Hellmann-Feynman theorem does not apply for X-DF/VQE in places where it usually would be exact. Take, for example, a case where the VQE ansatz covers all available variational free parameters. Consider also a perturbation that does not change the underlying Gaussian basis functions – perhaps an electric field perturbation. In that case, the Hellmann-Feynman theorem would hold and the derivative of the energy is equivalent to the one-particle density matrix traced against the one-body electric field

derivative integrals (i.e., dipole integrals). Here, for a perturbation in the x direction,

$$\frac{\mathrm{d}E}{\mathrm{d}\varepsilon_x} = \sum_{pq} \gamma_{pq} (p|\hat{\mu}_x|q). \tag{43}$$

However, in the X-DF/VQE case, we do not assume that the full density matrices are available, only the eigenbasis density matrices. As a result, application of the Hellmann-Feynman in the case analogous to what is shown above is no longer exact,

$$\frac{\mathrm{d}E}{\mathrm{d}\varepsilon_x} \neq \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\varepsilon_x} + \sum_k \frac{\mathrm{d}\mathcal{F}_k^{\varnothing}}{\mathrm{d}\varepsilon_x} \omega_k^{\varnothing}. \tag{44}$$

The loss of exactness can be traced to the representation of the density matrix in the eigenbasis. This representation of the density is a projection of the full density to the minimal amount required to recover the energy. This representation of the density alone is insufficient in general for obtaining properties other than the energy.

To illustrate how the projection of the density affects property computations, consider a general example,

$$o = \sum_{pq} A_{pq} B_{pq}. \tag{45}$$

Now, if we eigendecompose matrix $\mathbf{A}$,

$$A_{pq} = \sum_k U_{pk} a_k U_{qk}, \tag{46}$$

we can write the original summation in the eigenbasis of $\mathbf{A}$ as

$$o = \sum_k a_k \underbrace{\left[\sum_{pq} U_{pk} B_{pq} U_{qk} \right]}_{B_{kk}}. \tag{47}$$

Notice that we now have a dot product between the eigenvalues of $\mathbf{A}$ and the diagonal elements of $\mathbf{B}$ after transformation to the eigenbasis of $\mathbf{A}$. Note that unless $[\mathbf{A}, \mathbf{B}] = 0$, $\mathbf{B}$ is not diagonal in the eigenbasis of $\mathbf{A}$. Now, if we transform back to the original basis using (46) and $\tilde{B}_{pq} = \sum_k U_{pk} B_{kk} U_{qk}$, we find

$$o = \sum_{pq} A_{pq} \tilde{B}_{pq}. \tag{48}$$

In this case, $\mathbf{B} \neq \tilde{\mathbf{B}}$ and we have shown that only a projection of the matrix $\mathbf{B}$ is required to maintain equality. If we consider a second matrix, $\mathbf{C}$, where $[\mathbf{A}, \mathbf{C}] \neq 0$, following the same logic we find that

$$\sum_{pq} C_{pq} B_{pq} \neq \sum_{pq} C_{pq} \tilde{B}_{pq}. \tag{49}$$

This is because the projection of $\mathbf{B}$ is lossy.

What we have shown in general above is what we must contend with when we work in terms of the eigenbasis density matrices (i.e., (36)). We only retained enough information in the density matrices to compute the energy. The remainder of the information that is needed must be obtained by other means and it turns out that taking derivatives with respect to the parameters of the Givens operators is a suitable approach. Another way of stating this problem is that by writing the energy in terms of only the diagonal elements of the density matrix, we have lost some of the usual orbital invariances.

III. Lagrangian Formulation of the Gradient

We must consider the response of several non-variational parameters introduced in the X-DF method in order to evaluate the analytic derivative of the energy. These parameters are the $U_{pk}^{\varnothing}$, U_{pk}^{t} and V_{pq}^{t} tensors that define the X-DF matrix elements. In addition, the X-DF/VQE derivative also depends on $\theta_g^{\varnothing}$ and θ_g^t, which parameterize the Givens operators used to transform to the eigenbasis of each t-leaf. Finally, the derivative of the energy also depends on the usual quantities such as the underlying molecular orbitals and the one- and two-electron integrals in the atomic orbital basis.

Treating the response of these non-variational parameters to the perturbation in an efficient manner is essential to the development of a formulation of the analytic derivative that can be applied to practical chemical problems. From the chain rule, it appears that the response of all of these free parameters to the perturbation is required to evaluate the derivative of the energy,

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}R} = \frac{\partial E}{\partial R} + \frac{\partial E}{\partial U_{pk}^{\varnothing}} \frac{\mathrm{d}U_{pk}^{\varnothing}}{\mathrm{d}R} + \frac{\partial E}{\partial U_{pk}^{t}} \frac{\mathrm{d}U_{pk}^{t}}{\mathrm{d}R} + \frac{\partial E}{\partial V_{pq}^{t}} \frac{\mathrm{d}V_{pq}^{t}}{\mathrm{d}R} \\ + \frac{\partial E}{\partial \theta_g^{\varnothing}} \frac{\mathrm{d}\theta_g^{\varnothing}}{\mathrm{d}R} + \frac{\partial E}{\partial \theta_g^{t}} \frac{\mathrm{d}\theta_g^{t}}{\mathrm{d}R} + \frac{\partial E}{\partial C_{\mu p}} \frac{\mathrm{d}C_{\mu p}}{\mathrm{d}R}.\end{aligned} \tag{50}$$

While direct evaluation of this equation will produce the desired result, the computational cost will scale with the number of perturbations. There are several approaches to simplifying such equations to remove the explicit perturbation dependence that go by several names. Perhaps the most systematic approach to removing perturbation dependence is to construct a Lagrangian. The value of the Lagrangian is equal to the energy (the quantity we wish to differentiate) and it contains undetermined multipliers which allow partial derivatives with respect to the non-variational parameters to be set to zero. That is, we enforce

$$\begin{aligned}\mathcal{L} = E, \quad \frac{\partial \mathcal{L}}{\partial U_{pk}^{\varnothing}} = 0, \quad \frac{\partial \mathcal{L}}{\partial U_{pk}^{t}} = 0, \quad \frac{\partial \mathcal{L}}{\partial V_{pq}^{t}} = 0, \\ \frac{\partial \mathcal{L}}{\partial \theta_g^{\varnothing}} = 0, \quad \frac{\partial \mathcal{L}}{\partial \theta_g^{t}} = 0, \quad \frac{\partial \mathcal{L}}{\partial C_{\mu p}} = 0.\end{aligned} \tag{51}$$

Such a Lagrangian can be constructed by adding constraints that are satisfied by the definition of each of these non-variational quantities. The complete Lagrangian we will use to define the analytic gradient is

$$\begin{aligned}\mathcal{L} = \mathcal{E} &+ \sum_k \mathcal{F}^{\varnothing}_k \omega^{\varnothing}_k + \sum_t^T \sum_{kl} Z^t_{kl} \omega^t_{kl} \\ &+ \sum_{p>k} \eta^{\varnothing}_{pk} \left[\left(\prod_g \hat{G}_g(\theta^{\varnothing}_g) \right)_{pk} - U^{\varnothing}_{pk} \right] \\ &+ \sum_t^T \sum_{p>k} \eta^t_{pk} \left[\left(\prod_g \hat{G}_g(\theta^t_g) \right)_{pk} - U^t_{pk} \right] \\ &+ \sum_{k>k'} \mu^{\varnothing}_{kk'} \mathcal{F}^{\varnothing}_{kk'} + \sum_t^T \sum_{k>k'} \mu^t_{kk'} \lambda^t_{kk'} + \sum_{t>t'} \nu^{tt'} g^{tt'} \\ &+ \sum_{p>q} z_{pq} f_{pq} + \sum_{p \leq q} x_{pq} [\delta_{pq} - (p|q)]. \end{aligned} \tag{52}$$

Here, we have introduced seven sets of undetermined multipliers. The η multipliers use the definition of the Givens operators (31) to constrain the values of $\{\theta\}$. The μ multipliers use the representation of off-diagonal elements of $\mathcal{F}^{\varnothing}$ and λ^t in their eigenbases to constrain $U^{\varnothing}$ and U^t to be the eigenvectors of these matrices. We use the ν multipliers to constrain V to be the eigenvectors of the two-electron integrals, by using the off-diagonal elements of the two-electron integrals in its eigenbasis as a constraint. Here, we use T to denote where the factorization of the two-electron integrals has been truncated. The final two multipliers, z and x, constrain the convergence condition and orthonormality of the underlying molecular orbitals. Here, our constrain is that the Fock matrix, f_{pq}, is diagonal and that the overlap of the molecular orbitals, $(p|q)$, is an identity matrix. These constraints and equations defining the undetermined multipliers will be discussed in later sections of this document.

A. Constraints on the Givens Operator

Let us say we have an antisymmetric matrix of size N,

$$X_{rs} = -X_{sr} \;\forall\; s,\; r. \tag{53}$$

Through the exponential map, this generates a special orthogonal matrix,

$$O_{pq} \equiv \left[\exp(\hat{X}) \right]_{pq} \in \mathcal{SO}(N). \tag{54}$$

Through a Givens decomposition, we can define the special orthogonal matrix U_{pk}. The pivot structure of the Givens matrices, e.g., rectangle [55], diamond [12], or otherwise, can be chosen freely and is considered to be fixed,

$$O_{pq}[\{\theta_g\}] \equiv \left[\prod_g \hat{G}_g(\theta_g) \right]_{pq} \in \mathcal{SO}(N) \tag{55}$$

Note that here $\hat{G}_g(\theta_g)$ is the one-body image of the g-th Givens operator.

We are able to use this to construct Givens decompositions that represent $U^{\varnothing}_{pk}$ and U^t_{pk}.

$$\left[\prod_g \hat{G}_g(\theta^{\varnothing}_g) \right]^{\varnothing}_{pk} = U^{\varnothing}_{pk} \tag{56}$$

$$\left[\prod_g \hat{G}_g(\theta^t_g) \right]^{t}_{pk} = U^t_{pk} \tag{57}$$

These equalities form the basis of the constraints that enter the X-DF/VQE Lagrangian (52). To solve for the undetermined η multipliers, we set the partial derivatives of the Lagrangian with respect to θ to zero. That is, the equations for $\eta^{\varnothing}_{pk}$ are defined by $\partial\mathcal{L}/\partial\theta^{\varnothing}_g = 0$ and the equations for η^t_{pk} are defined by $\partial\mathcal{L}/\partial\theta^t_g = 0$.

In the Lagrangian, only the energy (36) and the constraint terms themselves carry explicit dependence on $\theta^{\varnothing}_g$ and θ^t_g. Therefore, solving for the $\eta^{\varnothing}_{pk}$ and η^t_{pk} undetermined multipliers is the first step in making the Lagrangian stationary. Equations for the Lagrange multipliers are determined by,

$$\sum_{p>k} \eta^{\varnothing}_{pk} \underbrace{\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}\theta_{g'}} \left[\prod_g \hat{G}_g(\theta^{\varnothing}_g) \right]_{pk}}_{A^{\varnothing}_{g',pk}} = -\frac{\mathrm{d}E}{\mathrm{d}\theta^{\varnothing}_{g'}} \tag{58}$$

and

$$\sum_{p>k} \eta^t_{pk} \underbrace{\frac{\mathrm{d}\mathcal{O}}{\mathrm{d}\theta_{g'}} \left[\prod_g \hat{G}_g(\theta^t_g) \right]_{pk}}_{A^t_{g',pk}} = -\frac{\mathrm{d}E}{\mathrm{d}\theta^t_{g'}} \tag{59}$$

The matrices $A^{\varnothing}_{g,pk}$ and $A^t_{g,pk}$ are square of dimension $N(N-1)/2$ and can be inverted to obtain $\eta^{\varnothing}_{pk}$ and η^t_{pk}. Note that we treat each t-leaf separately, thus treating $A^t_{g,pk}$ as a stack of matrices rather than a third-order tensor. In practice, we find that these equations can be accurately solved by pseudoinverting the $\mathbf{A}$ matrices as they may become singular in cases with high symmetry depending on orbital ordering and the pivot structure of the givens decomposition.

B. Constraints on the X-DF Eigenbases

In the X-DF/VQE Lagrangian (52), the transformations to the eigenbases of the one- and two-body operators (i.e., the $\mathbf{U}$ matrices) can be constrained by using the definition of the matrices they diagonalize. That is, $U^{\varnothing}_{pk}$ contains the eigenvectors of $\mathcal{F}_{pq}$. Therefore, we can

use the fact that, when tranformed to its eigenbasis, the off-diagonal matrix elements of this matrix are zero (viz. $\mathcal{F}^{\varnothing}_{kk'} = 0 \; \forall \; k \neq k'$). Similarly, this holds for each t-leaf of the two-body operator. In this case, U^t_{pk} are the eigenvectors of V^t_{pq} with eigenvalues λ^t_k. If we define a new quantity,

$$\lambda^t_{kk'} = \sum_{pq} U^t_{pk} V^t_{pq} U^t_{qk'}, \tag{60}$$

trivially, we can see that $\lambda^t_{kk'} = \lambda^t_k \delta_{kk'}$. However, this eigenbasis representation of V^t_{pq} is useful to define the constraint on U^t_{pk}. We use the $\mu^{\varnothing}_{kk'}$ and $\mu^t_{kk'}$ undetermined multipliers to set the partial derivatives of the Lagrangian with respect to $U^{\varnothing}_{pk}$ and U^t_{pk}, respectively, to zero. Here, one should note that the energy, the η constraint terms and the μ constraint terms carry explicit dependence on the $\mathbf{U}$ matrices. Therefore, solving for the μ undetermined multipliers is the next step after the η multipliers have been formed.

We will begin by deriving programmable equations for $\mu^{\varnothing}_{kk'}$ that make the Lagrangian stationary with respect to variation in $U^{\varnothing}_{pk}$. To accomplish this, we will set partial derivatives of the Lagrangian with respect to $U^{\varnothing}_{pk}$ to zero and transformed the resulting derivative by $U^{\varnothing}_{pk}$ to produce a result that is fully in the eigenbasis of $\mathcal{F}_{pq}$. The transformation by $U^{\varnothing}_{pk}$ is not formally required, but will simplify the resulting equations.

$$\begin{aligned} 0 = &\sum_r U^{\varnothing}_{rl'} \frac{\partial \mathcal{L}}{\partial U^{\varnothing}_{rl}} = \sum_r U^{\varnothing}_{rl'} \frac{\partial}{\partial U^{\varnothing}_{rl}} \sum_k \mathcal{F}^{\varnothing}_k \omega^{\varnothing}_k \\ &+ \sum_r U^{\varnothing}_{rl'} \frac{\partial}{\partial U^{\varnothing}_{rl}} \sum_{p>k} \eta^{\varnothing}_{pk} \left[\left(\prod_g \hat{G}^{\varnothing}(\theta^{\varnothing}_g) \right)_{pk} - U^{\varnothing}_{pk} \right] \\ &+ \sum_r U^{\varnothing}_{rl'} \frac{\partial}{\partial U^{\varnothing}_{rl}} \sum_{k>k'} \mu^{\varnothing}_{kk'} \mathcal{F}^{\varnothing}_{kk'} \end{aligned} \tag{61}$$

Now, we can rewrite the Lagrangian to expose the dependence on $U^{\varnothing}_{pk}$.

$$\begin{aligned} 0 = &\sum_r U^{\varnothing}_{rl'} \frac{\partial}{\partial U^{\varnothing}_{rl}} \sum_k \sum_{pq} U^{\varnothing}_{pk} U^{\varnothing}_{qk} \mathcal{F}_{pq} \omega^{\varnothing}_k \\ &+ \sum_r U^{\varnothing}_{rl'} \frac{\partial}{\partial U^{\varnothing}_{rl}} \sum_{p>k} \eta^{\varnothing}_{pk} \left[\left(\prod_g \hat{G}^{\varnothing}(\theta^{\varnothing}_g) \right)_{pk} - U^{\varnothing}_{pk} \right] \\ &+ \sum_r U^{\varnothing}_{rl'} \frac{\partial}{\partial U^{\varnothing}_{rl}} \sum_{k>k'} \sum_{pq} \mu^{\varnothing}_{kk'} U^{\varnothing}_{pk} U^{\varnothing}_{qk'} \mathcal{F}^{\varnothing}_{pq} \end{aligned} \tag{62}$$

Taking the derivatives and transforming, to the eigenbasis we get:

$$\begin{aligned} 0 = &2\delta_{ll'} \mathcal{F}^{\varnothing}_l \omega^{\varnothing}_l - \eta^{\varnothing}_{l'l} \\ &+ \begin{cases} \mu^{\varnothing}_{ll'} \mathcal{F}^{\varnothing}_l & \text{if } \; l > l' \\ \mu^{\varnothing}_{l'l} \mathcal{F}^{\varnothing}_l & \text{if } \; l' > l \end{cases}. \end{aligned} \tag{63}$$

We have transformed the molecular orbital index of the $\eta^{\varnothing}_{pk}$ multipliers to the eigenbasis.

$$\eta^{\varnothing}_{kk'} = \sum_p U^{\varnothing}_{pk} \eta^{\varnothing}_{pk'} \tag{64}$$

Here, we have treated $\eta^{\varnothing}_{pk}$ as a square matrix containing the values of the multipliers defined by Eqn. 58 in the lower triangle and zeros elsewhere. Notice that we have a full matrix representing the derivatives of the Lagrangian, while we only have a lower triangular matrix of undetermined multipliers. We can define the multipliers, $\mu^{\varnothing}_{ll'}$, by equating the upper and lower triangles of the Lagrangian derivatives – both of which must equal zero.

$$\begin{aligned} 2\delta_{ll'} \mathcal{F}^{\varnothing}_l \omega^{\varnothing}_l - \eta^{\varnothing}_{l'l} + \mu^{\varnothing}_{ll'} \mathcal{F}^{\varnothing}_l = \\ 2\delta_{ll'} \mathcal{F}^{\varnothing}_l \omega^{\varnothing}_l - \eta^{\varnothing}_{ll'} + \mu^{\varnothing}_{ll'} \mathcal{F}^{\varnothing}_{l'} \end{aligned} \tag{65}$$

When we solve for the multipliers, the contributions from the eigenbasis density matrix cancel and only the $\eta^{\varnothing}_{ll'}$ contribute to the right-hand side of the system of linear equations defining $\mu^{\varnothing}_{ll'}$.

$$\mu^{\varnothing}_{ll'} = \frac{\eta^{\varnothing}_{ll'} - \eta^{\varnothing}_{l'l}}{\mathcal{F}^{\varnothing}_{l'} - \mathcal{F}^{\varnothing}_l} \tag{66}$$

With the $\mu^{\varnothing}_{ll'}$ multipliers defined, the Lagrangian is stationary with respect to variation in $U^{\varnothing}_{pk}$.

Following along similar lines, we can derive an expression for $\mu^t_{kk'}$ to make the Lagrangian stationary with respect to variation in U^t_{pk}. Again, we start by setting derivatives of the Lagrangian with respect to U^t_{pk} to zero.

$$\begin{aligned} 0 = &\sum_r U^{t'}_{rm'} \frac{\partial \mathcal{L}}{\partial U^{t'}_{rm}} = \sum_r U^{t'}_{rm'} \frac{\partial}{\partial U^{t'}_{rm}} \sum_t^T \sum_{kl} Z^t_{kl} \omega^t_{kl} \\ &+ \sum_r U^{t'}_{rm'} \frac{\partial}{\partial U^{t'}_{rm}} \sum_t^T \sum_{p>k} \eta^t_{pk} \left[\left(\prod_g \hat{G}^t(\theta^t_g) \right)_{pk} - U^t_{pk} \right] \\ &+ \sum_r U^{t'}_{rm'} \frac{\partial}{\partial U^{t'}_{rm}} \sum_t^T \sum_{k>k'} \mu^t_{kk'} \lambda^t_{kk'} \end{aligned} \tag{67}$$

Expanding to show explicitly the dependence on U^t_{pk}, we get:

$$\begin{aligned} 0 = &\sum_r U^{t'}_{rm'} \frac{\partial}{\partial U^{t'}_{rm}} \Bigg(\\ &\sum_t^T \sum_{kl} \sum_{pp'qq'} U^t_{pk} U^t_{p'k} V^t_{pp'} g^t V^t_{qq'} U^t_{ql} U^t_{q'l} \omega^t_{kl} \\ &+ \sum_t^T \sum_{p>k} \eta^t_{pk} \left[\left(\prod_g \hat{G}^t(\theta^t_g) \right)_{pk} - U^t_{pk} \right] \\ &+ \sum_t^T \sum_{k>k'} \sum_{pp'} \mu^t_{kk'} U^t_{pk} U^t_{p'k'} V^t_{pp'} \Bigg). \end{aligned} \tag{68}$$

At this point, an important observation to make is that we only need to take derivatives corresponding to U^t_{pk} with $t \leq T$; derivatives with respect to U^t_{pk} with $t > T$ are trivially zero. Now, we can take the partial derivatives and transform the result back to the eigenbasis.

$$\begin{aligned} 0 =& 4\delta_{mm'} \sum_k Z^{t'}_{km} \omega^{t'}_{km} - \eta^{t'}_{m'm} \\ &+ \begin{cases} \mu^{t'}_{mm'} \lambda^{t'}_m & \text{if } m > m' \\ \mu^{t'}_{m'm} \lambda^{t'}_m & \text{if } m' > m \end{cases}. \end{aligned} \tag{69}$$

As we did previously, we will transform the $\eta^t_{pk'}$ multipliers into the eigenbasis.

$$\eta^t_{kk'} = \sum_p U^t_{pk} \eta^t_{pk'} \tag{70}$$

Again, the $\mu^t_{mm'}$ multipliers are defined by a system of linear equations.

$$\begin{aligned} 4\delta_{ll'} \sum_k Z^t_{kl} \omega^t_{kl} - \eta^t_{l'l} + \mu^t_{ll'} \lambda^t_l = \\ 4\delta_{ll'} \sum_k Z^t_{kl} \omega^t_{kl} - \eta^t_{ll'} + \mu^t_{ll'} \lambda^t_{l'} \end{aligned} \tag{71}$$

These linear equations have a simple analytic solution.

$$\mu^t_{ll'} = \frac{\eta^t_{ll'} - \eta^t_{l'l}}{\lambda^t_{l'} - \lambda^t_l} \tag{72}$$

Solving for these multipliers makes the Lagrangian stationary with respect to variation in U^t_{pk}. In both cases ($U^{\varnothing}_{pk}$ and U^t_{pk}), the multipliers are defined separately for each t-leaf.

C. Constraints on the Two-Electron Integral Eigendecomposition

The $\nu^{tt'}$ multipliers depend on derivatives of the two-body contribution to the energy as well as the μ^t_{pk} multipliers. These multipliers couple the t-leaves together and, in the case of truncated X-DF they account for rotations between t-leaves included in the definition of the Hamiltonian and those that were excluded. These multipliers make the Lagrangian stationary with respect to variation in the eigenvectors of the two-electron integrals, V^t_{pq}. To begin, we will set the derivatives of the Lagrangian with respect to V^t_{pq} to zero.

$$\begin{aligned} 0 = \sum_{rr'} V^{u'}_{rr'} \frac{\partial \mathcal{L}}{\partial V^u_{rr'}} = \sum_{rr'} V^{u'}_{rr'} \frac{\partial}{\partial V^u_{rr'}} \Bigg(\\ \sum_t^T \sum_{kl} Z^t_{kl} \omega^t_{kl} + \sum_t^T \sum_{k>k'} \mu^t_{kk'} \lambda^t_{kk'} + \sum_{t>t'} \nu^{tt'} g^{tt'} \Bigg) \end{aligned} \tag{73}$$

We must fully expand these terms to expose the explicit dependence on V^t_{pq}. Consider $\lambda^t_{kk'}$

$$\lambda^t_{kk'} = \sum_{pp'} U^t_{pk} V^t_{pp'} U^t_{p'k'} \tag{74}$$

and g^t

$$g^t = \sum_{pp'qq'} V^t_{pp'} (pp'|qq') V^t_{qq'}. \tag{75}$$

Some useful partial derivatives are:

$$\sum_{rr'} V^{u'}_{rr'} \frac{\partial \lambda^t_{kk'}}{\partial V^u_{rr'}} = \delta_{tu} \sum_{rr'} V^{u'}_{rr'} U^t_{rk} U^t_{r'k'} \tag{76}$$

and

$$\sum_{rr'} V^{u'}_{rr'} \frac{\partial g^t}{\partial V^u_{rr'}} = 2\delta_{tu} g^{tu'} \tag{77}$$

These are combined in the definition of Z^t_{kl}

$$Z^t_{kl} = \sum_{pp'qq'} \sum_{rr'ss'} U^t_{pk} V^t_{pp'} U^t_{p'k} V^t_{rr'} (rr'|ss') V^t_{ss'} U^t_{ql} V^t_{qq'} U^t_{q'l}, \tag{78}$$

and its partial derivatives.

$$\begin{aligned} \sum_{rr'} V^{u'}_{rr'} \frac{\partial Z^t_{kl}}{\partial V^u_{rr'}} = 2\delta_{tu} \lambda^t_k g^{tu'} \lambda^t_l \\ + \delta_{tu} \sum_{rr'} \Big[U^t_{rk} V^{u'}_{rr'} U^t_{r'k} g^t \lambda^t_l + \lambda^t_k g^t U^t_{rl} V^{u'}_{rr'} U^t_{r'l} \Big] \end{aligned} \tag{79}$$

Now, we can define a useful intermediate.

$$\begin{aligned} R^{u'u} &= \sum_{rr'} V^{u'}_{rr'} \frac{\partial}{\partial V^u_{rr'}} \sum_t^T \Big[\sum_{kl} Z^t_{kl} \omega^t_{kl} + \sum_{k>k'} \mu^t_{kk'} \lambda^t_{kk'} \Big] \\ &= 2 \sum_{rr'} \sum_{kl} U^u_{rk} V^{u'}_{rr'} U^u_{r'k} g^u \lambda^u_l \omega^u_{kl} \\ &+ \sum_{k>k'} \sum_{rr'} \mu^u_{kk'} U^u_{rk} U^u_{r'k'} V^{u'}_{rr'} \quad \forall \quad u \leq T \end{aligned} \tag{80}$$

$$R^{u'u} = 0 \quad \forall \quad u > T \tag{81}$$

From here, the $\nu^{tt'}$ multipliers are defined by the eigenvector response of V^t_{pq} (as was the case for the $\mu^{\varnothing}_{kk'}$ and $\mu^t_{kk'}$ multipliers). As a result, the $\nu^{tt'}$ multipliers have a simple analytic form.

$$\nu^{tt'} = \frac{R_{tt'} - R_{t't}}{g^{t'} - g^t} \tag{82}$$

Once these multipliers are determined, the Lagrangian has been made stationary with respect to V^t_{pq}. At this point, the Lagrangian is stationary with respect to variation in any of the quantities defining the X-DF.

D. Constraints on the Molecular Orbitals

The final step before the derivative of the energy can be evaluated is to make the Lagrangian stationary with respect to variation in the underlying molecular orbitals (in this case, those determined from FON-HF). Techniques to evaluate the derivative of configuration interaction (CI) energies constructed from FON-HF orbitals are well-known [50] and we will not repeat the derivation here. The key to leveraging these existing approaches in the context of X-DF/VQE is to construct relaxed one- and two-body reduced density matrices within the active space. These matrices are analogous to those obtained from, for example, a complete active space CI wavefunction. These density matrices are used to define the right-hand side of the coupled-perturbed FON-HF (CP-FON-HF) equations (i.e., those that define the z_{pq} multipliers). Subsequently, a combination of the density matrices and the z_{pq} multipliers are used to define the x_{pq} multipliers. Once these multipliers are determined, the Lagrangian has been made stationary with respect to all nonvariational parameters, and the derivative of the energy can be evaluated.

An operational definition of the one- and two-body density matrices can be obtained by differentiating the stationary Lagrangian with respect to the one- and two-body integral matrices. In what follows, it is convenient to define an identity matrix with a rank equal to the number of active orbitals,

$$I_{pq} = \delta_{pq}. \tag{83}$$

We will define the one-body density matrix as the derivative of the Lagrangian (which has been made stationary with respect to all the X-DF non-variational parameters: $\theta_g^{\varnothing}$, θ_g^t, $U_{pk}^{\varnothing}$, U_{pk}^t, V_{pq}^t) with respect to the frozen core operator, $(p|h_c|q)$. We denote this Lagrangian (which omits the z_{pq} and x_{pq} multipliers) as $\mathcal{L}_{\text{X-DF}}$.

$$\frac{\partial \mathcal{L}_{\text{X-DF}}}{\partial (p|h_c|q)} = \gamma_{pq} = I_{pq} + \sum_k U_{pk}^{\varnothing} \omega_k^{\varnothing} U_{qk}^{\varnothing} + \sum_{kk'} U_{pk}^{\varnothing} \mu_{kk'}^{\varnothing} U_{qk'}^{\varnothing} \tag{84}$$

Here, the identity matrix appears as derivatives of $\mathcal{E}$. In the eigenbasis, the density matrix $\omega_k^{\varnothing}$ accounts for the diagonal contributions. Interestingly, we can see that the Lagrange multipliers, $\mu_{kk'}^{\varnothing}$, account, *exactly*, for the off-diagonal density matrix elements in the eigenbasis. This illustrates how a series of nested Lagrange multipliers – first solving for $\eta_{pk}^{\varnothing}$ and then for $\mu_{kk'}^{\varnothing}$ – can be used to indirectly recover the full one-body density (that is, without evaluating $\gamma_{pq} = \langle \Psi | \hat{E}_{pq} | \Psi \rangle$ as would be done in a classical electronic structure program).

The two-body density matrix can be defined in a similar manner.

$$\begin{aligned}\frac{\partial \mathcal{L}_{\text{X-DF}}}{\partial (pq|rs)} = \Gamma_{pqrs} &= \frac{1}{2} I_{pq} I_{rs} - \frac{1}{8} I_{pr} I_{qs} - \frac{1}{8} I_{ps} I_{qr} \\ &+ \bar{\gamma}_{pq} I_{rs} - \frac{1}{4} \bar{\gamma}_{pr} I_{qs} - \frac{1}{4} \bar{\gamma}_{ps} I_{qr} \\ &+ \sum_t^T \sum_{kl} V_{pq}^t \lambda_k^t \omega_{kl}^t \lambda_l^t V_{rs}^t + \sum_{tu} V_{pq}^t \nu^{tu} V_{rs}^u \end{aligned} \tag{85}$$

Here, we define $\gamma_{pq} = I_{pq} + \bar{\gamma}_{pq}$; this allows the two-electron contributions from the effective one-body operator, $\mathcal{F}_{pq}$, to be incorporated into the two-body density matrix. The remainder of the contributions to the two-body density matrix come from the eigenbasis density, ω_{kl}^t, and the Lagrange multipliers, ν^{tu}.

As defined above, the density matrices have less symmetry than may have been expected. However, in the subsequent steps involving derivatives with respect to the molecular orbital coefficients, these density matrices will be multiplied by and contracted with one- and two-body integral tensors that possess the usual 2-fold and 8-fold symmetries. As a result, we can symmetrize the density matrices to exhibit the same 2-fold and 8-fold symmetry.

$$\tilde{\gamma}_{pq} = \frac{1}{2} \left(\gamma_{pq} + \gamma_{qp} \right) \tag{86}$$

$$\begin{aligned}\tilde{\Gamma}_{pqrs} = \frac{1}{8} \Big(&\Gamma_{pqrs} + \Gamma_{pqsr} + \Gamma_{qprs} + \Gamma_{qpsr} \\ &+ \Gamma_{rspq} + \Gamma_{srpq} + \Gamma_{rsqp} + \Gamma_{srqp} \Big) \end{aligned} \tag{87}$$

These symmetrized density matrices can be used directly in existing formulation of the CP-FON-HF response equations to evaluate the z_{pq} and x_{pq} multipliers.

E. Evaluation of the Gradient

Once all of the Lagrange multipliers have been determined and the Lagrangian is fully stationary with respect to all non-variational parameters, the derivative of the energy can be evaluated by taking the derivative of the Lagrangian with respect to the perturbation of interest, ξ.

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}\xi} \equiv \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\xi} &= \sum_{pq} \hat{\gamma}_{pq} (p|\hat{h}|q)^{\xi} + \sum_{pqrs} \hat{\Gamma}_{pqrs} (pq|rs)^{\xi} \\ &+ \sum_{p>q} z_{pq} f_{pq}^{\xi} - \sum_{p \leq q} x_{pq} (p|q)^{\xi} \end{aligned} \tag{88}$$

We denote the derivative integrals with superscript ξ, that is:

$$(p|q)^{\xi} = \frac{\mathrm{d}}{\mathrm{d}\xi} (p|q) = (\frac{\mathrm{d}}{\mathrm{d}\xi} p|q) + (p|\frac{\mathrm{d}}{\mathrm{d}\xi} q) \tag{89}$$

Here, we have absorbed contributions from the frozen core determinant (i.e., E_c and $(p|\hat{h}_c|q)$) into the $\hat{\gamma}_{pq}$ and $\hat{\Gamma}_{pqrs}$ density matrices. In many cases, it is possible to absorb the z_{pq} multipliers into these density matrices as well. For generality, we contract the z_{pq} multipliers against the derivative Fock matrix elements instead. This allows one to switch between molecular orbitals obtained from Hartree-Fock (which is the case where the z_{pq} multipliers can be absorbed into the one- and two-body density matrices) or those from Kohn-Sham density functional theory (KS-DFT) or even those obtained from semiempirical theories. All that matters is that the coupled-perturbed response equations used to define z_{pq} and x_{pq} are consistent with the self-consistent field method used to determine the molecular orbitals and that the derivative Fock matrix is also consistent with that method.

F. Parameter-shift rules of parameter-locked gate ensembles

Calculating gradients of quantum operations on chip has always been a key part for variational quantum algorithms. Finite difference as a method of evaluating these is flawed due to imperfections and limited precision of measurement on these devices, so parameter shift rules [56–60] were proposed as a way of obtaining the gradient of a generative gate by evaluating the gate at specific angles and using the knowledge one has over the respective objective function. This has been extended and generalized to gates with arbitrary generators and used to not only to calculate the local derivative but extract the entire local objective function for optimisation purposes [60]. Here we want to briefly discuss the case where one gate is composed of multiple of these gates that are parameter-locked to each other and act on distinct qubits/subspaces, as this applies to our spin-adapted orbital rotation gate that is a composition of two Givens rotation gates acting on the alpha and beta qubits respectively.

Using the general shift-rules, one finds an 8-shift rule for the gradient of the device which seems not-optimal as it is comprised of two 2-shift gates. We can recover a 4-shift rule by unlocking the parameters for the calculation of the gradient and using the fact that the individual derivatives of the gates decouple.

To see this, quickly reminder of the derivative of the single Givens rotation:

$$E(\theta) = \langle\Psi|G^\dagger(\theta)\hat{O}G(\theta)|\Psi\rangle \tag{90}$$

$$\partial_\theta E(\theta) = \langle\Psi|G^\dagger(\theta)\hat{O}\partial_\theta G(\theta)|\Psi\rangle + \text{h.c.} \tag{91}$$

We absorbed every gate before the gate of interest into the $\langle\psi|$ and every gate after it into the observable $\hat{O}$ enforce each sub-gate to act on distinct subspaces over the commutation rules. This can be evaluated by a two shift parameter rule. A look at the orbital rotation gate with the same premises:

$$F(\theta) = \langle\Psi|G_2^\dagger(\theta)G_1^\dagger(\theta)\hat{O}G_1(\theta)G_2(\theta)|\Psi\rangle \tag{92}$$

Evaluating the partial derivative via the product rule gives us

$$\begin{aligned}\partial_\theta F(\theta) = &\langle\Psi|G_2^\dagger(\theta)G_1^\dagger(\theta)\hat{O}G_1(\theta)\partial_{\theta'}G_2(\theta')|\Psi\rangle \\ &+ \langle\Psi|G_2^\dagger(\theta)G_1^\dagger(\theta)\hat{O}\partial_{\theta'}G_1(\theta')G_2(\theta)|\Psi\rangle \\ &+ \text{h.c.}\end{aligned} \tag{93}$$

One can now recover the two individual shift rules by rearranging and lifting the parameter-lock between the two Givens rotations to evaluate the two shift rules.

One can now see that each of these terms can now be evaluated with the parameter-shift rule of the individual gate. This leaves us with the following 4-shift rule:

$$F_{\text{ul}}(\theta,\theta') = \langle\Psi|G_2^\dagger(\theta')G_1^\dagger(\theta)\hat{O}G_1(\theta)G_2(\theta')|\Psi\rangle \tag{94}$$

$$\begin{aligned}\partial_\theta F(\theta) = \frac{1}{2}[&\text{F}_{\text{ul}}\left(\theta+\frac{\pi}{2},\theta\right) - \text{F}_{\text{ul}}\left(\theta-\frac{\pi}{2},\theta\right) \\ &+ \text{F}_{\text{ul}}\left(\theta,\theta+\frac{\pi}{2}\right) - \text{F}_{\text{ul}}\left(\theta,\theta-\frac{\pi}{2}\right)]\end{aligned} \tag{95}$$

On a sidenote, this decoupling does not allow for the construction of the local function and only recovers the local derivative at the given angle.

IV. Numerical Demonstration

To test the correctness of the above derivation, we compute the analytic derivative of the ground state energy of butadiene. We use an active space containing 4 electrons in the 4 π and π^* orbitals. The molecular orbitals were obtained from FON-HF/6-31G** using Gaussian broadening of the orbital energy levels with a temperature parameter of 0.3 a.u. We have chosen a geometry of butadiene that is near equilibrium, but slightly distorted to break symmetry. The Cartesian coordinates of the nuclei are given below in Ångströms.

C	25.72300	29.68500	26.90100
C	25.51000	29.88900	25.49000
C	26.07700	30.59500	27.82800
C	25.16300	28.92900	24.60400
H	25.62000	28.64500	27.18200
H	25.33300	30.87700	25.08600
H	26.26500	30.17800	28.85200
H	26.32100	31.69500	27.63100
H	24.81000	29.04000	23.56900
H	25.10400	27.88700	24.94600

There are four cases of particular interest that we consider below. The VQE wavefunction can be converged to

the exact result or it can be truncated. The X-DF can be exact (i.e., all t-leaves are retained) or t-leaves with small eigenvalues can be neglected. This leads to four cases where we will demonstrate the correctness of our analytic gradient formulation.

A. Exact X-DF and Converged VQE

The easiest case to test the correctness of our analytic gradients is where the X-DF is exact and the VQE wavefunction has been converged to the exact ground state energy. This allow direct comparison between classical complete active space configuration interaction (CASCI) analytic gradients and the X-DF/VQE analytic gradient. The numerical results are shown below in units of Hartree/Bohr.

CASCI gradient:

C	0.0087640	0.0338484	-0.0230318
C	0.0338931	-0.0447006	0.0585443
C	-0.0228010	-0.0204001	-0.0339556
C	0.0201028	0.0442291	-0.0012467
H	-0.0159444	-0.0116912	0.0023734
H	-0.0281456	0.0057919	-0.0161052
H	0.0108257	-0.0283045	0.0318503
H	0.0111623	0.0430530	-0.0067838
H	-0.0091125	-0.0101414	-0.0188318
H	-0.0087443	-0.0116847	0.0071869

X-DF/VQE gradient:

C	0.0087640	0.0338484	-0.0230317
C	0.0338931	-0.0447006	0.0585443
C	-0.0228010	-0.0204001	-0.0339556
C	0.0201028	0.0442291	-0.0012467
H	-0.0159444	-0.0116912	0.0023734
H	-0.0281456	0.0057919	-0.0161052
H	0.0108257	-0.0283045	0.0318503
H	0.0111623	0.0430530	-0.0067838
H	-0.0091125	-0.0101414	-0.0188318
H	-0.0087443	-0.0116847	0.0071869

In this case, we achieve agreement of 10^{-7} Hartree/Bohr or better in all of the nuclear degrees of freedom. This result indicates that many of the terms in our Lagrangian are correct, but it does not fully test the formulation because the X-DF has not been truncated – all 10 of the t-leaves are included in the factorization.

B. Truncated X-DF and Converged VQE

It is essential that the analytic gradients be correct when the X-DF is truncated. Here, we discard all eigenvalues of the electron repulsion integral matrix that are below 10^{-1} a.u. This results in only 4 of 10 t-leaves being retained. In this case, we do not have the analogous classical implementation of the gradient available, so we compare against numerical differentiation of the X-DF/VQE energy. We use a 5-point central difference formula with a step size of 10^{-3} Bohr to obtain the numerical derivative.

Numerical X-DF/VQE gradient:

C	0.0024931	0.0253168	-0.0298848
C	0.0351976	-0.0349578	0.0654167
C	-0.0138414	-0.0122360	-0.0250024
C	0.0158476	0.0359590	-0.0099373
H	-0.0148655	-0.0104737	0.0006804
H	-0.0264798	0.0026624	-0.0136967
H	0.0100783	-0.0265528	0.0300495
H	0.0085186	0.0421931	-0.0055591
H	-0.0074004	-0.0105833	-0.0194204
H	-0.0095482	-0.0113273	0.0073544

Analytical X-DF/VQE gradient:

C	0.0024931	0.0253165	-0.0298849
C	0.0351976	-0.0349580	0.0654169
C	-0.0138413	-0.0122359	-0.0250024
C	0.0158476	0.0359592	-0.0099374
H	-0.0148654	-0.0104737	0.0006803
H	-0.0264798	0.0026622	-0.0136968
H	0.0100782	-0.0265528	0.0300495
H	0.0085186	0.0421931	-0.0055592
H	-0.0074003	-0.0105833	-0.0194203
H	-0.0095482	-0.0113273	0.0073545

Here, we achieve agreement of 10^{-6} Hartree/Bohr or better between analytical and numerical gradients. This indicates that our formulation of the X-DF/VQE gradient is correct when the X-DF is truncated. Had there been errors in the gradient formulation, the aggressive truncation we applied should have made them readily apparent.

C. Exact X-DF and Approximate VQE

Another case where our analytic gradient formulation should be correct is when an approximate (i.e., not converged to the exact ground state) VQE wavefunction is used. Our assumption is only that the wavefunction has been variationally optimized, not that it produces the exact energy. In this case, we again compare our analytic gradients against numerical gradients (using the same recipe as above). Note that due the the presence of many local minima in the VQE parameter space, we locate a minimum in the parameter space and use those parameters to seed the calculations at displaced geometries to avoid finding different minima at every geometry.

Numerical X-DF/VQE gradient:

C	-0.0018739	0.0152072	-0.0156463
C	0.0340339	-0.0220893	0.0421282
C	-0.0193957	-0.0108778	-0.0176749
C	0.0124084	0.0339842	-0.0071089
H	-0.0080573	-0.0080723	-0.0010542

```
H   -0.0203369   -0.0023167   -0.0109582
H    0.0086448   -0.0252825    0.0288439
H    0.0125202    0.0424496   -0.0070197
H   -0.0112387   -0.0112398   -0.0170473
H   -0.0067050   -0.0117621    0.0055379
```

Analytical X-DF/VQE gradient:

```
C   -0.0018740    0.0152069   -0.0156461
C    0.0340339   -0.0220895    0.0421281
C   -0.0193956   -0.0108778   -0.0176750
C    0.0124085    0.0339845   -0.0071090
H   -0.0080573   -0.0080723   -0.0010544
H   -0.0203368   -0.0023169   -0.0109584
H    0.0086447   -0.0252825    0.0288439
H    0.0125203    0.0424496   -0.0070198
H   -0.0112387   -0.0112398   -0.0170473
H   -0.0067050   -0.0117621    0.0055379
```

Once again we achieve agreement of better than 10^{-6} Hartree/Bohr between analytical and numerical gradients indicating the correctness of our gradient formulation.

D. Truncated X-DF and Approximate VQE

A final test is that the analytic gradients are correct in cases where both the X-DF is truncated and the VQE energy is an approximation of the exact ground state energy. For this test, we retain 4 of 10 t-leaves in the X-DF.

Numerical X-DF/VQE gradient:

```
C   -0.0088850    0.0045834   -0.0160591
C    0.0339181   -0.0123153    0.0518398
C   -0.0095220   -0.0020111   -0.0103695
C    0.0077376    0.0243913   -0.0198183
H   -0.0054178   -0.0065542   -0.0049439
H   -0.0184659   -0.0055151   -0.0092365
H    0.0070285   -0.0221339    0.0254742
H    0.0105200    0.0415747   -0.0049236
H   -0.0087623   -0.0114716   -0.0180051
H   -0.0081513   -0.0105478    0.0060424
```

Analytical X-DF/VQE gradient:

```
C   -0.0088851    0.0045831   -0.0160591
C    0.0339182   -0.0123156    0.0518400
C   -0.0095219   -0.0020110   -0.0103695
C    0.0077377    0.0243915   -0.0198184
H   -0.0054177   -0.0065542   -0.0049441
H   -0.0184658   -0.0055153   -0.0092367
H    0.0070284   -0.0221340    0.0254742
H    0.0105200    0.0415748   -0.0049237
H   -0.0087623   -0.0114715   -0.0180051
H   -0.0081513   -0.0105478    0.0060424
```

This test shows a similar level of agreement between analytical and numerical gradients.

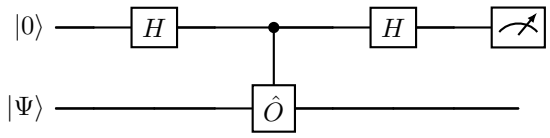


Figure 3: Hadamard test circuit.

3 Purification-based Error-Mitigation Experimental Benchmarking

3.1 Preliminaries on Error-Mitigation

The VQE class of algorithms is mainly categorized as a NISQ set of methods. This is due to its typically shallow circuits and simpler gate set [3] relative to the QPE. Therefore, in order to mitigate the errors that result from the noisy nature of the devices, several error-mitigation techniques have been proposed, such as symmetry-conservation-based approaches [9] in a given symmetry subspace like the particle number or spin-squared operators provided that the ansatz in question preserves these symmetries such as that in [3]. Another class of methods evaluate the desired observables under varying noise levels and statistically infer the noise-free version such as zero-noise extrapolation [22] and probabilistic error cancellation [7]. We focus here specifically on two main prominent methods, namely echo-verification (previously branded as verified phase estimation) [45, 42] and virtual distillation [26, 42].

Like all error-mitigation methods, the goal of the echo-verification is, is to provide an estimate of the noise-free observable $\hat{O}$. In our case, this observable represents the molecular Hamiltonian $\hat{\mathcal{H}}$. Intuitively, the echo-verification technique is an extension over the famous Hadamard test technique whereby, as we will show shortly, all states that are orthogonal to the input state does not contribute to the Hadamard test estimator of the observable. To show this, let us consider the de $|\Psi\rangle$ be a state and we are interested in calculating $\langle\Psi|\hat{O}|\Psi\rangle$. It is well established that using the Hadamard test [41] (See FIG 3), we have

$$|0\rangle|\Psi\rangle \xrightarrow{H} |+\rangle|\Psi\rangle \xrightarrow{C-\hat{O}} \frac{1}{\sqrt{2}}(|0\rangle|\Psi\rangle + |1\rangle\hat{O}|\Psi\rangle) \xrightarrow{H} \frac{1}{2}(|0\rangle|\left(\hat{I}+\hat{O}\right)\Psi\rangle + |1\rangle\left(\hat{I}-\hat{O}\right)|\Psi\rangle). \tag{55}$$

Now, noting p_0 and p_1, the probability of measuring $|0\rangle$ and $|1\rangle$ respectively, we have

$$p_0 = \frac{1}{4}\|(\left(\hat{I}+\hat{O}\right)|\Psi\rangle\|^2 = \frac{1}{4}\langle\Psi|\left(\hat{I}+\hat{O}^\dagger\right)\hat{I}\left(\hat{I}+\hat{O}\right)|\Psi\rangle = \frac{1}{4}\langle\Psi|\left(\hat{I}+\hat{O}^\dagger+\hat{O}+\hat{O}^\dagger\hat{O}\right)|\Psi\rangle \tag{56}$$

$$= \frac{1}{2}\langle\Psi|\left(\hat{I}+Re(\hat{O})\right)|\Psi\rangle = \frac{1}{2}\left(1+Re(\langle\Psi|\hat{O}|\Psi\rangle)\right). \tag{57}$$

Similarly,

$$p_1 = \frac{1}{2}\left(1-Re(\langle\Psi|\hat{O}|\Psi\rangle)\right) \tag{58}$$

which implies that

$$p_0 - p_1 = Re(\langle\Psi|\hat{O}|\Psi\rangle). \tag{59}$$

Now, the identity operator in the middle of Eq (56) can be decomposed into

$$\hat{I} = |\Psi\rangle\langle\Psi| + |\Psi\rangle_\perp\langle\Psi|_\perp. \tag{60}$$

Now, if we look at the contribution in the projector $|\Psi\rangle\langle\Psi|$ of p_0 and p_1, we have

$$o_1 \coloneqq \frac{1}{4}\langle\Psi|\left(\hat{I}+\hat{O}^\dagger\right)|\Psi\rangle\langle\Psi|\left(\hat{I}+\hat{O}\right)|\Psi\rangle = \frac{1}{4}\left(1+2Re(\langle\Psi|\hat{O}|\Psi\rangle)+|\langle\Psi|\hat{O}|\Psi\rangle|^2\right) \tag{61}$$

and in p_1

$$o_2 \coloneqq \frac{1}{4}\langle\Psi|\left(\hat{I}-\hat{O}^\dagger\right)|\Psi\rangle\langle\Psi|\left(\hat{I}-\hat{O}\right)|\Psi\rangle = \frac{1}{4}\left(1-2Re(\langle\Psi|\hat{O}|\Psi\rangle)+|\langle\Psi|\hat{O}|\Psi\rangle|^2\right) \tag{62}$$

which leads to

$$o_1 - o_2 = Re(\langle\Psi|\hat{O}|\Psi\rangle). \tag{63}$$

Therefore, all the contribution from the orthogonal projectors is null. Statistically this means that they cancel out on average. This is effectively the intuition behind the echo-verification which when measuring the control

qubit consider only the results conditioned on the system register begin in the input state $|\Psi\rangle$. It is important to note that, unlike post-selection, the measurement results are averaged with respect to the total number of samples. The error-mitigation property in this stems from the fact that in case of errors the test is more likely to fail. This is due to the stark contrast between the sizes of the Hilbert subspaces of states that pass and fail the verification. The former is spanned by $\{|0\rangle|\Psi\rangle$
$, |1\rangle|\Psi\rangle\}$ and is, consequently, of dimension 2. Therefore, the complement Hilbert subspace is of dimension $2^{n+1}-2$. Now, in the case of VQE, the molecular Hamiltonian is decomposed $\hat{\mathcal{H}} = \sum \hat{H}_s$ into of fast-forwardable terms i.e. terms whose time-evolution operator can be implemented in $O(1)$ depth this includes Pauli strings for instance. As stated in [45], such decomposition for sparse row computable Hamiltonians is always possible. After applying the ansatz, a controlled time-evolution under $\hat{H}_s$ is performed. Lastly, the ansatz is inversed and the check is done to test weather the vacuum state is returned. At the end we have an estimate of $\langle\psi(\vec{\theta})|e^{-i\hat{H}_s t}|\psi(\vec{\theta})\rangle$ which can be used to infer $\langle\psi(\vec{\theta})|\hat{H}_s|\psi(\vec{\theta})\rangle$ because if we note by

$$\hat{H}_s = \sum e_i^s |e_i^s\rangle\langle e_i^s| \tag{64}$$

$$|\psi(\vec{\theta})\rangle = \sum \alpha_i |e_i^s\rangle \tag{65}$$

the spectral decomposition of $\hat{H}_s$ and the expansion of $|\psi(\vec{\theta})\rangle$ in its eigenbasis, we then have

$$\langle\psi(\vec{\theta})|e^{-i\hat{H}_s t}|\psi(\vec{\theta})\rangle = \sum |\alpha_i|^2 e^{-ie_i^s t}. \tag{66}$$

Using signal processing one can determine the amplitudes $|e_i^s|^2$ using the evaluations at different time steps t_i. For instance, the imaginary part of the derivative of $\langle\psi(\vec{\theta})|e^{-i\hat{H}_s t}|\psi(\vec{\theta})\rangle$ at $t = 0$ corresponds exactly to $\langle\psi(\vec{\theta}|\hat{H}_s|\psi(\vec{\theta})\rangle$:

$$\frac{\langle\psi(\vec{\theta})|e^{-i\hat{H}_s t}|\psi(\vec{\theta})\rangle}{dt} = \sum |\alpha_i|^2 \frac{de^{-ie_i^s t}}{dt} = \sum |\alpha_i|^2 e_i^s e^{-ie_i^s t} \tag{67}$$

$$\implies \frac{\langle\psi(\vec{\theta})|e^{-i\hat{H}_s t}|\psi(\vec{\theta})\rangle}{dt}(0) = \sum |\alpha_i|^2 e_i^s = \langle\psi(\vec{\theta})|\hat{H}_s|\psi(\vec{\theta})\rangle \tag{68}$$

which can be evaluated using parameter-shift rules for example [77]. Concretely in [42], we used Pauli strings as H_s summands, the spectrum is then $\{-1, 1\}$. As shown in the experiment results, for both the molecular Hamiltonian and RG cases, the error-mitigated expectations values were x order of magnitudes better compared to raw and post-selected results. For details see [42]. The error-mitigation quality of this method stems from the fact that the measurement that passes the check are exponentially fewer in contrast to the rest of the Hilbert space. More specifically, we have two states that pass the check which are $\{|0\rangle|0_s\rangle$
$, |1\rangle|0_s\rangle\}$. Therefore, it is exponentially more likely for errors to be distributed on the rest of the Hilbert space and therefore rejected.

On the other hand, the virtual distillation [26] uses spatial redundancy, as opposed to echo-verification that uses time-redundancy, to produce the error-mitigated expectation values of $\langle\Psi|\hat{O}|\Psi\rangle$ using M noisy copies to virtually distill ρ^m. This has the effect of suppressing all coherent and incoherent errors exponentially in the number of copies m. It is worth noting that ρ^m is effectively not physically prepared on the device, hence the qualification as "virtual". The error-mitigiated approximation of $\langle\Psi|\hat{O}|\Psi\rangle$ is

$$\frac{\mathrm{Tr}(\hat{O}\rho^M)}{\mathrm{Tr}(\rho^M)}. \tag{69}$$

In [26], they exploit the fact that using M copies of ρ, we have

$$\mathrm{Tr}(\hat{O}\rho^M) = \mathrm{Tr}(\hat{O}S^{(M)}\left(\bigotimes^M \rho\right)) \tag{70}$$

where $S^{(M)}$ is the shift operator defined as:

$$S^{(M)}|\psi_0\rangle|\psi_1\rangle..|\psi_{M-1}\rangle = |\psi_1\rangle|\psi_2\rangle..|\psi_0\rangle \tag{71}$$

and using a symmetrized version of the observable $\hat{O}$, defined as

$$\hat{O}^{(M)} = \frac{1}{M}\sum \hat{O}^i. \tag{72}$$

The shift operator $\hat{S}^{(M)}$ commutes with $\hat{O}^{(M)}$ and can then be simultaneously diagonalized. Therefore, estimating the numerator and denominator of Eq (69) can be achieved from measurement in the same basis. Moreover, it is shown in [26] that the diagonalizing basis, when $\hat{O}^i = \hat{Z}^i$, is

$$B_i^{(q)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, i = 0..2n-1, q = 0..M/2-1 \tag{73}$$

where the index i indicates the number of qubits and q the pair of circuits (copies of ρ). Other observables can thus be deduced by applying the respective diagonalizing rotation before applying B_i

3.2 Purification-based Quantum Error mitigation of Pair-correlated Electron Simulations

Bibliographic Information:

(Double-check bib standards and parctices) O'Brien, T. E., Anselmetti, G., Gkritsis, F., Elfving, V. E., Polla, S., Huggins, **O. Oumarou**, W. J., ... & Vidal, G. (2023). Purification-based quantum error mitigation of pair-correlated electron simulations. Nature Physics, 19(12), 1787-1792.

Summary:

We perform one of the largest (at the time) Variational Quantum Eigenvalue (VQE) experiments to benchmark two of the most prominent error-mitigation techniques, namely echo-verification and virtual distillation on Google's Sycamore device. The two error-mitigation techniques have "orthogonal" properties. Echo-verification duplicates the depth of the circuit by running the ansatz and its inverse, whereas the virtual distillation entangles two copies of the same ansatz. The first set of benchmarks was performed on the evaluation of energy of a pre-optimized UpCCD VQE ansatz. Both methods achieve similar performance, with two orders of magnitude better accuracy compared to the unmitigated energy evaluation. The second set of benchmarks was conducted on evaluating ground state energies of the Richardson-Gaudin model, where we similarly observe on average two orders of magnitude better accuracy.

Contribution:

O.O. co-proved the unfeasibility of classically simulating the UpCCD ansatz considered in the manuscript by proving BQP completeness of the nearest-neighbor Givens-swap circuits on the dual-rail encoding, and wrote the corresponding section in the supplementary informations.

nature physics

Article

https://doi.org/10.1038/s41567-023-02240-y

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

Received: 31 October 2022

Accepted: 6 September 2023

Published online: 12 October 2023

Check for updates

A list of authors and their affiliations appears at the end of the paper

An important measure of the development of quantum computing platforms has been the simulation of increasingly complex physical systems. Before fault-tolerant quantum computing, robust error-mitigation strategies were necessary to continue this growth. Here, we validate recently introduced error-mitigation strategies that exploit the expectation that the ideal output of a quantum algorithm would be a pure state. We consider the task of simulating electron systems in the seniority-zero subspace where all electrons are paired with their opposite spin. This affords a computational stepping stone to a fully correlated model. We compare the performance of error mitigations on the basis of doubling quantum resources in time or in space 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 such as postselection. We study how the gain from error mitigation scales with the system size and observe a polynomial suppression of error with increased resources. Extrapolation of our results indicates that substantial hardware improvements will be required for classically intractable variational chemistry simulations.

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[2–6], attention has focused on algorithms requiring only short-depth quantum circuits, such as the variational quantum eigensolver (VQE)[7,8]. Theoretical developments in ansatz design[8–13] and measurement optimization[14–18] have enabled small to midscale VQE experiments[11,19–26]. A key target of variational quantum algorithms has been the electronic structure problem in chemistry[8,11,19–23,27]. 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 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 in the Hartree–Fock (mean-field) approximation were implemented for system sizes up to 12 qubits in ref. 23, and to our knowledge, this study 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[12,28–33]. This has the advantage of projecting a local fermionic problem onto a local qubit problem[12]. The S_0 ground state is not a priori classically efficiently simulatable[12] (although good approximate methods are known to exist for many problems[33–35]). 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[36], which are not scalable like error correction[1,37] but promise to substantially shrink experimental errors. Popular methods are based on postselection[38,39], rescaling[24,40,41], purification[23,42–44] and probabilistic cancellation[40,45]. Various schemes and combinations of error-mitigation techniques have been implemented in practice[20,22–26,46]. 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

e-mail: teobrien@google.com; christian.gogolin@covestro.com; babbush@google.com; nickrubin@google.com

https://doi.org/10.1038/s41567-023-02240-y

that were crucial to the success of ref. 23: efficient density-matrix purification by means of McWeeny iteration[47], and low-cost gradient estimation.

In this work, we compare the performance of three different error-mitigation techniques—postselection, echo verification (EV) and virtual distillation (VD)—on the problem of preparing ground states in the seniority-zero space, using up to 20 qubits of a superconducting quantum processor. Using either EV or postselected VD, we are able to accurately reproduce the ground-state energy and order parameter for an $N = 10$–qubit simulation of the Richardson–Gaudin (RG) model, improving over unmitigated estimates by one to two orders of magnitude. This demonstrates an improvement over classical pair coupled-cluster doubles (pCCD) and the non-interacting Bardeen–Cooper–Schrieffer (BCS) theory, neither of which are qualitatively correct over the entire range of coupling values considered. We study the scaling of the mean absolute error in energy and the order parameter with the system size and observe that EV and VD suppress errors by a factor N^{α} for values of α between 0.8 and 2.5. EV was further able to greatly mitigate error of six- and ten-qubit simulations of the ring opening of cyclobutene (CB). Although the stringent error requirements (<0.05 Hartree) to differentiate between mean-field and the exact solution could only be achieved for the six-qubit case, this still represents the largest VQE simulation of electronic structure for chemistry that we are aware of so far. From these data, we are able to estimate the minimum requirements for a beyond-classical quantum 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× less than our mean accuracy for the ten-qubit CB problem.

The RG model

We first prepare approximate ground states of the RG model on ten sites at half-filling across a range of coupling strengths g. This model is chosen as a well-known benchmark with well-understood exact and approximate solutions. We work in the seniority-zero space (where all electrons are paired with their opposite spin) and use a unitary pair coupled cluster doubles (UpCCD) variational ansatz[12] with parameters optimized in noiseless simulation. In Fig. 1a, we estimate the prepared states' energy using various error-mitigation techniques: postselection, EV and postselected virtual distillation (PS-VD). We compare these results to exact diagonalization in the S_0 subspace (also known as double occupied configuration interaction) and classical 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. 1b we plot the error between our experimental data and the UpCCD ground-state energy. The observed gain from error mitigating some estimate y can be quantified by the suppression factor

$$\eta_y = |y_{\text{raw}} - y_{\text{true}}|/|y_{\text{mit}} - y_{\text{true}}|. \quad (1)$$

We summarize this across our experimental data in Table 1. For the energy ($y = E$), we find PS-VQE consistently achieves $\eta_E \approx 2$, whereas EV achieves $\bar{\eta}_E = 85$ and $\max(\eta_E) = 460$ and VD achieves $\bar{\eta}_E = 60$ and $\max(\eta_E) = 140$. The residual error following EV or PS-VD fluctuates on a scale larger than statistical error bars, which we attribute 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[48–50]. This gives us a second physically relevant target to simulate and a qualitative feature (the cusp at $g = 0$) to resolve. The traditional order parameter for the BCS state, $\Delta_{\text{BCS}} = \frac{1}{2N}\sum_j(\langle a_{j\uparrow}a_{j\downarrow}\rangle + \langle a^{\dagger}_{j\downarrow}a^{\dagger}_{j\uparrow}\rangle)$, is zero on the RG Hamiltonian ground state due to number conservation. However, $\Delta = \frac{1}{N}\sum_{j,\sigma}\sqrt{\langle n^2_{j\sigma}\rangle - \langle n_{j\sigma}\rangle^2}$ satisfies $\Delta = \Delta_{\text{BCS}}$ for the BCS ground state of the Hamiltonian, giving a many-body-order parameter[49]. In Fig. 1c, we plot experimental estimates of Δ across the range of g values considered. In the absence of error mitigation, although 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 $g = 0$. We measure the performance of the error-mitigation techniques by plotting the error in Δ against the noise-free UpCCD result in Fig. 1d. The suppression (equation (1), $y = \Delta$) from EV and PS-VD during order-parameter estimation is slightly less than for energy estimation: for EV, we find $\bar{\eta}_\Delta = 32$ and $\max(\eta_\Delta) = 56$; and for PS-VD, we find $\bar{\eta}_\Delta = 18$ and $\max(\eta_\Delta) = 51$.

We now study the scaling of the studied error-mitigation techniques for the RG model over system sizes $N = 4, 6, 8, 10$. In Fig. 2a, we plot the absolute error in the energy estimates, averaged across all points in Fig. 1, and repeat this for different system sizes N. (Plots equivalent to Fig. 1 for different system sizes can be found in Supplementary Information Section VI.D.) We observe that the energy

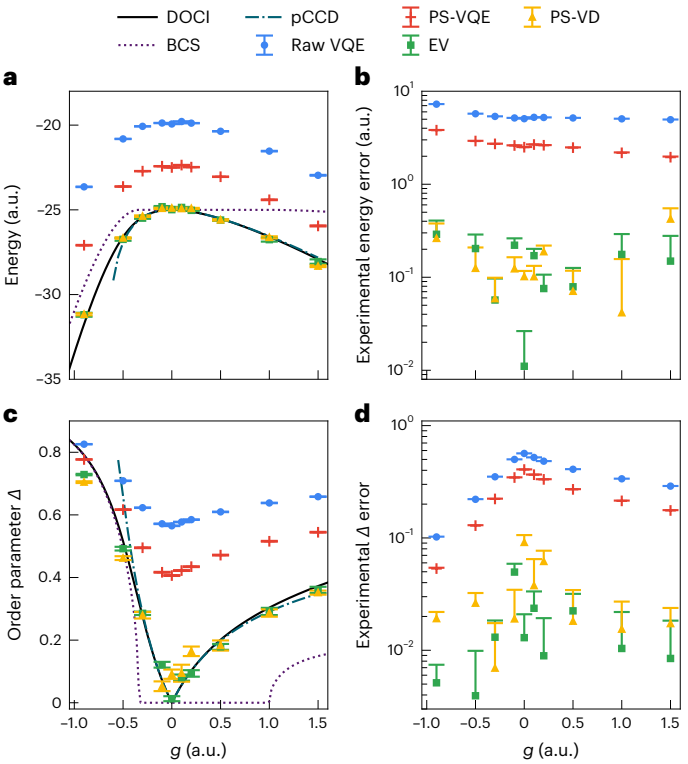


Fig. 1 | Digital quantum simulation of ground states of the RG model for ten spatial orbitals on a superconducting quantum device. a, Energy as a function of the coupling parameter g, estimated using various error-mitigation techniques (markers) and compared to classical models (lines). The classical pCCD results do not converge below a critical value, resulting in their cut-off. **b**, Log plot of experimental energy error (ignoring the model error from the UpCCD approximation). **c**, Many-body order parameter for the RG Hamiltonian (see text), again compared to classical models. **d**, Experimental error in estimating the superconducting order parameter versus the target state in the UpCCD approximation (again ignoring model error). Error bars show 1 standard deviation uncertainty from sampling noise, estimated by propagating variance (raw VQE, PS-VQE) or bootstrapping (EV, PS-VD); see Supplementary Information Section III for details. a.u., arbitrary units.

https://doi.org/10.1038/s41567-023-02240-y

Table 1 | Observed mean and maximum suppression factors $\bar{\eta}$ and $\max(\eta)$ (equation (1)) across the RG problems studied in Fig. 1 and the CB problems studied in Fig. 3, for all error-mitigation methods studied

System	Metric	PS-VQE	EV	PS-VD
RG (ten qubits, energy)	$\bar{\eta}_E$	2.1	86	54
	$\max(\eta_E)$	2.5	460	120
RG (ten qubits, order parameter)	$\bar{\eta}_\Delta$	1.6	32	18
	$\max(\eta_\Delta)$	1.9	56	51
CB (six qubits, energy)	$\bar{\eta}_E$	5.0	54	–
	$\max(\eta_E)$	5.2	140	–
CB (ten qubits, energy)	$\bar{\eta}_E$	2.9	38	–
	$\max(\eta_E)$	3.0	190	–

error scales sublinearly after applying EV or VD, demonstrating an asymptotic difference from raw VQE or PS-VQE of a factor of ~N^2. We similarly observe an ~N–N^2 gap between EV/VD and VQE/PS-VQE in the order-parameter error (Fig. 2b); the difference in the suppression factor is not statistically significant. Next, we extract various fidelity metrics from the experimental data in Fig. 2c, which approximate the (inverse) increased variance of our error-mitigated estimators[36]. The data for PS-VQE (red plus), EV (green square) and PS-VD (yellow triangle) directly correspond to the circuit overhead required by different error-mitigation techniques. These fidelity metrics allow us to estimate the required number of experiment repetitions to reduce statistical noise in the energy estimation to <0.1 a.u.(arbitrary units) in Fig. 2d, assuming ten-qubit fidelity is maintained. We observe a substantial gap between the experimental and theoretical sampling cost of EV (green squares). We attribute this to known overheads of EV that are not in the theoretical model (Supplementary Information Section IV.A).

CB ring opening

We further validated scalable error-mitigation protocols by simulating the conrotatory ring opening pathway for CB 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 electronic structure Hamiltonian. 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. Similar to the RG model, this was chosen as a well-known benchmark system with well-understood chemistry.

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 in mean-field theory. This is verified numerically for our seniority-zero model where the largest configuration interaction coefficient has an average value of 0.974(9) for six orbitals and 0.973(9) for ten orbitals, indicating a single-reference system. As such, our unitary pCCDs ansatz targets the dynamic correlation corrections to the mean field.

The average PS-VQE absolute error is 0.058 ± 0.006 and 0.395 ± 0.023 Hartree, and the mean suppression factor is $\bar{\eta}_E = 5.0$ and 2.9 for the six- and ten-orbital systems, respectively. The average echo-verified absolute error is 0.011 ± 0.005 and 0.064 ± 0.035 Hartree and the mean suppression factor (equation (1)) is $\bar{\eta}_E = 54$ and 38 for the six- and ten-orbital system respectively. Although there is notable improvement in energy across the reaction pathway for the ten-orbital system, the magnitude of the errors is larger than the 0.037 Hartree energy difference between CB and 1,3-butadiene. Furthermore, a visual inspection of Fig. 3 indicates high parallelity errors in the ten-orbital system. Given that the error bars on EV are smaller than the parallelity error (point scatter), we attribute the main source of error to device drift.

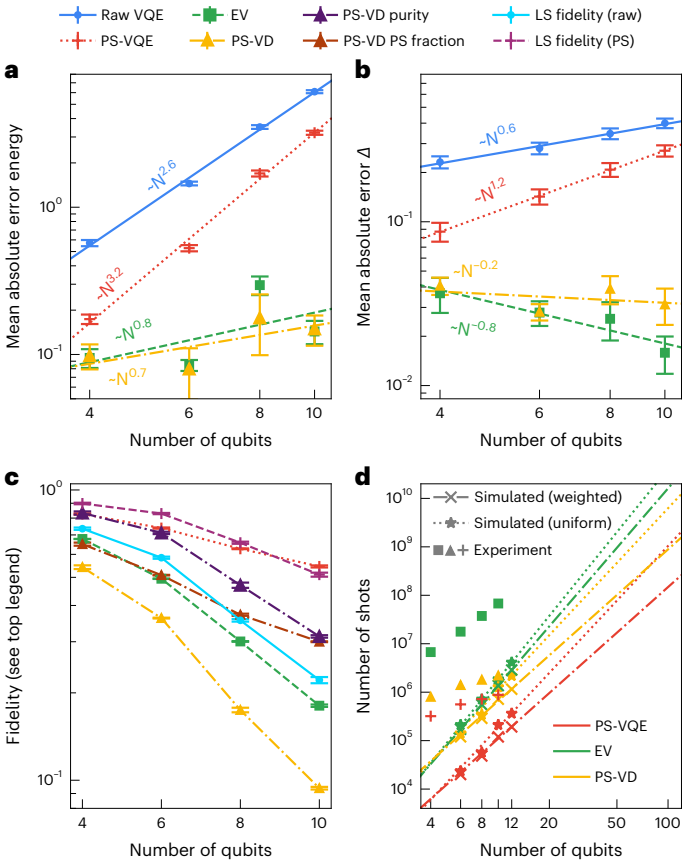


Fig. 2 | Scaling the simulation of the RG model to larger qubit counts. **a**, Experimental energy error (versus the UpCCD ground state) averaged over all points studied of the RG model. Error bars show sample standard deviation and lines a power-law fit. **b**, Experimental error in order parameter (versus the UpCCD ground state) averaged over all points studied of the RG model. Error bars and lines same as **a**. **c**, Different fidelity metrics for PS-VQE, EV, PS-VD and Loschmidt (LS) echo (see legend) averaged over all points studied of the RG model. **d**, Number of shots required for convergence at $g = -0.9$. Crosses and pluses give simulated estimations with two types of term grouping (see text for details) using observed fidelities of a ten-qubit experiment from **c**. Other symbols give experimental shots used.

Discussion and outlook

Many of the features of our experimental data can be understood physically. The similar performance of EV and PS-VD can be understood following the equivalence of the two techniques shown in refs. 46,51, which treat EV as purification of two states separated in time. However, this equivalence is broken by the further circuitry of the two techniques: EV requires Greenberger-Horne-Zeilinger (GHZ) state preparation, whereas VD requires bell-basis measurements and more routing. This doubles the susceptibility of the VD fidelity to measurement noise, to which we attribute the gap between this and the EV fidelity in Fig. 2c. Despite the fidelity gap, we observe PS-VD has a substantially smaller sampling cost than EV, which comes from known measurement bottlenecks[52] and overhead to reduce on-chip magnetic field fluctuations (Supplementary Information Section II.D). We can attribute the gap in performance between VD with and without

https://doi.org/10.1038/s41567-023-02240-y

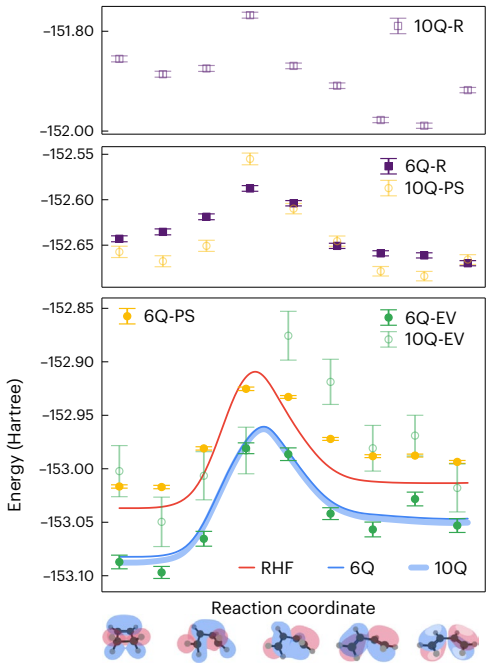


Fig. 3 | Conrotatory CB ring opening pathway simulated in the seniority-zero subspace. Comparison of raw VQE (purple, 'R'), PS-VQE (yellow, 'PS') and EV (green, 'EV') on an optimized unitary pair coupled-cluster ansatz. Darker (lighter) points correspond to six-qubit (6Q) (ten-qubit (10Q)) simulations. Error bars denote 1 s.d. calculated by either propagating error (VQE) or bootstrapping (EV). 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 ten orbitals (broad, lighter-blue line) and six orbitals (narrow, darker-blue line). The red curve is the RHF mean-field energy. Some of the data are plotted on a discontinuous and different scale to preserve the visual scale of the reaction energy along the reaction coordinate.

postselection to this further measurement circuitry (Supplementary Information Section VI.A): we suggest that the bell-basis measurements used are not robust to noise that does not conserve particle number and that the primary effect of postselection is to mitigate this noise source. The performance improvement of EV/VD over postselection is expected from numerical predictions in refs. 42,44 and analytic results of ref. 53 (although we show in Supplementary Information Section VI.B that this cannot be guaranteed). Finally, we can explain the qualitative feature in Fig. 1 around $g = 0$, where the energy error drops but the order parameter error increases. This comes from two coincidental effects: Δ is more sensitive to error around $\Delta = 0$, but at this point the energy becomes less dependent on off-diagonal terms, reducing its sensitivity to phase noise.

We now consider how the required number of experiment repetitions (shots) scales to classically intractable simulations, using the data in Fig. 2d. Taking a 50-qubit experiment as a proxy lower bound for a beyond-classical quantum computation, and assuming the ability to freely weight our shot distribution, we estimate that at least 10^8 or 10^9 shots would be required when using VD or EV, respectively. This is executable on current hardware in a wall-clock time of >1 hour or >10 hours, respectively. Including the difference between experiment and theory at ten qubits raises the cost of EV to 5×10^{10} shots, which would require several days to achieve. These numbers do not include the large multiplicative cost of variational optimization. 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.

Next, we consider what device fidelity would be required to scale to a beyond-classical experiment. To maintain circuit fidelity F over a depth $3N/2$, fully parallel circuit as N scales from 10 to 50 requires all error rates to drop by a factor of 25 ($\sim N^2$). As any reduction in F incurs a poly(F^{-1}) sampling cost[42,44], and as F scales exponentially in the error rate, and as $F \approx 10\%$ for PS-VD (Fig. 2d), 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 fidelities on all two-qubit gates $\lesssim 3 \times 10^{-4}$. This analysis likely precludes a large number of ansatzes with depth much greater than linear. For instance, successfully implementing a 50-qubit VQE with ansatz depth $3N^2/2$ with EV or VD would require error rates to drop ~1,000×. However, the sublinear scaling in the energy error following the application of EV or VD (Fig. 2a) indicates that a 25× decrease in error rate would be sufficient to deal with the bias in the resulting experiment; the dominant cost of error-mitigated quantum experiments comes from the sampling and fidelity requirements. Confirming this point in studies on other systems is a key target for future work.

We have observed that EV and VD suppress errors by one to two orders of magnitude on a range of quantum simulation problems using up to 20 superconducting qubits. We further observe that the error-suppression performance of these mitigation techniques increases with the number of qubits. This allows us to conclude that variance, not bias, will be the limiting factor in purification-based error-mitigation techniques going forward. As our error-mitigation protocols target expectation values themselves and are blind to the outer loops combining these to estimate energies or order parameters, we expect these results to hold for a wider range of problems than the VQE examples studied here. Any algorithm using circuits with similar structure, such as time evolution on spin chains, or brick-wall parameterized quantum circuits for quantum machine learning, should see similar gains in performance. Testing whether this extends to more generic circuits (for example, higher-dimensional or sparse circuits, or primitive circuits for fault tolerance[54]) is a clear target for future work.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41567-023-02240-y.

References

1. Acharya, R. et al. Suppressing quantum errors by scaling a surface code logical qubit. *Nature* **614**, 676–681 (2023).
2. von Burg, V. et al. Quantum computing enhanced computational catalysis. *Phys. Rev. Res.* **3**, 033055 (2021).
3. Lee, J. et al. Even more efficient quantum computations of chemistry through tensor hypercontraction. *PRX Quantum* **2**, 030305 (2021).
4. Gidney, C. & Ekerå, M. How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. *Quantum* **5**, 433 (2021).
5. Campbell, E. T. Early fault-tolerant simulations of the Hubbard model. *Quant. Sci. Technol.* **7**, 015007 (2021).
6. Berry, D. W. et al. Quantifying quantum advantage in topological data analysis. Preprint at https://arxiv.org/abs/2209.13581 (2022).
7. Peruzzo, A. et al. A variational eigenvalue solver on a quantum processor. *Nat. Commun.* **5**, 4213 (2014).
8. McClean, J. R., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **18**, 23023 (2016).

9. Wecker, D., Hastings, M. B. & Troyer, M. Progress towards practical quantum variational algorithms. *Phys. Rev. A* **92**, 042303 (2015).
10. Grimsley, H. R., Economou, S. E., Barnes, E. & Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* **10**, 3007 (2019).
11. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
12. Elfving, V. E., Millaruelo, M., Gámez, J. A. & Gogolin, C. Simulating quantum chemistry in the seniority-zero space on qubit-based quantum computers. *Phys. Rev. A* **103**, 032605 (2021).
13. Lee, J., Huggins, W. J., Head-Gordon, M. & Whaley, K. B. Generalized unitary coupled cluster wave functions for quantum computation. *J. Chem. Theory Comput.* **15**, 311 (2018).
14. Huggins, W. J. et al. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Inf.* **7**, 23 (2021).
15. Cotler, J. & Wilczek, F. Quantum overlapping tomography. *Phys. Rev. Lett.* **124**, 100401 (2020).
16. Bonet-Monroig, X., Babbush, R. & O'Brien, T. E. Nearly optimal measurement scheduling for partial tomography of quantum states. *Phys. Rev. X* **10**, 031064 (2020).
17. Verteletskyi, V., Yen, T.-C. & Izmaylov, A. F. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *J. Chem. Phys.* **152**, 124114 (2020).
18. Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020).
19. O'Malley, P. J. J. et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 31007 (2016).
20. Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019).
21. Hempel, C. et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X* **8**, 031022 (2018).
22. Sagastizabal, R. et al. Error mitigation by symmetry verification on a variational quantum eigensolver. *Phys. Rev. A* **100**, 010302 (2019).
23. Arute, F. et al. Hartree-Fock on a superconducting qubit quantum computer. *Science* **369**, 1084–1089 (2020).
24. Stanisic, S. et al. Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer. *Nat. Comm.* **13**, 5743 (2022).
25. Kim, Y. et al. Scalable error mitigation for noisy quantum circuits produces competitive expectation values. *Nat. Phys.* **19**, 752–759 (2023).
26. van den Berg, E., Minev, Z. K., Kandala, A. & Temme, K. Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors. *Nat. Phys.* **19**, 1116–1121 (2023).
27. Motta, M. et al. Quantum chemistry simulation of ground- and excited-state properties of the sulfonium cation on a superconducting quantum processor. *Chem. Sci.* **14**, 2915–2927 (2023).
28. Limacher, P. A. et al. A new mean-field method suitable for strongly correlated electrons: computationally facile antisymmetric products of nonorthogonal geminals. *J. Chem. Theory Comput.* **9**, 1394 (2013).
29. Boguslawski, K. et al. Efficient description of strongly correlated electrons with mean-field cost. *Phys. Rev. B* **89**, 201106 (2014).
30. Johnson, P. A. et al. Richardson-Gaudin mean-field for strong correlation in quantum chemistry. *J. Chem. Phys.* **153**, 104110 (2020).
31. Gunst, K., Van Neck, D., Limacher, P. A. & De Baerdemacker, S. The seniority quantum number in tensor network states. *SciPost Chem.* **1**, 001 (2021).
32. Kossoski, F., Damour, Y. & Loos, P.-F. Hierarchy configuration interaction: combining seniority number and excitation degree. *J. Phys. Chem. Lett.* **13**, 4342–4349 (2022).
33. Fecteau, C.-E. et al. Near-exact treatment of seniority-zero ground and excited states with a Richardson-Gaudin mean field. *J. Chem. Phys.* **156**, 194103 (2022).
34. Dukelsky, J., Roman, J. M. & Sierra, G. Comment on 'polynomial-time simulation of pairing models on a quantum computer'. *Phys. Rev. Lett.* **90**, 249803 (2003).
35. Dukelsky, J. Integrable Richardson-Gaudin models in mesoscopic physics. *J. Phys. Conf. Ser.* **338**, 012023 (2012).
36. Cai, Z. et al. Quantum error mitigation. *Rev. Mod. Phys.* (in the press).
37. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
38. McArdle, S., Yuan, X. & Benjamin, S. Error-mitigated digital quantum simulation. *Phys. Rev. Lett.* **122**, 180501 (2019).
39. Bonet-Monroig, X., Sagastizabal, R., Singh, M. & O'Brien, T. Low-cost error mitigation by symmetry verification. *Phys. Rev. A* **98**, 062339 (2018).
40. Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
41. Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
42. Huggins, W. J. et al. Virtual distillation for quantum error mitigation. *Phys. Rev. X* **11**, 041036 (2021).
43. Koczor, B. Exponential error suppression for near-term quantum devices. *Phys. Rev. X* **11**, 031057 (2021).
44. O'Brien, T. E. et al. Error mitigation via verified phase estimation. *PRX Quantum* **2**, 020317 (2021).
45. Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027 (2018).
46. Huo, M. & Li, Y. Dual-state purification for practical error mitigation. *Phys. Rev. A* **105**, 022427 (2022).
47. McWeeny, R. Some recent advances in density matrix theory. *Rev. Mod. Phys.* **35**, 668 (1963).
48. von Delft, J., Zaikin, A., Golubev, D. & Tichy, W. Parity-affected superconductivity in ultrasmall metallic grains. *Phys. Rev. Lett.* **77**, 3189 (1996).
49. Braun, F. & von Delft, J. Superconductivity in ultrasmall metallic grains. *Phys. Rev. B* **59**, 9527 (1999).
50. Dukelsky, J. & Sierra, G. The crossover from the bulk to the few-electron limit in ultrasmall metallic grains. *Phys. Rev. B* **61**, 12302 (1999).
51. Cai, Z. Resource-efficient purification-based quantum error mitigation. Preprint at https://arxiv.org/abs/2107.07279 (2021).
52. Polla, S., Anselmetti, G.-L. R. & O'Brien, T. E. Optimizing the information extracted by a single qubit measurement. *Phys. Rev. A*, **108**, 012403 (2023).
53. Koczor, B. The dominant eigenvector of a noisy quantum state. *New. J. Phys.* **23**, 123047 (2021).
54. Pivetaeu, C., Sutter, D., Bravyi, S., Gambetta, J. M. & Temme, K. Error mitigation for universal gates on encoded qubits. *Phys. Rev. Lett.* **127**, 200505 (2021).

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], J. C. Bardin[1,5], A. Bengtsson[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], 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], 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], T. Khattar[1], M. Khezri[1], M. Kieferová[1,8], S. Kim[1], P. V. Klimov[1], A. R. Klots[1], A. N. Korotkov[1,9], F. Kostritsa[1], J. M. Kreikebaum[1], D. Landhuis[1], P. Laptev[1], K.-M. 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], F. D. Malone[1], S. Mandrà[1], O. Martin[1], S. Martin[1], J. R. McClean[1], T. McCourt[1], M. McEwen[1,11], 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], 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], W. C. Smith[1], R. D. 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], D. Bacon[1], S. Boixo[1], Y. Chen[1], J. Hilton[1], J. Kelly[1], E. Lucero[1], A. Megrant[1], H. Neven[1], V. Smelyanskiy[1], C. Gogolin[3] ✉, R. Babbush[1] ✉ & N. C. Rubin[1] ✉

[1]Google Research, Mountain View, CA, USA. [2]Instituut-Lorentz, Universiteit Leiden, Leiden, the Netherlands. [3]Covestro Deutschland AG, Leverkusen, Germany. [4]PASQAL SAS, Massy, France. [5]Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, MA, USA. [6]Department of Physics, University of Connecticut, Storrs, CT, USA. [7]Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA. [8]QSI, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, New South Wales, 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, USA. ✉e-mail: teobrien@google.com; christian.gogolin@covestro.com; babbush@google.com; nickrubin@google.com

Methods

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[28,29,31,32] and as a starting point for modelling strong correlations from electron-pair states[55].

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[56]

$$\begin{aligned} P_p^\dagger &= a_{p\alpha}^\dagger a_{p\beta}^\dagger, \quad N = \sum_{p,\sigma} a_{p\sigma}^\dagger a_{p\sigma}, \\ \left[P_p, P_q^\dagger\right] &= (1-N_p)\delta_{p,q}, \quad \left[N_p, P_q\right] = -2P_q\delta_{p,q}, \end{aligned} \tag{2}$$

where p, q and α, β are orbital and spin indices, respectively, and δ is the Kronecker delta function. 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 RG or pairing model

$$\hat{H} = \sum_{p=1}^{N} \epsilon_p N_p + g \sum_{p\neq q=1}^{N} P_p^\dagger P_q, \tag{3}$$

with ε_p the single-particle energy for site p, and g the coupling strength. At small N this corresponds to a single strength all-to-all coupling of fermion pairs. In this Article, we work at half-filling by adding a chemical potential μ to the single-particle energies:

$$\epsilon_p = p - \mu, \quad \mu = \frac{1}{2}(N+1) \tag{4}$$

This is a model for a small superconducting grain when $g < 0$ (refs. 48–50), but with a g-dependent Debye frequency[49]. The second model is the electronic structure Hamiltonian (H_{elec}) projected into the S_0 subspace:

$$\begin{aligned} H_{S_0} &= P_{S_0} H_{\text{elec}} P_{S_0} = \sum_p (h_{p,p}) N_p \\ &+ \frac{1}{4} \sum_{p\neq q} (2V_{pqpq} - V_{pqqp}) N_p N_q + \sum_{pq} (V_{ppqq}) P_p^\dagger P_q, \end{aligned} \tag{5}$$

where h and V are the one- and two-body term coefficients. The all-to-all connected Heisenberg spin Hamiltonian is not known to be classically solvable in general, but good approximate methods exist. This is especially true for the RG model, which is often integrable[35], well-approximated by density-matrix renormalization group[34] 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[29,57–59] while full configuration interaction quantum Monte Carlo shows a reduced sign problem[60]. As such, although we have strong evidence that the quantum circuits used in this text are not classically simulatable (Supplementary Information Section II.A), 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.

The unitary pair coupled-cluster ansatz and energy estimation

In this work, we use a Trotterized UpCCD ansatz[12] compiled into a set of qubits in a ladder geometry with nearest-neighbour 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

$$U(\boldsymbol{\theta}) = \prod_{\ell}^{N_\ell} U_e(\theta^{e,\ell}) U_o(\theta^{o,\ell}) \tag{6}$$

$$U_o(\theta^{o,\ell}) = \prod_{n=0}^{N/2-1} \mathrm{GS}_{2n+1,(2n+2)\%N}\left(\theta^{o,l}_{2n+1,(2n+2)\%N}\right) \tag{7}$$

$$U_e(\theta^{e,\ell}) = \prod_{n=0}^{N/2-1} \mathrm{GS}_{2n,2n+1}\left(\theta^{e,l}_{2n,2n+1}\right) \tag{8}$$

where each $\mathrm{GS}_{ij}(\theta)$ is a Givens SWAP gate corresponding to the product of a Givens rotation gate by an angle θ on a pair labelled by qubits i and j followed by a SWAP operation[12]

$$\mathrm{GS}(\phi) = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{SWAP}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) & 0 \\ 0 & \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{Givens}}. \tag{9}$$

The GS gate corresponds to a coherent partial pair excitation (by the angle ϕ) followed by a pair SWAP. Given several layers N_ℓ in equation (6) 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 $|0101\ldots01\rangle$—for example, 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 $N/2$ (Supplementary Information Section II.C). Thus, in this work, we set $N_\ell = N/2$ for all systems. Each GS(θ) gate is compiled into a product of three controlled-Z (CZ) gates interleaved with tunable single-qubit microwave gates (Extended Data Fig. 1, top; see Supplementary Information Section II.B for decompositions).

To perform energy estimation on our two S_0 models, expectation values with respect to nearest-neighbour and non-nearest-neighbour qubits are required. The expectation value $\langle X_iX_j + Y_iY_j\rangle$ is estimated by performing a number-preserving diagonalization[16,23] 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-neighbour pairs across the rungs of the ladder in a similar fashion at the further cost of at most one SWAP operation. The full measurement protocol is detailed further in Supplementary Information Section II.C. 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 circuits exist, but they do not allow for postselection on particle number.

EV and VD

EV, introduced in ref. 44, 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 (refs. 46,52). 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 $|\phi\rangle\langle\psi|$ (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 $i|\phi\rangle\langle\psi| - i|\psi\rangle\langle\phi|$) 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 total, adding control-free EV to a variational circuit requires doubling the circuit depth and adding circuitry to prepare and unprepare GHZ states. This will reduce the fidelity of the final circuit from F to $< F^2$. As experiment runtime scales inversely with circuit fidelity, this will cost us slightly in the experiment.

To estimate expectation values by means of EV, we use the fact that

$$\langle\Psi|\phi\rangle\langle\psi|\Psi\rangle = \frac{1}{2}\langle\psi|O|\psi\rangle e^{i\phi}, \tag{10}$$

where $O|\phi\rangle = e^{i\phi}|\phi\rangle$. The expectation value $\langle\psi|O|\psi\rangle$ can be recovered from equation (10) as the other terms are known. The largest effect of noise on the echo-verified estimator (equation (10)) is to dampen $\langle\Psi|\phi\rangle\langle\psi|\Psi\rangle \to F\langle\Psi|\phi\rangle\langle\psi|\Psi\rangle$ (ref. 44). We can estimate F independently by removing O from the circuit, which yields a Loschmidt echo of the preparation unitary[61]. This is achieved in practice by removing a virtual Z rotation (Extended Data Fig. 1, bottom), making the estimated Loschmidt fidelity an accurate estimate of F. Further EV implementation details can be found in Supplementary Information Section II.D.

VD[42,43] is an error-mitigation technique that uses collective measurements of k copies of a state ρ to estimate expectation values with respect to $\rho^k/\mathrm{Tr}[\rho^k]$, where Tr is the trace. VD schemes are based on the observation that the cyclic shift operator $S^{(k)}$ is easily diagonalized and therefore can be measured, which yields, for example, for $k = 2$,

$$\mathrm{Tr}[\rho\otimes\rho S^{(2)}] = \mathrm{Tr}[\rho^2], \mathrm{Tr}[\rho\otimes\rho S^{(2)}O_s] = \mathrm{Tr}[\rho^2 O], \tag{11}$$

with $O_s = \frac{1}{2}(I\otimes O + O\otimes I)$, and I is the identity. $S^{(2)}$ can be simultaneously diagonalized with O_s when $O = Z_i$ by a GS(π/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 (Section 4.2) and following the above procedure. The VD circuit is only six two-qubit gates deeper than PS-VQE. However, as we double the number of qubits used, we more than double the total circuit size and reduce the total circuit fidelity again from F to $<F^2$.

As the GS(π/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 before 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 present results showing VD with postselection only (PS-VD). We compare VD with and without postselection in Supplementary Information Section VI.A.

Data availability

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

Code availability

Code to generate quantum circuits and process raw data is available in ReCirq: https://github.com/quantumlib/ReCirq.

References

55. Surján, P. R., Szabados, Á., Jeszenszki, P. & Zoboki, T. Strongly orthogonal geminals: size-extensive and variational reference states. *J. Math. Chem.* **50**, 534–551 (2012).
56. Ring, P. & Schuck, P. *The Nuclear Many-Body Problem* (Springer Science & Business Media, 2004).
57. Henderson, T. M., Bulik, I. W., Stein, T. & Scuseria, G. E. Seniority-based coupled cluster theory. *J. Chem. Phys.* **141**, 244104 (2014).
58. Stein, T., Henderson, T. M. & Scuseria, G. E. Seniority zero pair coupled cluster doubles theory. *J. Chem. Phys.* **140**, 214113 (2014).
59. Vu, N. & Eugene DePrince, A. Size-extensive seniority-zero energy functionals derived from configuration interaction with double excitations. *J. Chem. Phys.* **152**, 244103 (2020).
60. Shepherd, J. J., Henderson, T. M. & Scuseria, G. E. 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. *J. Chem. Phys.* **144**, 094112 (2016).
61. Mi, X. et al. Information scrambling in quantum circuits. *Science* **374**, 1479–1483 (2021).

Acknowledgements

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.

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 $2\times N$ grid. F.G. and C.G. designed the conjugate model gradient descent algorithm and preoptimized ansatz parameters for the ansatz. N.C.R. wrote the pair coupled-cluster code, preoptimized ansatz parameters and performed the classical chemistry calculations for the CB 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 coordinated 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.

Competing interests

The authors declare no competing interests.

Additional information

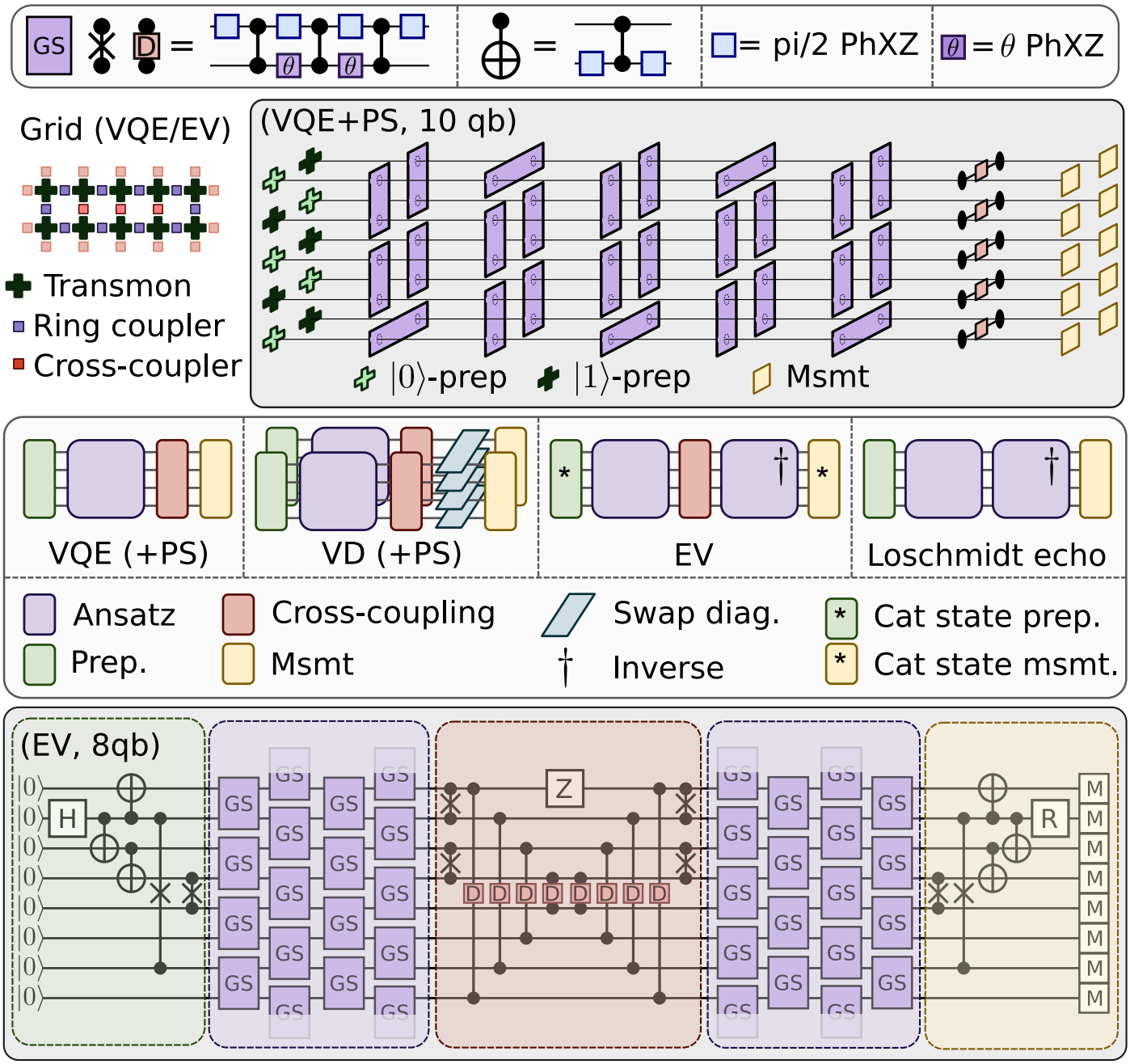
Extended data is available for this paper at https://doi.org/10.1038/s41567-023-02240-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41567-023-02240-y.

Correspondence and requests for materials should be addressed to T. E. O'Brien, C. Gogolin, R. Babbush or N. C. Rubin.

Peer review information *Nature Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.

https://doi.org/10.1038/s41567-023-02240-y



Extended Data 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 Supplementary Information Section II.B 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.

nature physics

Article

https://doi.org/10.1038/s41567-023-02240-y

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

In the format provided by the authors and unedited

I. 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 [1] and 0.008 speckle purity [1]. 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 [2, 3], 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.

II. FURTHER DETAILS OF THE UPCCD ANSATZ

A. 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 match-gate.

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

$$|\tilde{0}\rangle \coloneqq |01\rangle \tag{1}$$

$$|\tilde{1}\rangle \coloneqq |10\rangle. \tag{2}$$

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:

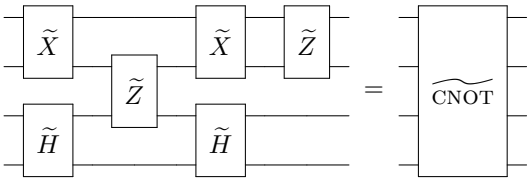
$$GS(\frac{\pi}{4}) = \widetilde{H} \tag{3}$$

$$GS(0) = \widetilde{X} \tag{4}$$

$$GS(\frac{\pi}{2}) = \widetilde{Z} \tag{5}$$

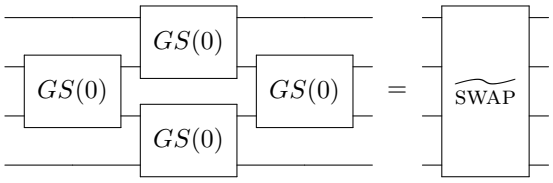
$$GS(0) \cdot GS(\theta) = \widetilde{RY}(\theta) \tag{6}$$

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



(7)

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



. (8)

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.

B. 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 [3]. 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 [1]

$$R(\alpha_x, \alpha_a, \alpha_z) = e^{i(\alpha_z+\alpha_a)Z} e^{i\alpha_x X} e^{-i\alpha_a Z} \tag{9}$$

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

$$\begin{aligned} GS_{i,j}(\theta) =& R_i(\tfrac{\pi}{4}, \tfrac{\pi}{4}, 0) \cdot CZ_{i,j} \\ & \cdot R_i(\tfrac{\pi}{4}, \tfrac{-\pi}{4}, 0) \cdot R_j(\tfrac{\pi}{4}+\tfrac{\theta}{2}, \tfrac{-\pi}{4}, \tfrac{\pi}{2}) \cdot CZ_{i,j} \\ & \cdot R_i(\tfrac{\pi}{4}, \tfrac{\pi}{4}, 0) \cdot R_j(\tfrac{\pi}{4}+\tfrac{\theta}{2}, \tfrac{-\pi}{4}, \tfrac{\pi}{2}) \cdot CZ_{i,j} \\ & \cdot R_i(\tfrac{\pi}{4}, \tfrac{-\pi}{4}, 0). \end{aligned} \tag{10}$$

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

$$\sqrt{\text{SWAP}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{e^{-i\frac{\pi}{4}}}{\sqrt{2}} & \frac{e^{i\frac{\pi}{4}}}{\sqrt{2}} & 0 \\ 0 & \frac{e^{i\frac{\pi}{4}}}{\sqrt{2}} & \frac{e^{-i\frac{\pi}{4}}}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{11}$$

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

$$GS_{i,j}(\theta) = \sqrt{\text{SWAP}}_{i,j} e^{i\frac{\theta}{2}Z_i} e^{-i\frac{\theta}{2}Z_j} \sqrt{\text{SWAP}}_{i,j}. \tag{12}$$

This is similar to the decomposition of a Givens rotation gate into two $\sqrt{I\text{SWAP}}$ and z-rotation gates [4]: 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

$$\text{SWAP} = \left(e^{-i\frac{\pi}{4}(YI+IY)} \sqrt{\text{iSWAP}} e^{i\frac{\pi}{4}(YI+IY)}\right) \sqrt{\text{iSWAP}} \left(e^{i\frac{\pi}{4}(XI+IX)} \sqrt{\text{iSWAP}} e^{-i\frac{\pi}{4}(XI+IX)}\right) \tag{13}$$

$$= \left(\exp\left[i\tfrac{\pi}{8}(YY+ZZ)\right]\right) \exp\left[i\tfrac{\pi}{8}(XX+YY)\right] \left(\exp\left[i\tfrac{\pi}{8}(XX+ZZ)\right]\right). \tag{14}$$

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

$$G(\alpha) = \exp\left[i\tfrac{\pi}{8}(\cos(\alpha)YY + \sin(\alpha)XY)\right] \exp\left[i\tfrac{\pi}{8}(XX+YY+2ZZ)\right] \exp\left[i\tfrac{\pi}{8}(\cos(\alpha)XX - \sin(\alpha)XY\right]. \tag{15}$$

This function can be expanded to give

$$G(\alpha) = g_1(\alpha)II + g_2(\alpha)(XX+YY) + g_3(\alpha)(IZ-ZI) + g_4(\alpha)ZZ, \tag{16}$$

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

$$\begin{aligned} g_1(\alpha) =& \frac{1}{\sqrt{2}} \Big\{ \cos^2(\tfrac{\pi}{8})(\cos^2(\tfrac{\pi}{8}) + i\sin^2(\tfrac{\pi}{8})) + i\cos(\alpha)\cos(\tfrac{\pi}{8})\sin(\tfrac{\pi}{8})\frac{(1+i)}{\sqrt{2}} \\ & + \cos^2(\alpha)\sin^2(\tfrac{\pi}{8})(i\cos^2(\tfrac{\pi}{8}) + \sin^2(\tfrac{\pi}{8})) + \sin^2(\alpha)\sin^2(\tfrac{\pi}{8})(\cos^2(\tfrac{\pi}{8}) + i\sin^2(\tfrac{\pi}{8})) \Big\} \end{aligned} \tag{17}$$

$$g_2(\alpha) = \frac{1+i}{4\sqrt{2}}(1+\cos(\alpha)) \tag{18}$$

$$\begin{aligned} g_3(\alpha) =& \frac{1}{\sqrt{2}} \Big\{ \sin(\alpha)\sin(\tfrac{\pi}{8})\cos(\tfrac{\pi}{8})\frac{(1+i)}{\sqrt{2}} + i\sin(\alpha)\cos(\alpha)\sin^2(\tfrac{\pi}{8})(\cos^2(\tfrac{\pi}{8}) + i\sin^2(\tfrac{\pi}{8})) \\ & - i\cos(\alpha)\sin(\alpha)\sin^2(\tfrac{\pi}{8})(i\cos^2(\tfrac{\pi}{8}) + \sin^2(\tfrac{\pi}{8})) \Big\} \end{aligned} \tag{19}$$

$$\begin{aligned} g_4(\alpha) =& \frac{1}{\sqrt{2}} \Big\{ \cos^2(\tfrac{\pi}{8})(i\cos^2(\tfrac{\pi}{8}) + \sin^2(\tfrac{\pi}{8})) - i\cos(\alpha)\cos(\tfrac{\pi}{8})\sin(\tfrac{\pi}{8})\frac{(1+i)}{\sqrt{2}} \\ & + \cos^2(\alpha)\sin^2(\tfrac{\pi}{8})(\cos^2(\tfrac{\pi}{8}) + i\sin^2(\tfrac{\pi}{8})) + \sin^2(\alpha)\sin^2(\tfrac{\pi}{8})(i\cos^2(\tfrac{\pi}{8}) + \sin^2(\tfrac{\pi}{8})) \Big\}. \end{aligned} \tag{20}$$

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 $G(\alpha)$,

$$\begin{aligned}\langle 01|G(\alpha)|01\rangle &= g_1(\alpha) - g_4(\alpha) + 2g_3(\alpha) \\ &= A(\alpha)e^{i\phi(\alpha)}e^{i\pi/4} \qquad (21)\\ \langle 10|G(\alpha)|10\rangle &= g_1(\alpha) - g_4(\alpha) - 2g_3(\alpha) \\ &= -A(\alpha)e^{-i\phi(\alpha)}e^{i\pi/4}, \qquad (22)\end{aligned}$$

which can be removed by shifting

$$G(\alpha) \to e^{i\phi/4(IZ-ZI)}G(\alpha)e^{i\phi/4(IZ-ZI)}. \qquad (23)$$

The precise value of ϕ here is

$$\phi = -\arctan\left\{ \frac{\sqrt{2}\left(\cos^2(\frac{\pi}{8}) - \sin^2(\frac{\pi}{8})\cos(2\alpha) - \frac{1}{\sqrt{2}}\cos(\alpha)\right)}{\sin(\alpha) + \sin(2\alpha)\sin^2(\frac{\pi}{8})}\right\}. \qquad (24)$$

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 $G(\alpha) \to e^{i\pi/2ZI}e^{i\pi/2IZ}G(\alpha)e^{i\pi/2ZI}e^{i\pi/2IZ}$ (again at no extra cost), and solve for $\sin(\theta) = 2g_2(\alpha)$.

C. 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

$$U(\theta) = \exp\left\{\sum_{p\in\text{unocc},q\in\text{occ}} \theta_{pq}a^\dagger_{p\alpha}a^\dagger_{p\beta}a_{q\alpha}a_{q\beta} - \text{h.c.}\right\}, \qquad (25)$$

which when mapped to qubits in the S_0 approximation, becomes

$$U(\theta) = \exp\left\{\sum_{p\in\text{unocc},q\in\text{occ}} \theta_{pq}X_pY_q - \text{h.c.}\right\}. \qquad (26)$$

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 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 of the methods and materials, 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. (26). 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. 1[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 $Z_i, Z_j \to D^+_{ij}, D^-_{ij}$, where we define

$$D^\pm_{ij} = \frac{1}{2}\Big[Z_i + Z_j \pm \big(X_iX_j + Y_iY_j\big)\Big] \qquad (27)$$

As we implement our ansatz on a $2\times N/2$ grid, qubit i is not only connected to qubit $(i\pm1)\%N$; 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 $i \to (i+k)\%N$; 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. 1[bottom]. Code that implements this scheduling has also been uploaded to ReCirq [5].

D. Scheduling of EV circuits

In this section we outline the additional experimental details required to implement EV in this work. We implemented control-free EV for $O = Z_i$, or Z_iZ_j, or D_{ij}^{+}(Eq. (27)), 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 on the cat state $\frac{1}{\sqrt{2}}(|0000\ldots\rangle + |0101\ldots\rangle)$. 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. 9 of the methods and materials 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 of the methods and materials, 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 $12N^2$ $(6N^2 + 6N)$

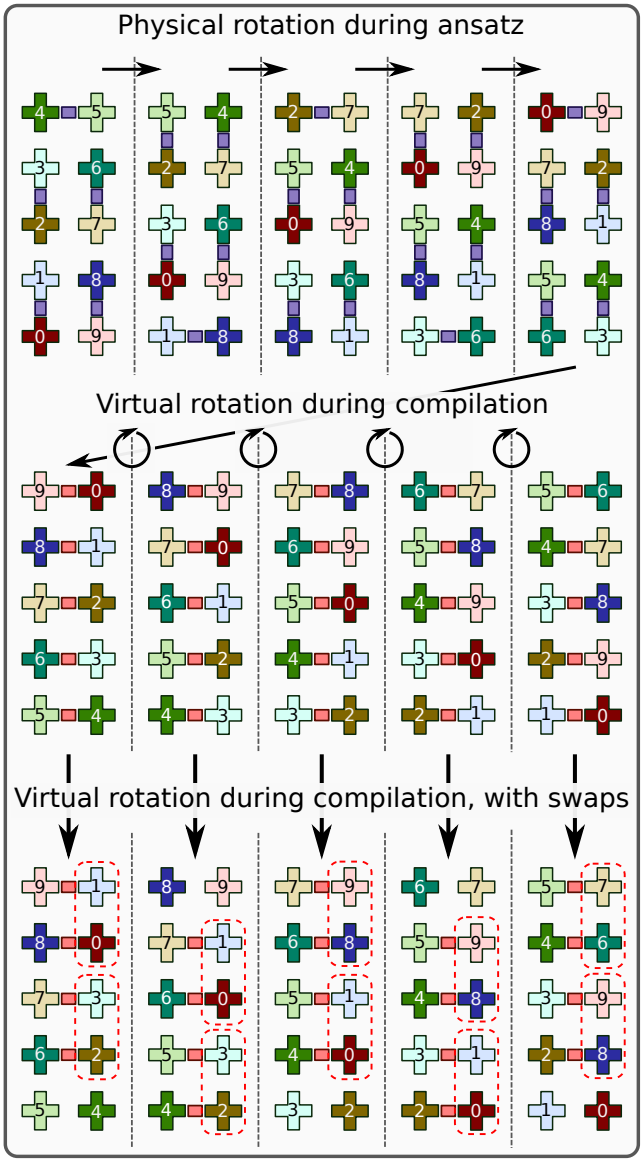


FIG. 1. Detail of the scheduling of pair excitation interactions during the UpCCD ansatz, and measurement scheduling. (top) The physical rotations during the execution of SWAP layers of the UpCCD ansatz, showing that all occupied (even index) and all unoccupied (odd index) orbitals are coupled (purple 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.

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. IV B).

III. DATA PROCESSING

A. 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

$$H = \sum_i c_i Q_i, \tag{28}$$

and estimate the expectation value of each such term

$$E = \text{Trace}[H\rho] = \sum_i c_i \text{Trace}[Q_i\rho]. \tag{29}$$

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

$$\Sigma_{i,j} = \text{Covar}\Big[\text{Trace}[Q_i\rho], \text{Trace}[Q_j\rho]\Big]. \tag{30}$$

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. (27)), 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. (29)

$$\text{Var}(E) = \sum_{i,j} c_i \Sigma_{i,j} c_j, \tag{31}$$

subject to Eq. (28) 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

$$\sum_i c_i q_{i,j} = h_j. \tag{32}$$

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

$$\mathcal{L} = \sum_{i,j} c_i \Sigma_{i,j} c_j - \sum_j \Big(\sum_i c_i q_{i,j} - h_j\Big)\lambda_j. \tag{33}$$

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

$$2\sum_j \Sigma_{i,j} c_j - \sum_j q_{i,j}\lambda_j = 0 \to c_j. \tag{34}$$

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

$$\mathbf{c} = \frac{1}{2}\Sigma^{-1} q \boldsymbol{\lambda}. \tag{35}$$

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

$$\mathcal{L} = -\frac{1}{4}\boldsymbol{\lambda}^T q^T \Sigma^{-1} q \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \mathbf{h}. \tag{36}$$

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

$$q^T\Sigma^{-1}q\boldsymbol{\lambda} = 2\mathbf{h} \to \boldsymbol{\lambda} = 2(q^T\Sigma^{-1}q)^*\mathbf{h} \tag{37}$$

$$\to \mathbf{c} = \Sigma^{-1}q(q^T\Sigma^{-1}q)^*\mathbf{h}, \tag{38}$$

where here, $*$ denotes the Moore-Penrose pseudoinverse of a matrix. The resulting contribution to the energy variance can be calculated to be

$$\text{Var}(E) = \mathbf{h}^T(q^T\Sigma^{-1}q)^*\mathbf{h}. \tag{39}$$

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

$$\mathbf{c} = q(q^Tq)^*\mathbf{h}. \tag{40}$$

The above theory neglects the allocation of shots to individual experiments, which changes $\Sigma_{i,j}$ and can be additionally optimized over to minimize Eq. (39). In our experiment we did not allocate shots except during EV, where the basis Q_i was chosen to be linearly independent, allowing for the use of standard results [6, 7] (see App. IV A). In numerical simulations (App. IV C), a decomposition depending only on $\{|c|_i\}_i$ was chosen in advance (see again App. IV A).

B. 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

$$\text{Covar}[\langle P_i\rangle\langle P_j\rangle] = \frac{\langle P_iP_j\rangle - \langle P_i\rangle\langle P_j\rangle}{M}. \tag{41}$$

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

$$\frac{\partial\Delta}{\partial\langle n_{j\sigma}\rangle} = \frac{1}{2N}\frac{1-2\langle n_{j\sigma}\rangle}{\sqrt{\langle n_{j\sigma}\rangle - \langle n_{j\sigma}\rangle^2}} \tag{42}$$

$$\text{Var}[\Delta] = \sum_{j\sigma}\frac{1}{2N}\frac{(1-2\langle n_{j\sigma}\rangle)^2}{\langle n_{j\sigma}\rangle - \langle n_{j\sigma}\rangle^2}\text{Var}[\langle n_{j\sigma}\rangle]. \tag{43}$$

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

The experimental covariances (Eq. (41) 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. 1 and Fig. 2 of the main text. 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. [8] 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 [9]. 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

$$\frac{M_0}{M} = 1 - |\langle U\rangle|^2, \tag{44}$$

where U is the operator to be estimated via EV, and $M = M_+ + M + - + M_0$. Assuming the fraction $1 - F_{\text{EV}}$ fails verification, we have $M_0 = M(1 - F_{\text{EV}}|\langle U\rangle|^2)$, 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 $M_0 \to M_0 - 2\min(M_+, M_-)$, $M_+, M_- \to M_+ + \min(M_+, M_-), M_- + \min(M_+, M_-)$.)

IV. 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 total time spent on the device, or *wall-clock time*. We allocate shots M (Eq. (41)) using either an even distribution or established techniques [6, 7] to minimize the final energy variance (Eq. (39)). This requires first that we fix a decomposition of our Hamiltonian into simultaneously measurable terms. This allows us to compare (Fig. 3 of the main text (top)) between optimal allocations and those used in the experiment. We also develop a wall clock model that captures overhead in the experiment. This work serves as input to the tuning of hyperparameters of the variational optimizers we present in App. V.

A. Hamiltonian decomposition schemes

In this section we define the different options chosen to decompose a Hamiltonian into terms Q_i for measurement (Eq. (28)). 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. II C. This mostly matches the scheme used for the estimation of expectation values Raw VQE, PS VQE and PS VD in Fig. 1 of the main text, and the scheme used for the PS VQE estimates in Fig. 2 of the main text.

- $XX + YY + IZ + ZI$ — here, we draw $Q_i \in \{Z_i, Z_iZ_j, D_{ij}^+\}$ ($D_{ij}^\pm$ is defined in Eq. (27)), 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 [8]. 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.(29)) 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 $X_j X_k + Y_j Y_k$ and combined using the techniques outlined in App. III A. 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. 3 of the main text 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. II D, when performing EV to estimate each Q_i, 12 unique circuits were run to cancel out a background magnetic field and depolarize readout noise.

B. 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 (41), 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 IV C).

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. (29). 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 [10]. (Note that as the three measurement strategies measure only linearly-independent operators, the $|c_i|$ can be fixed prior to measurement.)

C. Wall-clock model

The primary cost of a quantum algorithm is the wall-clock time actually spent on the device. Due to experimental overhead in loading circuits and returning results, this is not a function of just the total shot count. Circuits on sycamore hardware are run in batches; with each circuit in each batch being additionally repeated multiple times (see the cirq.Sampler.run_batch function in Ref. [11]). We find a good wall-clock time model depends on three parameters: the number a of circuit batches sent to the device, the number b of circuits across all batches, and the number of repeat experiments c. We attribute this respectively to the latency in sending and receiving batches from the device, the latency in switching between circuits on the device, and the execution of each circuit. Our final equation for the wall-clock time is a linear combination of the three parameters with empirically-estimated coefficients,

$$t_{\text{wall}} = a * 1s + b * 0.042s + c * 5 * 10^{-5} s. \tag{45}$$

Note that we do not consider effects of drift (or the need to re-calibrate) in this model; error rates are constant as a function of wall-clock time.

D. Wall-clock time extrapolation to large system sizes

Using the protocol described in the previous sections, we compute the cost of estimating the energy of the RG model at $g = -0.9$ in the state prepared by the Up-CCD ansatz with pre-optimized parameters to within a shot noise standard deviation of 0.1. (We do not yet include the increase in this standard deviation due to experimental error.) In Fig. 2, we plot this cost in terms of the number of shots (top) and in wall-clock time for superconducting hardware (bottom) for the Hamiltonian decompositions described in App. IV A and shot distribution schemes described in Section IV B. Computations are performed for $N = 6, 8, 10, 12$, which are well fit by a power law to extrapolate to larger system sizes. (This extrapolation assumes that the coefficients in Eq. (45) are independent of N.) As only shot noise was simulated, VQE and EV yield the same resource estimates.

To process the data of Fig. 2 to give Fig. 3[top] of the main text, we divide it by the 10-qubit experimental fidelity estimates F from Fig. 3[bottom-right] of the main text. This corresponds to assuming that the circuit fidelity is maintained as we go to larger system sizes, which requires decreasing error rates correspondingly. We argue in the main text that this will be necessary to prevent an exponential blow-up of the shot requirements studied here. The results in Fig. 3[top] of the main text do not account for any residual bias post-error mitigation [12], but this by definition cannot be fixed by increasing the runtime of the experiment.

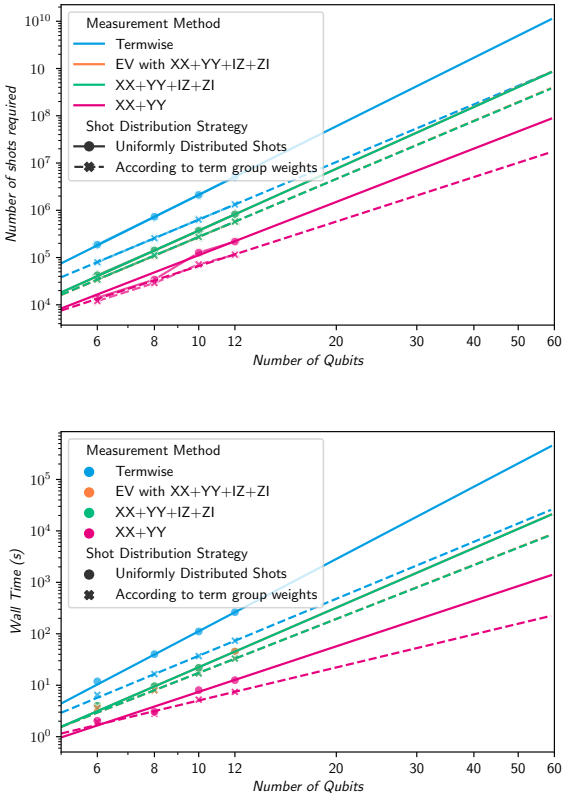


FIG. 2. Costs of estimating the ground-state energy of the RG Hamiltonian at $g = -0.9$ to within 0.1a.u.. (Top) Shot-count estimates as a function of the number of qubits, obtained by standard Lagrangian optimization [7] using term distributions defined in App. IV B and measurement strategies defined in App. IV A. (Bottom) Wall-clock time estimates, calculated from the data from (top) and Hamiltonian decomposition details, using the model described in App. IV C. The orange and green data points for the $XX + YY + IZ + ZI$ scheme with and without EV coincide.

V. VARIATIONAL OPTIMIZATION OF A 6-QUBIT EXPERIMENT

A. 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. IV C), it is beneficial if the optimizer can request a batch of cost-function evaluations at once before making a step. In [13], 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 [14]. For the special case of quadratic cost functions over an n dimensional space, conjugate gradient methods can be proven to converge in n iterations [15]. 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 [16] 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 γ, sample radius δ, maximum iterations n, number of evaluations per iteration k, rate decay exponent α, stability constant A, sample radius decay exponent ξ, tolerance ε, 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 δ'-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

B. 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 [17] or model-based [4] optimizers, including the Model Gradient Descent optimizer that we have based our conjugated Model Gradient Descent algorithm upon in the previous section.

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:

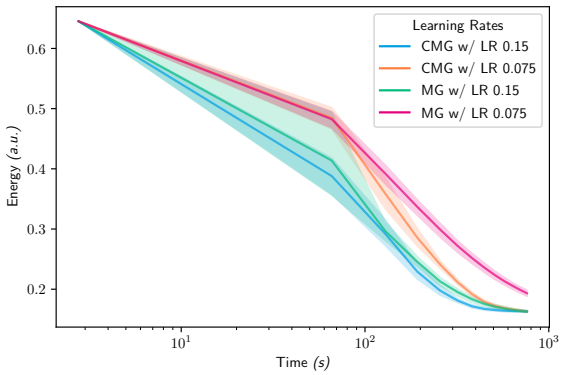


FIG. 3. In this figure, the distance between the energy that was calculated in the noiseless regime by the L-BFGS-B optimizer is plotted over the wall-time required to reach such resolution. The system displayed is a RG Hamiltonian for 6 qubits, with a coupling constant of $g = -0.9$, consisting of 10 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 [13].

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 3 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. 4 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.

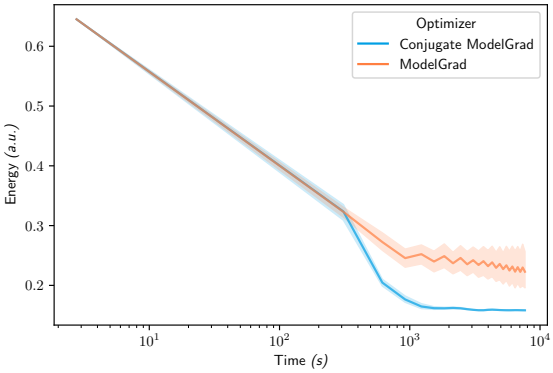


FIG. 4. Comparison of the two variants of the Model Gradient Descent optimizer, where in this case the number of samples used to construct the quadratic model one uses is higher, namely $k = 4(N+1)(N+2)$, allowing for the performance difference of the two optimizers to become more pronounced, with Conjugate 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.

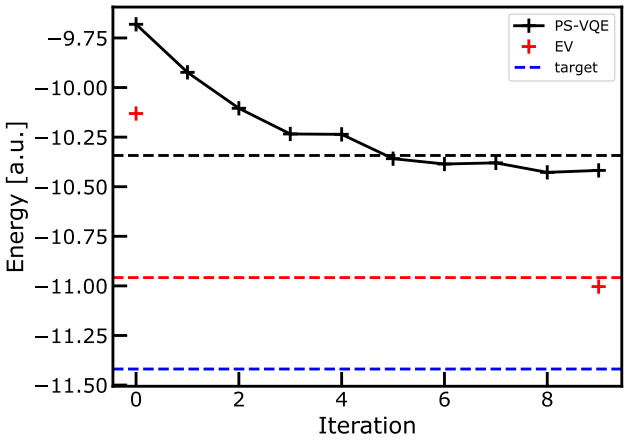


FIG. 5. Optimization trace of a 6 qubit RG simulation at $g = -0.9$, targeting the blue dashed line (ground state energy in the seniority-zero space). After an energy estimation using postselection (black dashed line) and echo verification (red dashed line) at optimized parameters from a noiseless simulation, 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).

C. 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 [18, 19]. 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. 8, which we attribute to poor calibration on the day. If the absolute gain in energy was replicated in Fig. 8, 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.

VI. ADDITIONAL RESULTS AND ANALYSIS

A. Virtual distillation with and without postselection

In this appendix, we show the effect that postselection has on virtual distillation. In Fig. 6, we repeat the plot of Fig. 1 of the main text, 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 $|\langle D_{ij}^{\pm} \rangle| \geq 1$. (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 $\langle D_{ij}^{\pm} \rangle$ to lie in $[-1, 1]$. Performing this (Fig. 6, 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

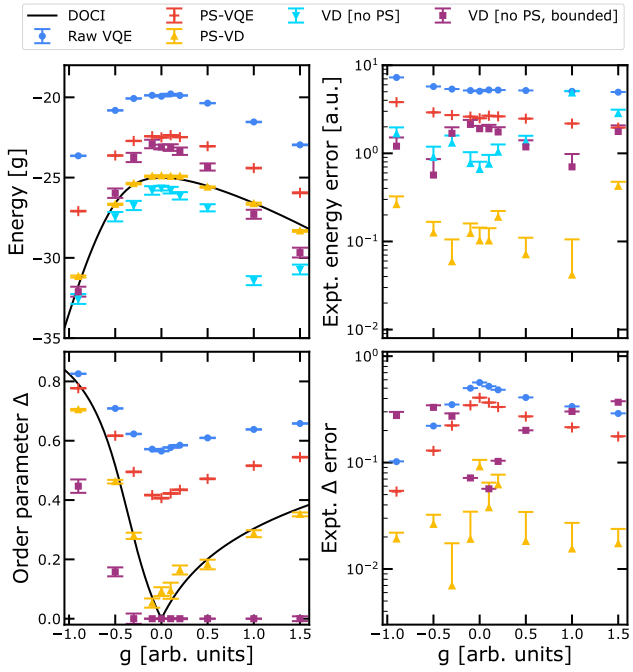


FIG. 6. Comparison of virtual distillation with postselection (yellow), without postselection (light blue), and without postselection but while bounding expectation values of Z_j and $D_{ij}^{\pm}$ in $[-1,1]$ (purple), for a 10-qubit simulation of ground states of the RG Hamiltonian across a range of g values. Raw VQE (blue), postselected VQE (red), and post-selected VD [PS VD] data same as in Fig. 1 of the main text. VQE error bars obtained by error propagation (1 standard deviation, see App. III B), VD error bars obtained by bootstrapping (1 standard deviation).

of generated estimates [7].) 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 incoherent 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

$$\rho_2 = \sum_{p,q,r,s} c_{p,q,r,s} |p\rangle\langle q| \otimes |r\rangle\langle s|, \tag{46}$$

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

$$\begin{aligned}&\mathrm{Tr}\big[\rho_2 S^{(2)}(Z_j \otimes I + I \otimes Z_j)\big] \\ &= \sum_{p,q,r,s} c_{p,q,r,s}\delta_{s,p}\delta_{r,q}\Big(\langle s|Z_j|s\rangle + \langle r|Z_j|r\rangle\Big),\end{aligned} \tag{47}$$

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 decoherent noise channels such as a T1 channel (with identical error rates for each qubit) will be completely mitigated as these off-diagonal elements do not exist. Coherent noise may cause some of these off-diagonal elements to appear, 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 VD can be first-order sensitive.) The above analysis has assumed that the postselection is perfect; readout noise would complicate the above.

B. Comparison of purification to postselection

A link between EV and VD has been previously established in E.g. [20, 21]. Both methods try to consume two approximate copies $\rho_1, \rho_2 \sim |\psi\rangle\langle\psi|$ (with $|\psi\rangle$ the target state) to estimate expectation values w.r.t. $\{\rho_1, \rho_2\}/2\mathrm{Trace}[\{\rho_1, \rho_2\}]$. (Here, $\{\cdot,\cdot\}$ is the anti-commutator.) The two copies are prepared in slightly different methods; EV prepares the second copy via an inverse preparation circuit (in this formalism), while VD uses two separate registers, so we do not expect $\rho_1 = \rho_2$ precisely. The error-mitigated estimator of the expectation value of some observable O then takes the form [20]

$$\widetilde{\langle O\rangle}_{\mathrm{pur}} = \frac{\mathrm{Trace}[O\{\rho_1, \rho_2\}]}{\mathrm{Trace}[\{\rho_1, \rho_2\}]} \tag{48}$$

This is similar in form but not identical to the estimator of an expectation value estimated following postselection on some symmetry projector P [22],

$$\widetilde{\langle O\rangle}_{\text{sym}} = \frac{\text{Trace}[OP\rho P]}{\text{Trace}[P\rho P]}. \tag{49}$$

In the main text we observed that projection on symmetries of a problem resulted in a far smaller error than purification. In this section we expand on the differences between the above two expressions in some example cases. We note that this connection has been studied before and generalized in Ref. [23].

It is simple to construct toy examples where postselection greatly improves on purification and vice-versa. Purification does not correct for coherent error, so in principle any coherent error that breaks a symmetry is better-protected by verifying that symmetry. As a simple toy example, consider the state $|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ subject to a single-qubit Y rotation on qubit 0 by a small angle α. This maps $|\psi\rangle$ to

$$|\tilde{\psi}\rangle \to \frac{\cos(\alpha)}{\sqrt{2}}(|01\rangle + |10\rangle) + \frac{\sin(\alpha)}{\sqrt{2}}(|11\rangle - |00\rangle). \tag{50}$$

One can see that in this case (assuming identical errors on both copies of $\rho_1 = \rho_2 = |\tilde{\psi}\rangle\langle\tilde{\psi}|$), purification has no effect whatsoever, while postselection on number perfectly corrects the error. On the other hand, postselection has no effect on errors that commute with the projector P. Incoherent phase noise on qubit 0 with probability p sends the above state $|\psi\rangle$ to

$$\rho = (1-p/2)|\psi\rangle\langle\psi| + p/2|\bar{\psi}\rangle\langle\bar{\psi}|, \tag{51}$$

where $|\bar{\psi}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ is the orthogonal state in the $|10\rangle, |01\rangle$ subspace. Purification here suppresses error to $\sim p^2/4$ (up to a normalization factor), but if P projects onto the half-filling subspace, $P\rho P = \rho$, and postselection has no effect.

It is natural to ask at this point how the above examples relate to the performance observed in the main text; why do we not see a significant contribution from the coherent rotation example given? One main reason for this is likely that incoherent noise plays a larger role in superconducting hardware than coherent error; most gates are decoherence-limited. We also expect PS-VD to suppress coherent errors that break postselection to at least second order (following the discussion in App. VI A). In the above example (Eq. (50)), postselecting two copies of the state $|\tilde{\psi}\rangle$ into the 2-particle sector yields

$$\frac{1}{Z}\Big(\cos^2(\alpha)|\psi\rangle|\psi\rangle - \sin^2(\alpha)\big[|1100\rangle + |0011\rangle\big]\Big), \tag{52}$$

with Z a normalization constant. The product of the swap operator $S^{(2)}$ and any number-conserving operator $O\otimes I$ does not mix the $|\psi\rangle|\psi\rangle$ and $[|1100\rangle + |0011\rangle]$ states, so any corrections to the expectation values are of the order $\sin^4(\alpha)$. (The coefficient $\frac{1}{Z}\cos^2(\alpha)$ is accounted for in the normalization by Trace$[\rho^2]$.) This justifies our claim that coherent error is suppressed quadratically (corrections from this error channel to the expectation value of a number-conserving operator can be seen from Eq. (50) to be $\sin^2(\alpha)$). PS-VD is still not as powerful in this case as postselection, which removes the error exactly. However, it implies that sources of coherent, number-conserving error need to be far larger than all sources of incoherent error for PS-VQE to outperform PS-VD in mitigating a real experiment.

We can generalize the above under some assumptions. Let us diagonalize our noisy state as

$$\rho = \sum_j p_j |\rho_j\rangle\langle\rho_j|. \tag{53}$$

Let us assume that incoherent noise does not bias us towards a particular state; i.e. for $j > 1$, $p_j \sim 2^{-N}$. This implies that $\rho^2/\text{Trace}[\rho^2] = |\rho_0\rangle\langle\rho_0| + \mathcal{O}(2^{-N})$. The main effect of coherent noise is to shift the dominant eigenvector $|\rho_0\rangle$ away from the target state $|\psi\rangle$. Let us write

$$|\rho_0\rangle = a|\psi\rangle + b|\tilde{\psi}_P\rangle + c|\tilde{\psi}_{\bar{P}}\rangle, \tag{54}$$

where $|\tilde{\psi}_P\rangle$ are the corrections that do not break postselection, and $\tilde{\psi}_{\bar{P}}$ are the corrections that do. The condition for PS-VQE to outperform PS-VD in the large-N limit is then roughly

$$c^4 \geq \sum_{j>0} p_j^2 \langle\rho_j|P|\rho_j\rangle. \tag{55}$$

(In the absence of postselection, c^4 can be replaced by c^2.)

C. Distribution of errors in Pauli expectation value estimation

In Fig. 7, we plot a histogram of the error in estimating the expectation values of Pauli operators, taking data from Fig. 1 of the main text, Fig. 8, and Fig. 2 of the main text. 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 the main text 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 $O(N^2)$; 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

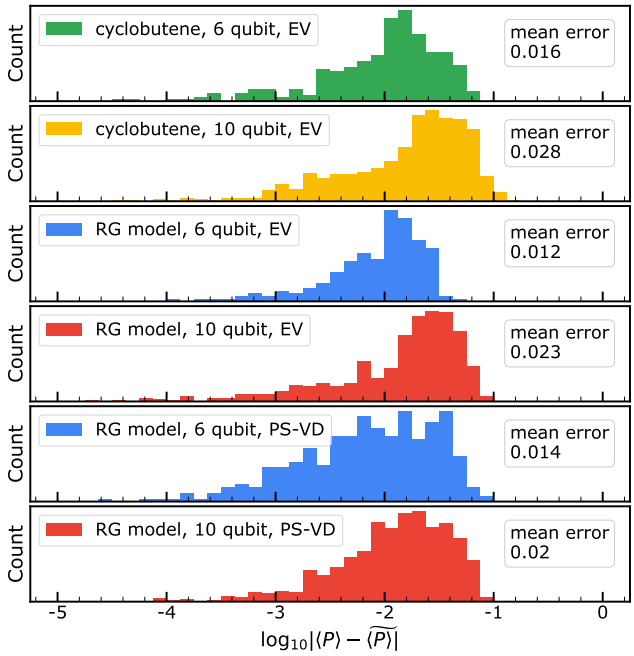


FIG. 7. Histogram of the expectation value error in Pauli operators $P = Z_j$, $P = Z_i Z_j$ or $P = D^{\pm}_{ij}$ (Eq. (27)), across all data taken for the experiment mentioned. Mean error across the dataset is given.

gain in energy of $2.5-3.3\times$, which lies in between the observation of Fig. 2 of the main text and Fig. 3 of the main text. 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.

D. Smaller studies of the RG Hamiltonian

In Fig. 8, 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. 1 of the main text, save for the number of qubits and shots used. Aggregated data from these figures was used to generate the scaling plots of Fig. 3 of the main text.

[1] 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).

[2] 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).

[3] 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.

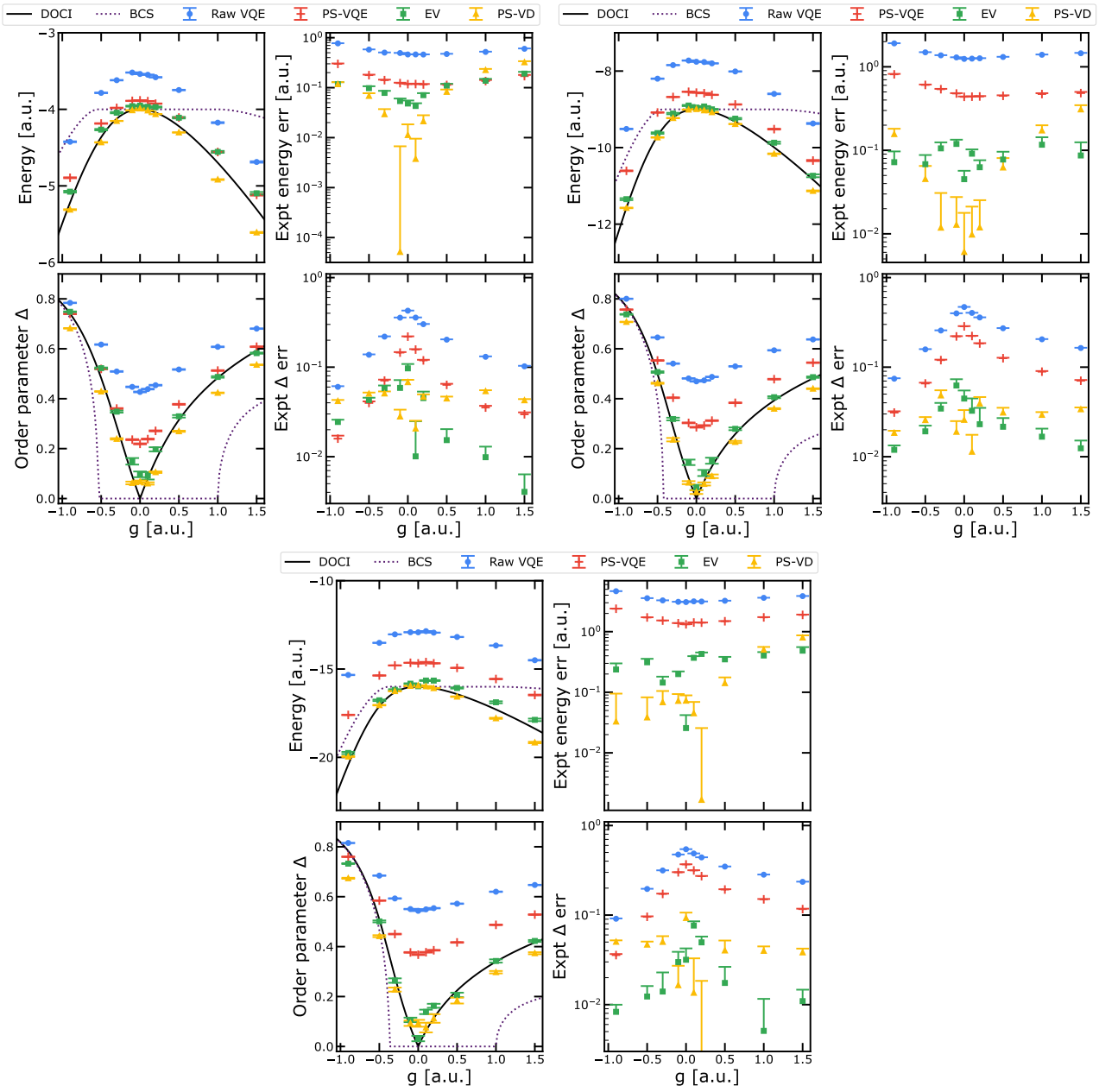


FIG. 8. Identical experiment to Fig. 1 of the main text, but for 4 (top-left), 6(top-right), and 8 (bottom) qubits instead of 10. Aggregate data is used in Fig. 3 of the main text. (top-left) Energy plot for the RG system as a function of the coupling parameter g, for an unmitigated state preparation [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).

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).

[4] 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).

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

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

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

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

[9] 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).

[10] 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).

[11] Cirq Developers, Cirq.

[12] 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).

[13] 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).

[14] 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).

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

[16] 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.

[17] 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 fermi-hubbard model using a scalable algorithm on a quantum computer, ArXiv:2112.02025 (2021).

[18] 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).

[19] 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).

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

[21] Z. Cai, Resource-efficient purification-based quantum error mitigation, ArXiv:2107.07279 (2021).

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

[23] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo, Generalized quantum subspace expansion, Phys. Rev. Lett. **129**, 020502 (2022).

4 Relaxed Reduced Density Matrices and Energy Gradient within Quantum Krylov Subspace Diagonalization Framework

In this section, we shall propose an efficient (in terms of the asymptotic scaling of the depth) method to calculate the first-order total derivatives of QKSD energies. We start by introducing preliminary results namely we present the QKSD algorithm, and then layout the the cost that arises to a quadratic scaling (in the Krylov subspace dimension D) when computing the gradient of the QKSD energy. Then, we introduce quantum signal processing which is used to realize the circuits in our algorithm. We conclude by presenting our approach that reduces the scaling to O(1) distinct circuit per relaxed RDM.

4.1 Quantum Krylov Subspace Diagonalization

The quantum Krylov subspace diagonalization is a set of methods that construct a subspace, called the Krylov subspace $\mathcal{K} \coloneqq span(\{|\phi_k\rangle, k \in \{0, \dots, D-1\}\})$ using a set of states and outputs the lowest energy achievable in said subspace as the approximation of the ground state energy. To draw an analogy to VQE, the normalized linear combination of the generating state is the ansatz, where the linear combination coefficients are the ansatz parameters. However, in the QKSD the global energy minimum with respect to the ansatz parameters is analytically determined as we shall present. However, first, there are several methods to generate different types of states that span the Krylov subspace. For instance in [16], time-evolution under the Hamiltonian with equally spaced time stamps is used to generate the set of Krylov states $\{|\phi_k\rangle \coloneqq \exp -ik\Delta t|\phi_0\rangle, k \in \{0, \dots, D-1\}\}$. Similarly in [40] imaginary time evolution is used. In [79] other functions of the Hamiltonian are used to construct the Krylov subspace. The choice of one or the other depends on characteristics including but not limited to, the number of samples for a fixed accuracy ϵ, the complexity of the circuit preparing the Krylov states $|\phi_k\rangle$, of each method.

Before carrying on, let us first introduce some notations

$$T_m(x) \coloneqq \cos(m \arccos(x)), |x| \leq 1 \tag{74}$$

is the Chebychev polynomial of degree m defined over the disc of radius 1 and any matrix $A \in \mathbb{C}^{s\times s}$ with eigendecomposition $A = U \begin{pmatrix} a_0 & & & \\ & a_1 & & \\ & & \dots & \\ & & & a_{s-1} \end{pmatrix} U^{-1}$ and real spectrum $\{a_k \in \mathbb{R}, k \in \{0, \dots, s-1\}\}$ such that $\|A\| \leq 1$, we define

$$T_m(A) \coloneqq U \begin{pmatrix} T_m(a_0) & & & \\ & T_m(a_1) & & \\ & & \dots & \\ & & & T_m(a_{s-1}) \end{pmatrix} U^{-1}. \tag{75}$$

In our work we considered the Chebychev polynomials of the Hamiltonians to span the Krylov subspace where

$$\hat{H} \coloneqq \hat{\mathcal{H}}/\lambda \tag{76}$$

$$|\psi_k\rangle \coloneqq T_k(\hat{H})|\psi_0\rangle, k \in \{0, \dots, D-1\} \tag{77}$$

$$\mathcal{K} \coloneqq span(\{|\psi_k\rangle, k \in \{0, \dots, D-1\}\}). \tag{78}$$

This choice stems from the fact that the Chebychev polynomials can be prepared with no approximation error at finite depth contrary to time evolution for instance or other non polynomial functions. This leads to a very important property which preserving the Hankel form of the very important matrix used in of the QKSD namely, the matrix of overlaps of the Krylov states with respect to the Hamiltonian that

$$\widetilde{H}_{ij} \coloneqq \langle\phi_i|\hat{H}|\phi_j\rangle. \tag{79}$$

Hence, such matrix can be constructed with a linear number of entries as opposed to the other cases where a quadratic number of elements is needed. Furthermore, in [29] a more efficient method, that bypasses Hadamard tests to construct $\widetilde{H}$ and $\widetilde{S}$, is proposed. Before presenting how the global minimum energy within the Krylov subspace $\mathcal{K}$ is determined, we first define

$$\widetilde{S}_{ij} = \langle\phi_i|\phi_j\rangle \tag{80}$$

which similarly has a Hankel form. The optimal energy and state within the Krylov subspace K is obtained by solving the general eigenvalue problem

$$\widetilde{H}c^m = E_k\widetilde{S}c^m \tag{81}$$

where E_0 and α^0 are the optimal energy and its corresponding ansatz parameters. The process flow is described in the quantum Krylov box in FIG 4. The S matrix can be ill conditioned which makes solving the general eigenvalue in a numerically stable fashion a challenging task. This is often resolved with thresholding where all eigenvalues of the $S = U^\dagger \Lambda U$ matrix below a given threshold are discarded before solving Eq (81). The threshold value scales as $O(\sqrt{M}^{-1})$ [29] where M is the shot-budget for estimating each entry in H and S. One of the influential factors on the complexity of the circuits in the QKSD is the dimension of the quantum krylov subspace D as it directly influence the max-depth of the method. The scaling of the dimension D of the Krylov subspace as a function of the ground state energy precision is hence an important property. As shown in [2, 63, 19, 32], the Krylov method converges rapidly with the Krylov subspace dimension D. However, due to noise in the evaluation of the matrices $\widetilde{H}$ and $\widetilde{S}$, resulting from finite samples, for instance, such scaling can be affected and in [29] the upper-bound provided goes from logarithmic to linear scaling in the sought precision.

4.2 First-Order Derivative of QKSD Energies

In [49], we show that calculating the Krylov energy derivative using the "naive" direct approach results in an $O(D^2)$ different circuit per RDM element. To show this, we considered the monomial basis and showed that to calculate the derivative with respect to a given parameter, one would need to evaluate $\frac{d\langle\psi_0|\hat{H}^{i+j+1}|\psi_0\rangle}{d\theta} = \sum_k^{i+j-1}\langle\psi_0|\hat{H}^{k-1}\frac{d\hat{H}}{d\theta}\hat{H}^{i+j+1-k}|\psi_0\rangle$, where $i,j \in [0, D-1]$. To address this issue, we make the key observation that the expansion of the Krylov ground state in $\mathcal{K}$, can be written as a polynomial function of $\hat{H}$ of degree exactly $D-1$ applied to the initial state $|\psi_0\rangle$

$$|\Psi_0\rangle \coloneqq \underbrace{\left(\sum_k^{D-1} c_i^0 T_i(\hat{H})\right)}_{f(\hat{H})}|\psi_0\rangle \coloneqq f(\hat{H})|\psi_0\rangle. \tag{82}$$

Thus, by defining

$$\eta \coloneqq max(|f|, x \in [-1, 1]) \tag{83}$$

f/η can be exactly block-encoded with no systematic error or bias using QSP [35, 17]. Moreover, the QSP circuit has similar resources to that of $T_{D-1}(H)$. We show in [49], that the RDMs γ_{pq} and Γ_{pqrs} can be coherently calculated, that is, with no post-selection, and in $O(1)$ observables (per RDM). This constitutes a saving factor of $O(D^2)$ per relaxed RDM compared to the naive way. Furthermore, we proposed a coherent way of calculating the relaxed RDMs that bypasses the post-selection used in the post-selection-based measurement method and whose variance and shot budget depend on the success probability of the QSP block-encoding. By noting $\mathcal{U}_\Phi$ the unitary that block-encodes $g(\hat{H})$, we show that

$$\gamma_{pq} = \lambda\eta^2(\langle\psi_0|\mathcal{U}_\Phi^*\hat{E}_{pq}\mathcal{U}_\Phi|\psi_0\rangle + \langle\psi_0|\mathcal{U}_\Phi^*\hat{R}\hat{E}_{pq}\mathcal{U}_\Phi|\psi_0\rangle) \tag{84}$$

$$\Gamma_{pqrs} = \lambda\eta^2\left(\langle\psi_0|\mathcal{U}_\Phi^*\hat{E}_{pq}\mathcal{U}_\Phi|\psi_0\rangle + \langle\psi_0|\mathcal{U}_\Phi^*\hat{R}(\hat{E}_{pq}\hat{E}_{rs} - \delta\hat{E}_{qr})\mathcal{U}_\Phi|\psi_0\rangle\right). \tag{85}$$

Our benchmarks show that the variance of the coherent method is around 2x smaller than that of the post-selection method and subsequently the shot-budget for a given accuracy is similarly $2\times$ times less.s

4.3 Molecular Properties from Quantum Krylov Subspace Diagonalization

Bibliographic Information

Oumarou, O., Ollitrault, P. J., Cortes, C. L., Scheurer, M., Parrish, R. M., & Gogolin, C. (2025). Molecular Properties from Quantum Krylov Subspace Diagonalization. arXiv preprint arXiv:2501.05286.

Summary

By virtue of its shallow circuits, the quantum Krylov subspace diagonalization is proposed as a family of NISQ & early fault-Tolerant quantum methods. It produces the optimal energy within a subspace spanned by chosen states (Krylov subspace) by solving the general eigenvalue problem involving the matrices of overlaps

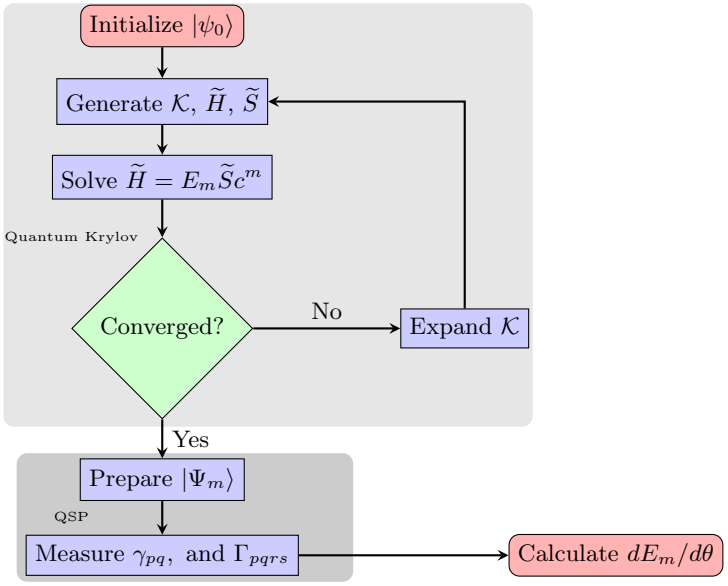


Figure 4: The quantum Krylov box describes the QKSD procedure; The QSP box describes the block-encoding of the function preparing the the quantum Krylov ground state.

and overlaps with respect to the Hamiltonian of the generators. We propose an efficient analytical method for calculating first-order derivatives of energies obtained through the quantum Krylov method. We focus on the relaxed reduced density matrices as the essential non-classical quantity used for obtaining other quantities such as the nuclear gradient. We use quantum signal processing to overcome the bottleneck of calculating the relaxed RDMs, reducing the number of observables from a square dependence on the Krylov subspace dimension $O(D^2)$ to a constant $O(1)$ per relaxed RDM. Furthermore, we propose a coherent method for the measurement of said RDMs alternative to the post-selection procedure that arises naturally from the equations. We perform our benchmark on the H_2O molecular system with 8 spin orbitals and first compare the variance of the nuclear gradient produced by the relaxed RDMs measured with the coherent method and the post-selection method. We note that our coherent approach has a smaller variance by a factor of 2. The last set of benchmarks traces the behavior of an important quantity, namely the normalization factor η, which is the infinity norm of the function expanded by the general eigenvector coefficient in the Chebychev basis, and the distance to the energy of the ground state as a function of the regularization threshold. We conclude that we can maintain the values of η low enough for a better shot budget without sacrificing the energy accuracy.

Contribution

O.O. co-derived the analytical expressions of the relaxed RDM, developed an improved circuit for the coherent measurement approach, implemented the quantum Krylov algorithm and the QSP iterative angle solver code (which has been open-sourced [1] and is now a included as a feature in the open-source quantum toolkit Pennylane), implemented and tested the QSP circuits required for relaxed RDMs computation, generated the numerical benchmark data and co-wrote the manuscript.

[1] https://github.com/PennyLaneAI/pennylane/pull/6694

Molecular Properties from Quantum Krylov Subspace Diagonalization

Oumarou Oumarou [1], Pauline J. Ollitrault [2], Cristian L. Cortes [2], Maximilian Scheurer [1], Robert M. Parrish [2], and Christian Gogolin [1]

[1] *Covestro Deutschland AG, Leverkusen 51373, Germany*
[2] *QC Ware Corp, Palo Alto, USA and Paris, France*

(Dated: January 10, 2025)

Quantum Krylov subspace diagonalization is a prominent candidate for early fault tolerant quantum simulation of many-body and molecular systems, but so far the focus has been mainly on computing ground-state energies. We go beyond this by deriving analytical first-order derivatives for quantum Krylov methods and show how to obtain relaxed one and two particle reduced density matrices of the Krylov eigenstates. The direct approach to measuring these matrices requires a number of distinct measurement that scales quadratically with the Krylov dimension D. Here, we show how to reduce this scaling to a constant. This is done by leveraging quantum signal processing to prepare Krylov eigenstates, including exited states, in depth linear in D. We also compare several measurement schemes for efficiently obtaining the expectation value of an operator with states prepared using quantum signal processing. We validate our approach by computing the nuclear gradient of a small molecule and estimating its variance.

I. INTRODUCTION

Predicting the energies and properties of complex, large molecules remains a challenging yet crucial goal. The advent of quantum computers has introduced a fundamentally new approach to solving this problem, promising feasible time scales for determining the ground-state energy of industrially significant molecules, such as P450 or FeMoco, through the quantum phase estimation (QPE) algorithm [1–10]. While these estimates are promising, they still require substantial quantum resources that are not yet available. Over the past decade, considerable effort has been devoted to minimizing these resource requirements. A prominent direction has been the optimization of key components of QPE, including state initialization [11–14], Hamiltonian simulation algorithms [15–17], the reduction of ancilla qubits [18–21], or optimizing circuit depth and the energy estimator [22]. It is worth noting that the literature on these topics is extensive; here, we cite only a selection of representative works for each of the mentioned area.

Alternatively, efforts to minimize resource requirements have driven the development of entirely new algorithms. These alternatives often exchange QPE's substantial circuit depth for additional incoherent steps and measurement processes, making them more suitable for near-term quantum devices [23, 24]. One particularly promising approach is quantum Krylov subspace diagonalization (QKSD), which maps the Hamiltonian onto a low-dimensional Krylov subspace [25–27]. This method reduces the computational demands while maintaining the accuracy required for practical applications. The quantum computer is used to obtain the matrix elements of the Hamiltonian within this Krylov basis, and a low-dimensional generalized eigenvalue problem is then solved on a classical computer to find the best approximation of the lowest-lying eigenstates and their corresponding energies.

This approach is advantageous for several reasons. First, it requires shallower quantum circuits than QPE while avoiding the challenging parameter optimization problem inherent to other variational methods, such as the variational quantum eigensolver (VQE) [28, 29]. The accuracy of the ground-state approximation is also known to converge rapidly with the subspace dimension [30, 31]. Moreover, the Hamiltonian and overlap matrices within the Krylov subspace exhibit a Hankel structure, meaning that they are defined by a number of elements that grows linearly with subspace dimension. These last two features effectively reduce the number of calls to the quantum processing unit (QPU). Leveraging these benefits, this algorithm has been successfully implemented on a quantum computer with up to 57 qubits, one of the largest quantum computing experiments to solve a quantum many-body problem to date [32].

The most recent theoretical studies of QKSD have largely focused on understanding the impact of noise on the stability and accuracy of the quantum Krylov methods. Ref. [33] addresses error bounds for real-time Krylov algorithms. A linear scaling of ground-state energy error with input noise is demonstrated and methods to manage variational violations are proposed. Complementing this, Ref. [34] investigates the impact of finite sampling errors in QKSD, proposing strategies such as optimal thresholding to mitigate ill-conditioning and improve algorithm resilience. Numerical simulations of the one-dimensional Hubbard model underscores the method's practical accuracy and error predictability.

One point that remains absent from the literature is the calculation of energy derivatives in QKSD, which is crucial for extracting molecular properties beyond just energy values. An important example, which will be the focus of this work, is the first order derivative of the energy with respect to the nuclear coordinates. It enables analyses of the potential energy surface such as geometry optimization or reaction pathway simulation [35–45]. While computing

such derivatives using finite differences is straightforward, this approach often suffers from inefficiency and inaccuracy due to the need for multiple evaluations and susceptibility to numerical instability. This makes analytical approaches essential for practical and accurate quantum chemistry simulations [46–48]. Analytical derivatives have been studied in the context of QPE [46, 49, 50], VQE [48–52] and in extensions of VQE to excited states [47, 52].

In this work, we derive a theoretical framework for obtaining analytical nuclear energy gradients with QKSD via the one and two-particle reduced density matrices (RDMs). We consider a Krylov basis composed of monomial functions of the Hamiltonian, i.e., spanned by applying powers of the Hamiltonian to a fixed reference state. We show that while the number of calls to the QPU scales linearly with the subspace dimension for energy calculations, obtaining a single RDM element naively requires a quadratic number of QPU executions in the Krylov space dimension. We propose an alternative approach that reduces the number of calls to the QPU while keeping the same circuit depth as required for computing energies and discuss its practical implementation.

The rest of this paper is structured as follows: Section II begins by deriving the expressions for QKSD energies and gradients, highlighting the quadratic increase in resources required for gradient estimation. We then discuss a method to circumvent this scaling by directly preparing the QKSD ground-state, detailing the practical implementation using quantum signal processing (QSP) in Section III. In Section IV, we explore two approaches for measuring the RDMs on the prepared QKSD ground-state. Finally, in Section V, we compare the efficiency of these measurement methods through numerical simulations.

II. QUANTUM KRYLOV ENERGY AND GRADIENTS

A. Energies in the Krylov subspace

Let $\hat{H}$ be a given Hamiltonian and $|\psi_0\rangle$ be a fixed reference state such that $|\psi_0\rangle$ is not orthogonal to the eigenstates of $\hat{H}$ that are of interest, e.g., the ground-state. We define the Krylov subspace $\mathcal{K}$ of dimension D as the linear span $\mathcal{K}$ of the Krylov states $|\psi_i\rangle, i = 0, \ldots, D-1$. The Krylov subspace is then defined as:

$$|\psi_k\rangle \coloneqq \hat{H}^k |\psi_0\rangle \tag{1}$$

$$\mathcal{K} \coloneqq \mathrm{span}(\{|\psi_k\rangle\}_{k=0}^{D-1}). \tag{2}$$

The Krylov space spanned by monomials is equivalent to the Krylov space spanned by any family of polynomials of the Hamiltonian of increasing degree that contains all powers up to the maximum power, such as Chebychev polynomials [53], which we will use later. This choice is motivated by the fact that firstly, these polynomial functions can be prepared with no approximation error at finite depth using QSP. Secondly, differentiating the Krylov energy in any polynomial basis reduces at the end to differentiating monomials.

There are alternative ways in which the Krylov space can be constructed. Ref. [25] for instance uses time-evolution under the Hamiltonian of the reference state $|\psi_0\rangle$ for different times $k\Delta t$ such that $|\psi_k\rangle \coloneqq e^{-ikH\Delta t}|\psi_0\rangle$. In a similar fashion, in Ref. [27] the Krylov states are generated using imaginary time evolution. Ideally, one would want the matrix $\langle\psi_i|\hat{H}|\psi_j\rangle, 0 \leq i, j \leq D-1$ to be a Toeplitz matrix, but if approximate, i.e., trotterized, (imaginary) time evolution is used, the Toeplitz property is not necessarily preserved. Other functions of the Hamiltonian that are beyond time evolution (real or imaginary) have also been applied to generate the Krylov subspace [54]. The choice of one or the other is guided by their inherent properties such as number of samples needed for a target accuracy, the complexity of preparing and determining expectation values and overlaps between the Krylov states.

Let $\hat{H} = \sum_j \lambda_j |\lambda_j\rangle\langle\lambda_j|$ be the spectral decomposition of $\hat{H}$ with λ_j (possibly degenerate) and ordered in ascending order. Given an appropriately chosen reference state $|\phi_0\rangle$, the ground-state energy, and in fact also subsets of excited state energies, are approximated by eigenvalues E_m within the Krylov subspace $\mathcal{K}$. These energies can be analytically determined by solving the generalized eigenvalue problem

$$\widetilde{H}c^m = E_m \widetilde{S}c^m \tag{3}$$

where

$$\widetilde{H}_{ij} \coloneqq \langle\psi_i|\hat{H}|\psi_j\rangle = \langle\psi_0|\hat{H}^{i+j+1}|\psi_0\rangle \tag{4}$$

$$\widetilde{S}_{ij} \coloneqq \langle\psi_i|\psi_j\rangle = \langle\psi_0|\hat{H}^{i+j}|\psi_0\rangle = \widetilde{H}_{i,j-1}. \tag{5}$$

Since the overlap matrix $\widetilde{S} = U\Lambda U^\dagger = (U\sqrt{\Lambda})(U\sqrt{\Lambda})^\dagger$ is symmetric and positive definite, the lowest energy in the Krylov subspace is equivalently the lowest eigenvalue of the canonical orthonormalized generalized eigenvalue

problem [55, 56]:

$$((U\sqrt{\Lambda}^{-1})^\dagger \widetilde{H} U\sqrt{\Lambda}^{-1})y^m = E_m y^m \tag{6}$$

We can thus obtain the eigenstates $c^m = U\sqrt{\Lambda}^{-1}y^m$ and eigenvalues

$$E_m = \sum_{i,j} c_i^m c_j^m \widetilde{H}_{ij}, \tag{7}$$

of the generalized eigenvalue problem by solving (6).

The matrices $\widetilde{H}$ and $\widetilde{S}$ are of Hankel form, i.e., their rows are related by a shift as in the matrix

$$A = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_1 & a_2 & & & \vdots \\ a_2 & & & & a_{2n-4} \\ \vdots & & & a_{2n-4} & a_{2n-3} \\ a_{n-1} & \cdots & a_{2n-4} & a_{2n-3} & a_{2n-2} \end{pmatrix}. \tag{8}$$

Therefore, $\widetilde{H}$ and $\widetilde{S}$ are completely defined by $2D-1$ and $2D-2$ entries, respectively (since $\widetilde{S}_{00} = 1$ by definition). This implies that only a linear (in D) number of distinct quantities need to be measured on the quantum device to obtain energies via the generalized eigenvalue problem. This is a great strength of the quantum Krylov approach. However, as we shall see, this linear scaling in D is lost when trying to obtain energy derivatives in a straightforward way.

B. Derivatives of the Krylov energy

From Eq. (3) the expression of the derivative of E_0 with respect to any parameter θ defining the Hamiltonian is (see Appendix A for a proof)

$$\frac{dE_0}{d\theta} = \frac{1}{\sum_{i,j=0}^{D-1} c_i^0 c_j^0 \widetilde{S}_{ij}} \sum_{ij} c_i^0 c_j^0 \left(\frac{d\widetilde{H}_{ij}}{d\theta} - E_0 \frac{d\widetilde{S}_{ij}}{d\theta} \right). \tag{9}$$

Explicitly writing the dependence of the Hamiltonian on the parameter θ as $\hat{H} = \hat{H}(\theta)$ and using the product rule we get

$$\frac{d\widetilde{H}_{ij}}{d\theta} = \sum_{k=1}^{i+j+1} \langle\psi_0| \hat{H}(\theta)^{k-1} \frac{\partial \hat{H}(\theta)}{\partial\theta} \hat{H}(\theta)^{i+j+1-k} |\psi_0\rangle. \tag{10}$$

Note that to simplify the notation, we only included the Hellmann-Feynman term and omitted the response terms stemming from the Hartree-Fock molecular orbital variations as these are well established and re-deriving them is out of the scope of this work (see Ref. [57] and Section D of Appendix *III* of Ref. [48]). If $\partial\hat{H}(\theta)/\partial\theta$ and $\hat{H}(\theta)$ do not commute, $\partial\widetilde{H}_{ij}/\partial\theta$ is still a Hankel matrix, but the i,j-th matrix element now requires evaluating the expectation value of $i+j$ distinct non-hermitian operators. A similar observation is made for the matrix elements $\partial\widetilde{S}_{ij}/\partial\theta$, since $\widetilde{S}_{ij} = \widetilde{H}_{i,j-1}$. This leaves one with the unsatisfying situation of having to measure $\mathcal{O}(D^2)$ distinct expectation values to evaluate gradients with the quantum device. Since precision is expected to increase exponentially with D [58] (albeit with potential impact from sampling noise) one one might assume this renders the scaling issue negligible. However, for practically relevant D, typically around 10 to 20, the difference between linear scaling and quadratic scaling becomes significantly pronounced and cannot be overlooked.

Here we are particularly interested in obtaining the nuclear gradients, i.e., the derivatives of the molecular electronic ground-state energy with respect to the position of the nuclei or, in other words, the forces acting on the atoms. In this case, the Hamiltonian is of the form

$$\hat{\mathcal{H}}_{\text{mol}} \coloneqq E_{\text{nuc}} + \sum_{pq}^{N} k_{pq}\hat{E}_{pq} + \frac{1}{2}\sum_{pqrs}^{N} g_{pqrs}\hat{E}_{pq}\hat{E}_{rs}, \tag{11}$$

where N is the number of spin orbitals, and the one-particle excitation operators are defined as $\hat{E}_{pq} \coloneqq \hat{a}_p^\dagger \hat{a}_q$, with $\hat{a}^\dagger$ and $\hat{a}$ as the fermionic creation and annihilation operators, respectively. The constant term, E_{nuc}, corresponds to the nuclear repulsion energy, g_{pqrs} denotes the two-electron integrals, and k_{pq} is the effective one-body operator [48]. The nuclear gradient can then be expressed using the chain rule,

$$\frac{dE_0}{dx} = \sum_{pq} \frac{dE_0}{dk_{pq}} \frac{dk_{pq}}{dx} + \sum_{pqrs} \frac{dE_0}{dg_{pqrs}} \frac{dg_{pqrs}}{dx}. \tag{12}$$

The expression for dE_0/dk_{pq} can be found from Eq. (9), and Eq. (10) can also be rewritten as

$$\frac{d\widetilde{H}_{ij}}{dk_{pq}} = \sum_{k=1}^{i+j+1} \langle \psi_0 | \hat{H}^{k-1} \hat{E}_{pq} \hat{H}^{i+j+1-k} | \psi_0 \rangle . \tag{13}$$

Note that similar expressions can be obtained for the derivatives with respect to the two-electron integrals. To avoid performing $\mathcal{O}(D^2)$ distinct overlap measurements $\langle \psi_0 | H^k \hat{E}_{pq} H^j | \psi_0 \rangle$, one can instead prepare the quantum Krylov ground-state $|\Psi_0\rangle = \sum_{i=0}^{D-1} c_i^0 \hat{H}^i |\psi_0\rangle$ on the QPU and use the Hellmann-Feynman theorem to replace Eq. (9) by

$$\gamma_{pq} := \frac{dE_0}{dk_{pq}} = \langle \Psi_0 | \hat{E}_{pq} | \Psi_0 \rangle \tag{14}$$

and

$$\Gamma_{pqrs} := \frac{dE_0}{dg_{pqrs}} = \langle \Psi_0 | \hat{E}_{pq} \hat{E}_{rs} | \Psi_0 \rangle . \tag{15}$$

Indeed, preparing $|\Psi_0\rangle$ with QSP has an inherent success probability $p_{\text{success}} < 1$ that depends on several factors such as the dimension of the Krylov space D. Another important consideration is the nature of the polynomial itself. Polynomials that are ill-conditioned or degenerate can be challenging particularly using methods that require roots solving of high degree polynomials [59]. We proceed with this approach for the calculation for the 1- and 2-body RDMs mainly because, as previously mentioned, the number of distinct measurements needed reduces from $O(D^2)$ to $O(1)$. Regarding the variance properties we benchmark the QSP based approach and one would need similar benchmarks for the verbose overlap approach to comment on how the two approaches compare. In the following section we detail how to prepare $|\Psi_0\rangle$ in practice. In Section IV we then show how to efficiently evaluate γ_{pq} and Γ_{pqrs}.

III. QUANTUM KRYLOV EIGENSTATE PREPARATION

As stated above we can prepare $|\Psi_0\rangle$, resulting from the quantum Krylov algorithm, to bypass the $\mathcal{O}(D^2)$ term measurements needed to calculate the nuclear gradients. This can be achieved with QSP [17, 50, 60]. In this section, we first present the foundational oracles used to block-encode a Hamiltonian in its linear combination of unitaries (LCU) form. Then we present the QSP circuit and detail how to generate the appropriate gate parameters to block-encode a given polynomial.

A. Block encoding and Qubitization

The electronic structure Hamiltonian expressed in a molecular orbital basis, after mapping to a N-qubit space (system register) via the Jordan-Wigner mapping, can be written as a linear combination of products $\hat{P}_k$ of Pauli operators

$$\hat{\mathcal{H}} = \sum_k \alpha_k \hat{P}_k, \tag{16}$$

where the coefficients α_k are real. Without loss of generality, we can consider that all $\alpha_k \geq 0$ by absorbing phase factors into the $\hat{P}_k$ and we use the subscripts a and s to mark the ancillary and system register qubits. The operator $\hat{G}$ acts on an ancilla register as

$$\hat{G} |0\rangle_a := |G\rangle_a = \sum_k \sqrt{\frac{\alpha_k}{\lambda_{\text{LCU}}}} |k\rangle_a \tag{17}$$

where $\lambda_{\text{LCU}} \coloneqq \sum_k \alpha_k$.

The operator $\hat{U}$ is then defined on the full set of qubits (system and ancilla registers) as

$$\hat{U} \left|k\right\rangle_a \left|\psi_0\right\rangle_s = \left|k\right\rangle_a \hat{P}_k \left|\psi_0\right\rangle_s . \tag{18}$$

Combining these operators, we obtain the block encoding of $\hat{H} \coloneqq \hat{\mathcal{H}}/\lambda_{\text{LCU}}$ in the subspace where all the ancilla qubits are in the $|0\rangle$ state,

$$\left\langle 0\right|_a \hat{U}_H \left|0\right\rangle_a \left|\psi_0\right\rangle_s = \left\langle G\right|_a \hat{U} \left|G\right\rangle_a \left|\psi_0\right\rangle_s = \hat{H} \left|\psi_0\right\rangle_s . \tag{19}$$

where $\hat{U}_H \coloneqq \hat{G}^\dagger \hat{U} \hat{G}$. Given the spectral decomposition

$$\hat{H} = \sum_j \lambda_j \left|\lambda_j\right\rangle_s \left\langle\lambda_j\right|_s \tag{20}$$

with $\lambda_j \in [-1, 1]$, we have that for each eigenstate $|\lambda_j\rangle$

$$\hat{U} \left|G\right\rangle_a \left|\lambda_j\right\rangle_s = \lambda_j \left|G\right\rangle_a \left|\lambda_j\right\rangle_s + \sqrt{1-\lambda_j^2} \left|G_j^\perp\right\rangle \tag{21}$$

Since $\hat{U}$ is the product of controlled-$\hat{P}_k$ operators, it is self inverse, i.e $\hat{U}^2 = \hat{\mathbb{1}}$. Therefore, by defining the reflection around $|G\rangle$

$$\hat{R}_G \coloneqq (2 \left|G\right\rangle_a \left\langle G\right|_a - \hat{\mathbb{1}}_a) \otimes \hat{\mathbb{1}}_s, \tag{22}$$

we know from Corollary 9 of Ref. [17], that the iterate $\hat{W} = \hat{R}_G \hat{U}$ leaves the two-dimensional subspaces $\mathcal{B}_j$ spanned by the state $\left|G\right\rangle_a \left|\lambda_j\right\rangle_s$ and $\hat{W} \left|G\right\rangle_a \left|\lambda_j\right\rangle_s$ invariant, i.e., $\text{span}(\left|G\right\rangle_a \left|\lambda_j\right\rangle_s , \hat{W} \left|G\right\rangle_a \left|\lambda_j\right\rangle_s) = \text{span}(\left|G\right\rangle_a \left|\lambda_j\right\rangle_s , \left|G_j^\perp\right\rangle)$. In this two-dimensional basis the iterate reads

$$[\hat{W}]_{\mathcal{B}_|} = \begin{pmatrix} \lambda_j & -\sqrt{1-\lambda_j^2} \\ \sqrt{1-\lambda_j^2} & \lambda_j \end{pmatrix}. \tag{23}$$

We see that $[\hat{W}]_{\mathcal{B}_|} = \hat{R}_Y(2 \arccos(\lambda_j)) = e^{i \arccos(\lambda_j) \hat{\sigma}_Y}$. Multiple applications of the iterate therefore result in a highly structured behavior. This concept, called qubitization, is used in QSP to construct the block-encoding of a wide set of polynomial functions.

B. Quantum Signal Processing

In what follows we define $\hat{W}(x) \coloneqq e^{i \arccos(x) \hat{\sigma}_Y}$. The QSP theorem [17] states that given an integer $d > 0$ and two complex polynomials $P(x)$ and $Q(x)$ satisfying the following conditions,

1. $\deg(P) \leq d$, $\deg(Q) \leq d-1$,
2. parity of $P = d \mod 2$, parity of $Q = d-1 \mod 2$ and
3. $|P(x)|^2 + (1-x^2)|Q(x)|^2 = 1$,

there exist phase factors $\Phi = (\phi_0, \ldots, \phi_d) \in [-\pi, \pi)^{d+1}$ such that

$$\hat{U}_\Phi(x) = e^{i(\phi_0 - \frac{\pi}{4})\hat{\sigma}_z} \Big(\prod_{k=1}^{d-1} \hat{W}(x) \, \mathrm{e}^{i\phi_k \hat{\sigma}_z} \Big) \, \mathrm{e}^{i(\phi_d + \frac{\pi}{4})\hat{\sigma}_z} = \begin{pmatrix} P(x) & i\sqrt{1-x^2} Q(x) \\ i\sqrt{1-x^2} Q^*(x) & P^*(x) \end{pmatrix}. \tag{24}$$

The corresponding circuit to $\hat{U}_\Phi(x)$ in Eq. (24) is laid out in Figure 1, in which, for improved readability, we have defined $\tilde{\phi}_i \coloneqq \phi_i$, $i = 1, \ldots, d-1$, $\tilde{\phi}_0 \coloneqq \phi_0 - \pi/4$ and $\tilde{\phi}_d \coloneqq \phi_d + \pi/4$.

While in general QSP applies to complex polynomial, we restrict ourselves to the real case since the electronic structure Hamiltonian and thus the Krylov space and ground-state are real. To construct the block-encoding of a general real polynomial $f(x)$, the polynomial is first decomposed into its odd and even components: $f = 1/2(f_{\text{odd}} + f_{\text{even}})$. Note that the explicit dependence on x is now omitted to simplify the notation. The next step involves

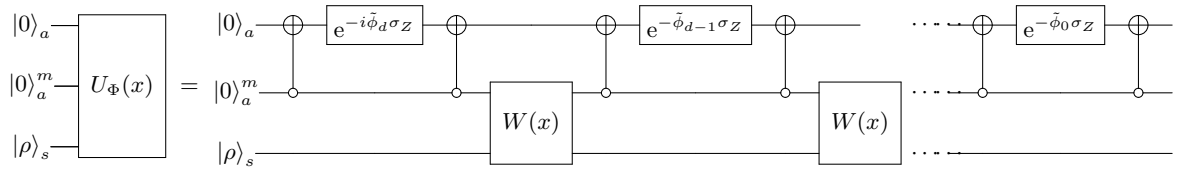


FIG. 1. Quantum circuit for block-encoding a polynomial $P(x)$ of fixed parity using quantum signal processing as defined in Eq. (24).

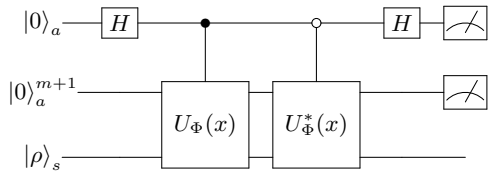


FIG. 2. Quantum circuit template for block-encoding the real part of $P(x)$ presented in Eq. (24). where U_Φ the circuit depicted in Figure 1 for block-encoding P and P^* respectively.

determining the complex polynomials $P_{\text{even(odd)}}$ and $Q_{\text{even(odd)}}$ that satisfy the conditions outlined above, ensuring that the real part of $P_{\text{even(odd)}}$ corresponds to $f_{\text{odd(even)}}$. Note that the existence of such polynomials is guaranteed [61]. Then the phase factors, Φ, must be determined either analytically or numerically. In the numerical approach [59], Φ is obtained by minimizing $||\langle 0|U_\Phi|0\rangle - P_{\text{odd(even)}}||$ over a grid of points $x_k = \cos(\frac{(2j-1)\pi}{4\tilde{d}})$ where $\tilde{d} = \lceil \frac{d+1}{2} \rceil$ and $j \in [1, \tilde{d}]$. This method is indeed more numerically stable for high degree cases compared to the analytical approaches [61] as it does not require root finding methods. Once the phase factors are determined, the block-encoding of $f_{\text{odd(even)}}$ is constructed by introducing an extra ancilla qubit initialized in the $|+\rangle$ state to control the unitaries U_Φ and U_Φ^* as illustrated in Figure 2). This produces $1/2(P_{\text{odd(even)}} + P^*_{\text{odd(even)}}) = f_{\text{odd(even)}}$. Finally, the block-encoding of f is obtained by adding an extra qubit and by applying the same procedure to combine the block-encodings of f_{odd} and f_{even}.

IV. EVALUATION OF GRADIENTS ON THE QSP PREPARED EIGENSTATE

For the rest of the manuscript, we use Chebyshev polynomials as our Krylov basis as the Chebyshev basis is more convenient in the context of qubitization. As previously mentioned, both the monomials of Eq. (1) and the Chebyshev polynomials span the same subspace and therefore the Krylov method yields the same ground-state in both cases. The straightforward differentiation of the Krylov energy also produces the same increase of complexity to a $O(D^2)$ scaling. By solving the generalized eigenvalue problem in the Krylov subspace one obtains the eigenenergies and the coefficients c_i^m of the eigenstates. Without loss of generality, we proceed next by fixing $m = 0$ i.e. to the ground-state and energy approximation case. The quantum state corresponding to the Krylov ground-state energy is then given by $|\Psi_0\rangle = \sum_{i=0}^{D-1} c_i^0 T_i(\hat{H})|\psi_0\rangle$. The function $\sum_{i=0}^{D-1} c_i^0 T_i(H)$ needs to be appropriately normalized before QSP can be applied [59]. Let us then define the infinity norm of the polynomial function defining the ground-state over $[-1, 1]$ as:

$$\eta \coloneqq \max(|\sum_{i=0}^{D-1} c_i^0 T_i(x)|, |x| \leq 1). \tag{25}$$

We then set

$$f(x) = \frac{1}{\eta} \sum_{i=0}^{D-1} c_i^0 T_i(x), \tag{26}$$

which satisfies the condition $|f| \leq 1$. The coefficients $c_i^0/\eta, i = 0, \ldots, D-1$ are used, as described in Section III B, to determine the corresponding phase angles Φ of the QSP circuit. Let $\hat{\mathcal{U}}_\Phi$ be the unitary of the corresponding QSP

circuit,

$$\hat{\mathcal{U}}_\Phi \left|0\right\rangle_a \left|\psi_0\right\rangle_s = f(\hat{H}) \left|0\right\rangle_a \left|\psi_0\right\rangle_s + \beta \left|\perp\right\rangle \tag{27}$$

where $|\beta|^2 = 1 - \|f(\hat{H}) \left|\psi_0\right\rangle_s\|^2$ and $\langle 0|_a \otimes I \left|\perp\right\rangle = 0$.

The ground-state can then be re-written as $\left|\Psi_0\right\rangle_s = \eta \langle 0|_a \hat{\mathcal{U}}_\Phi \left|0\right\rangle_a \left|\psi_0\right\rangle_s$ and therefore the energy expression is $E_0 = \lambda_{\mathrm{LCU}} \langle\Psi_0|_s \hat{H} \left|\Psi_0\right\rangle_s = \lambda_{\mathrm{LCU}} \eta^2 \langle\psi_0|_s \hat{\mathcal{U}}_\Phi^* \left|0\right\rangle_a \langle 0|_a \hat{H} \hat{\mathcal{U}}_\Phi \left|\psi_0\right\rangle_s$ and therefore we have for the derivative

$$\frac{\partial E_0}{\partial k_{pq}} = \lambda_{\mathrm{LCU}} \eta^2 \langle\psi_0|_s \hat{\mathcal{U}}_\Phi^* \left|0\right\rangle_a \langle 0|_a \hat{E}_{pq} \hat{\mathcal{U}}_\Phi \left|\psi_0\right\rangle_s . \tag{28}$$

Using the Jordan-Wigner mapping, the fermionic operator is replaced by a linear combination of Pauli strings, hence the derivative of Eq. (28) is ultimately a linear combination of expectation values with respect to Pauli strings $\hat{P}_\nu$ with coefficients ω_ν

$$\frac{\partial E_0}{\partial k_{pq}} = \lambda_{\mathrm{LCU}} \eta^2 \sum_\nu \omega_\nu \langle\psi_0|_s \hat{\mathcal{U}}_\Phi^* \left|0\right\rangle_a \langle 0|_a \hat{P}_\nu \hat{\mathcal{U}}_\Phi \left|\psi_0\right\rangle_s . \tag{29}$$

One can now proceed in two different ways to calculate the nuclear gradients. In the first approach, each term of the previous equation can be measured by post selecting the measurements to those where the ancilla qubits are in the zero state. However, this has a success probability $\mathrm{P_{suc}} \coloneqq \|f(H) \left|\psi_0\right\rangle_s\|^2$.

We therefore propose an alternative coherent, i.e. without post-selecting on the ancilla regitser, way of measuring the expectation value of $\hat{P}_\nu$. This can be achieved by measuring the observables $\hat{P}_\nu$ and $\hat{R}_0 \hat{P}_\nu$ of the state produced by the QSP circuit $\hat{U}_\Phi \left|0\right\rangle \left|\phi_0\right\rangle$:

$$o_1 \coloneqq \langle\psi_0|_s \langle 0|_a \hat{\mathcal{U}}_\Phi^* \hat{P}_\nu \hat{\mathcal{U}}_\Phi \left|0\right\rangle_a \left|\psi_0\right\rangle_s \tag{30}$$

and

$$o_2 \coloneqq \langle\psi_0|_s \langle 0|_a \hat{\mathcal{U}}_\Phi^* \hat{R}_0 \hat{P}_\nu \hat{\mathcal{U}}_\Phi \left|0\right\rangle_a \left|\psi_0\right\rangle_s \tag{31}$$

It can then be easily demonstrated (see Appendix B) that the sum of both observables yields:

$$2\langle\Psi_0|_s \hat{P}_\nu |\Psi_0\rangle_s = \eta^2 (o_1 + o_2). \tag{32}$$

It is worth noting that both $\hat{P}_\nu$ and $\hat{R}_\nu$ commute and therefore the observables $\hat{P}_\nu$ and $\hat{R}_0 \hat{P}_\nu$ can be jointly measured. In the following section, we compare both approaches by benchmarking their statistical properties via numerical simulations.

The QSP based eigenstate preparation routine developed above opens up the possibility to combine quantum Krylov methods with other techniques for the more efficient or more accurate quantum simulation of molecules. Instead of decomposing the Hamiltonian into Pauli words and then measuring them separately, one and two-particle RDMs can be obtained from compressed representations of the Hamiltonian, such as different flavors of double factorization [10, 48, 62]. This should further reduce the variance and thus number of shots needed to compute, e.g., nuclear gradients to a desired precision. Alternatively, one can take a classical shadow [63–66] of the state and then estimate RDMs from that shadow. One can even use such a shadow of a Krylov eigenstate to unbias fermionic quantum Monte-Carlo computations [67] or tailor split-amplitude coupled cluster methods [68] and thereby, e.g., augment the results with dynamic correlation corrections. We leave a detailed investigation of the potential of these method combinations to future work and in the following restrict ourselves to a numerical comparison of the methods to compute nuclear gradients via Pauli decompositions of molecular Hamiltonians.

V. NUCLEAR GRADIENT BENCHMARKS

In this section we benchmark the efficiency of the nuclear gradient calculation on the quantum Krylov ground-state through numerical examples. To this aim, we consider the H_2O molecule in an active space of 4 electrons in 4 molecular orbitals using cc-pVDZ atomic orbital basis set. We use PySCF [69] for generating the one and two-electron integrals and PennyLane [70] to simulate the circuits.

As per Section III B, we employ Chebyshev polynomials, applied to the Hartree-Fock state, to build the Krylov basis. We simulate the measurement of the matrix elements $\widetilde{H}_{ij} \coloneqq \langle\psi_0| T_i(\hat{H}) \hat{H} T_j(\hat{H}) \left|\psi_0\right\rangle$ and $\widetilde{S}_{ij} \coloneqq \langle\psi_0| T_i(\hat{H}) T_j(\hat{H}) \left|\psi_0\right\rangle$

as described in section II Ref. [53]. The Krylov ground-state coefficients, c^0, are obtained by solving the generalized eigenvalue problem as described in Section II Eq. (6). We regularize the overlap matrix $\widetilde{S}$ by fixing a threshold, s and by discarding all eigenvalues and their corresponding eigenvectors of $\widetilde{S}$ that are below s.

First, we compare the performance of measuring the one and two-particle RDMs using the post-selection and the coherent approach of Eq. (32) presented in Section IV. We assess how each method influences the efficiency of calculating the nuclear gradient as per Eq. (12). To do so we first determine the analytical variance of the RDMs' matrix elements, γ_{pq} and Γ_{pqrs}, with PennyLane. We then draw samples $(\gamma, \Gamma)_i$, with $1 \leq i \leq 100$, where each matrix element of the RDMs is obtained from a normal distribution defined by the analytical variance and by a mean taken to be the exact γ_{pq} (or Γ_{pqrs}). We compute the nuclear gradient for each of the 100 realizations and compute the ensemble mean and variance. In Figure 3, we plot the total variance, i.e., the sum of the ensemble variances of all the elements of the nuclear gradient vectors for both methods (post-selection and the Eq. (32) approach). We record the total variance's evolution as a function of the Krylov dimension, D. We observe that the latter approach results in a lower variance compared to post-selection, on average by about a factor of approximately two. This directly implies the same factor is carried between the total shot budget of both methods for a given target precision ϵ. The optimal shot distribution should take into account the variances of both p_0 (30) and p_1 (31). In all cases the variance of the RDM has a quartic dependence on the infinity norm η, i.e.,

$$\mathrm{Var}[\gamma_{pq}] = \eta^4 \mathrm{Var}[\langle\psi_0| f(\hat{H}) \hat{E}_{pq} f(\hat{H}) |\psi_0\rangle], \tag{33}$$

and similarly for Γ_{qprs}. It is therefore crucial to control the infinity norm η to prevent a significant increase in the variance of the RDMs elements, which would result in a substantially higher measurement cost. In Figure 4a, we plot the infinity norm η as a function of the Krylov subspace dimension for different regularization threshold s. We first observe that the value of η generally decreases with the Krylov subspace dimension. More importantly, we see that setting a low threshold s leads to high η, especially when D is small. For small D and s, the increase in measurements due to η becomes impractical, even for the small molecular system studied here. However, enlarging s has a negative effect on the energy accuracy.

In Figure 4b, we plot the absolute difference Δ between the true ground-state energy, obtained via exact diagonalization of the active space Hamiltonian, and the QKSD ground-state energy as a function of D. This is shown for various values of s. As expected, the QKSD energy converges to the exact one as D increases, with smaller s values providing more accurate approximations of the ground-state energy. In practice one thus needs to chose s such that the energy error remains acceptable and then increase D, possibly even beyond the D needed for satisfactory energy precision, to reduce η. It will be important to study the scaling of all these quantities with the size and amount of correlation in a molecular system.

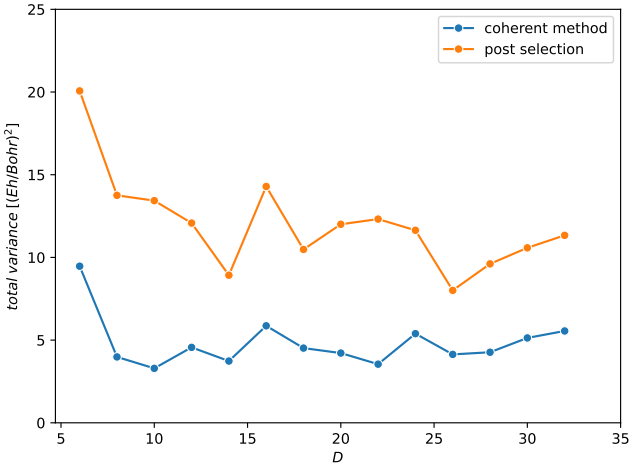
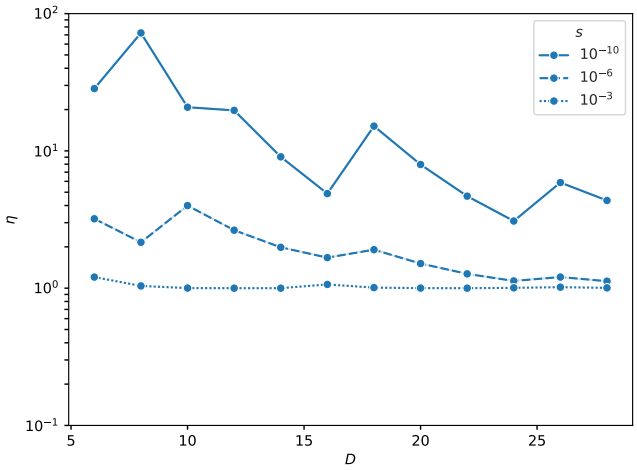
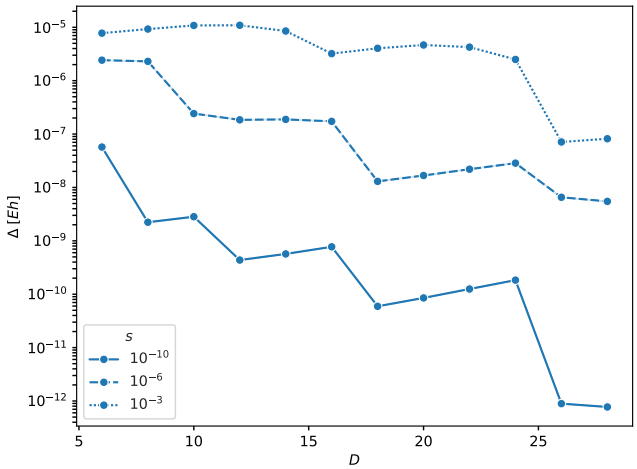


FIG. 3. Total variance, i.e. sum of the variance of all the nuclear gradient dE_0/dx vector components, as a function of the Krylov dimension. The blue and orange plot represent the variance of the coherent approach, that measures $\hat{P}_\nu$ and $\hat{R}_0\hat{P}_\nu$, and the post-selection approach respectively for $s = 10^{-3}$.



(a) Infinity norm η as a function of the Krylov dimension D. The line style represents the η values for different threshold cutoffs s.



(b) Absolute difference Δ between ground-state energy and E_0 as a function of the Krylov dimension D. The line style represents the infinity norm η for different threshold cutoffs s.

FIG. 4. FIG 4a infinity norm η and FIG. 4b the absolute difference of the Krylov minimal energy and the ground-state energy.

VI. CONCLUSION AND DISCUSSION

In this work, we presented a method for calculating nuclear energy gradients alongside energy estimates using the QKSD approach. We first gave the expression for the derivative of the QKSD eigenenergies with respect to any parameter of the Hamiltonian and showed that, while the energies can be obtained with a number of measurements scaling linearly with the Krylov space dimension D, the derivatives naively require $\mathcal{O}(D^2)$ QPU calls. To address this challenge, we proposed to directly prepareQKSD eigenstates $|\Psi_j\rangle$ and extract nuclear gradients by measuring its one- and two-particle RDMs. To efficiently obtain the coefficients of $|\Psi_j\rangle$ with respect to the Krylov basis for the use in QSP we work with a basis of Chebyshev polynomials. Importantly, these coefficients must be normalized by an infinity norm η before they can be used in QSP and the variance of the measured RDM elements scales like η^4. We find that mildly increasing the Krylov dimension D and regularization s used when solving the generalized Krylov eigenvalue problem can reduce η significantly. In a post selected approach the measurement of the RDMs elements also suffers from finite success probabilities, leading to discarded measurement outcomes and increased variance. To mitigate this, we proposed a coherent measurement scheme utilizing two jointly measurable observables, which improves the variance of the estimator. We demonstrated the enhanced performance of the proposed measurement

technique through numerical simulations.

Our numerical results highlight the impact of the generalized eigenvalue regularization threshold s on both the energy approximation and the variance of the RDMs elements. While tight thresholds yield a better energy precision, they also lead to larger η and therefore increase the measurement cost of the RDMs. It is therefore important to balance this parameter in QKSD calculations that are aimed at more than just computing energies. In practice, s is often chosen to scale linearly with the noise level. Ref. [34] demonstrated that in presence of sampling noise, i.e., a finite number of measurements M for each matrix element $\widetilde{H}_{ij}$ and $\widetilde{S}_{ij}$, the optimal value of s is $s_{\text{opt}} = e/\sqrt{M}$ where e is a factor that depends linearly on the dimension of the Hilbert space. In future work, it would be interesting to explore how, through this choice of s, the noise level affects η and therefore the efficiency of calculating properties in our QKSD and QSP framework.

Obtaining the RDMs via our coherent approach reduces the measurement cost per matrix element from $\mathcal{O}(D^2)$ to $\mathcal{O}(1)$ compared to the post-selection approach. It is important to realize that this is achieved without increasing circuit depth. This is because both approaches rely on QSP to implement the Chebyshev polynomials, and the depth of the QSP circuit is determined by the polynomial degree, which is D in both cases.

In this manuscript, we have focused on the nuclear gradients. However, the techniques developed are broadly applicable to any property that requires the measurement of the one- and/or two-particle RDMs including dipole moments or natural orbitals. In addition, our procedure can be used to compute excited state RDMs and gradients if a suitable reference state is used.

VII. ACKNOWLEDGMENTS

The authors thank Jérôme Gonthier for useful discussions and feedback on the manuscript. P.J.O. and R.M.P. own stock/options in QC Ware Corp.

[1] 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).
[2] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding electronic spectra in quantum circuits with linear t complexity, Physical Review X **8**, 041015 (2018).
[3] D. W. Berry, C. Gidney, M. Motta, J. R. McClean, and R. Babbush, Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization, Quantum **3**, 208 (2019).
[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] Y. Su, D. W. Berry, N. Wiebe, N. Rubin, and R. Babbush, Fault-tolerant quantum simulations of chemistry in first quantization, PRX Quantum **2**, 040332 (2021).
[6] I. H. Kim, Y.-H. Liu, S. Pallister, W. Pol, S. Roberts, and E. Lee, Fault-tolerant resource estimate for quantum chemical simulations: Case study on li-ion battery electrolyte molecules, Physical Review Research **4**, 023019 (2022).
[7] A. Delgado, P. A. Casares, R. Dos Reis, M. S. Zini, R. Campos, N. Cruz-Hernández, A.-C. Voigt, A. Lowe, S. Jahangiri, M. A. Martin-Delgado, *et al.*, Simulating key properties of lithium-ion batteries with a fault-tolerant quantum computer, Physical Review A **106**, 032428 (2022).
[8] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, Quantum computing enhanced computational catalysis, Physical Review Research **3**, 033055 (2021).
[9] 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, Proceedings of the National Academy of Sciences **119**, e2203533119 (2022).
[10] D. Rocca, C. L. Cortes, J. F. Gonthier, P. J. Ollitrault, R. M. Parrish, G.-L. Anselmetti, M. Degroote, N. Moll, R. Santagati, and M. Streif, Reducing the runtime of fault-tolerant quantum simulations in chemistry through symmetry-compressed double factorization, Journal of Chemical Theory and Computation (2024).
[11] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, *et al.*, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, Nature Communications **14**, 1952 (2023).
[12] D. Marti-Dafcik, N. Lee, H. G. Burton, and D. P. Tew, Spin-coupled molecular orbitals: chemical intuition meets quantum chemistry, arXiv preprint arXiv:2402.08858 https://doi.org/10.48550/arXiv.2402.08858 (2024).
[13] P. J. Ollitrault, C. L. Cortes, J. F. Gonthier, R. M. Parrish, D. Rocca, G.-L. Anselmetti, M. Degroote, N. Moll, R. Santagati, and M. Streif, Enhancing initial state overlap through orbital optimization for faster molecular electronic ground-state energy estimation, arXiv preprint arXiv:2404.08565 (2024).
[14] M. Mörchen, G. H. Low, T. Weymuth, H. Liu, M. Troyer, and M. Reiher, Classification of electronic structures and state preparation for quantum computation of reaction chemistry, arXiv preprint arXiv:2409.08910 (2024).

[15] D. W. Berry, A. M. Childs, and R. Kothari, Hamiltonian simulation with nearly optimal dependence on all parameters, in *2015 IEEE 56th annual symposium on foundations of computer science* (IEEE, 2015) pp. 792–809.

[16] G. H. Low and I. L. Chuang, Optimal hamiltonian simulation by quantum signal processing, Physical review letters **118**, 010501 (2017).

[17] G. H. Low and I. L. Chuang, Hamiltonian simulation by qubitization, Quantum **3**, 163 (2019).

[18] R. B. Griffiths and C.-S. Niu, Semiclassical fourier transform for quantum computation, Physical Review Letters **76**, 3228 (1996).

[19] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Entanglement-free heisenberg-limited phase estimation, Nature **450**, 393 (2007).

[20] L. Lin and Y. Tong, Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers, PRX Quantum **3**, 010318 (2022).

[21] Z. Ding and L. Lin, Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with applications to ground-state energy estimation, PRX Quantum **4**, 020331 (2023).

[22] A. Dutkiewicz, S. Polla, M. Scheurer, C. Gogolin, W. J. Huggins, and T. E. O'Brien, Error mitigation and circuit division for early fault-tolerant quantum phase estimation (2024), arXiv:2410.05369 [quant-ph].

[23] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, *et al.*, Noisy intermediate-scale quantum algorithms, Reviews of Modern Physics **94**, 015004 (2022).

[24] Y. Tong, Designing algorithms for estimating ground state properties on early fault-tolerant quantum computers, Quantum Views **6**, 65 (2022).

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

[26] K. Seki and S. Yunoki, Quantum power method by a superposition of time-evolved states, PRX Quantum **2**, 010333 (2021).

[27] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nature Physics **16**, 205–210 (2019).

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

[29] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature communications **9**, 4812 (2018).

[30] Y. Saad, On the rates of convergence of the lanczos and the block-lanczos methods, SIAM Journal on Numerical Analysis **17**, 687 (1980).

[31] E. N. Epperly, L. Lin, and Y. Nakatsukasa, A theory of quantum subspace diagonalization, SIAM Journal on Matrix Analysis and Applications **43**, 1263 (2022).

[32] N. Yoshioka, M. Amico, W. Kirby, P. Jurcevic, A. Dutt, B. Fuller, S. Garion, H. Haas, I. Hamamura, A. Ivrii, *et al.*, Diagonalization of large many-body hamiltonians on a quantum processor, arXiv preprint arXiv:2407.14431 (2024).

[33] W. Kirby, Analysis of quantum krylov algorithms with errors, Quantum **8**, 1457 (2024).

[34] G. Lee, D. Lee, and J. Huh, Sampling error analysis in quantum krylov subspace diagonalization, Quantum **8**, 1477 (2024).

[35] S. Bratož, Le calcul non empirique des constantes de force et des dérivées du moment dipolaire, in *Colloq. Int. CNRS*, Vol. 82 (1958) pp. 287–301.

[36] J. Gerratt and I. M. Mills, Force constants and dipole-moment derivatives of molecules from perturbed hartree–fock calculations. i, The Journal of Chemical Physics **49**, 1719 (1968).

[37] P. Pulay, Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules: I. theory, Molecular Physics **17**, 197 (1969).

[38] S. Kato and K. Morokuma, Energy gradient in a multi-configurational scf formalism and its application to geometry optimization of trimethylene diradicals, Chemical Physics Letters **65**, 19 (1979).

[39] J. D. Goddard, N. C. Handy, and H. F. Schaefer, Gradient techniques for open-shell restricted hartree–fock and multiconfiguration self-consistent-field methods, The Journal of Chemical Physics **71**, 1525 (1979).

[40] P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs, Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives, Journal of the American Chemical Society **101**, 2550 (1979).

[41] B. R. Brooks, W. D. Laidig, P. Saxe, J. D. Goddard, Y. Yamaguchi, and H. F. Schaefer, Analytic gradients from correlated wave functions via the two-particle density matrix and the unitary group approach, The Journal of Chemical Physics **72**, 4652 (1980).

[42] R. Krishnan, H. Schlegel, and J. Pople, Derivative studies in configuration–interaction theory, The Journal of Chemical Physics **72**, 4654 (1980).

[43] M. Dupuis, Energy derivatives for configuration interaction wave functions, The Journal of Chemical Physics **74**, 5758 (1981).

[44] H. Nakatsuji, T. Hayakawa, and M. Hada, Force in scf theories. mc scf and open-shell rhf theories, Chemical Physics Letters **80**, 94 (1981).

[45] H. Schaefer III and Y. Yamaguchi, Robert s. mulliken issue, J. Mol. Struct **135**, 369 (1986).

[46] I. Kassal and A. Aspuru-Guzik, Quantum algorithm for molecular properties and geometry optimization, The Journal of chemical physics **131** (2009).

[47] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martinez, Hybrid quantum/classical derivative theory: Analytical gradients and excited-state dynamics for the multistate contracted variational quantum eigensolver, arXiv preprint arXiv:1906.08728 (2019).

[48] E. G. Hohenstein, O. Oumarou, R. Al-Saadon, G.-L. R. Anselmetti, M. Scheurer, C. Gogolin, and R. M. Parrish, Efficient quantum analytic nuclear gradients with double factorization, The Journal of Chemical Physics **158**, 114119 (2023).

[49] T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, Calculating energy derivatives for quantum chemistry on a quantum computer, npj Quantum Information **5**, 10.1038/s41534-019-0213-4 (2019).

[50] T. E. O'Brien, M. Streif, N. C. Rubin, R. Santagati, Y. Su, W. J. Huggins, J. J. Goings, N. Moll, E. Kyoseva, M. Degroote, C. S. Tautermann, J. Lee, D. W. Berry, N. Wiebe, and R. Babbush, Efficient quantum computation of molecular forces and other energy gradients, Physical Review Research **4**, 10.1103/physrevresearch.4.043210 (2022).

[51] K. Mitarai, Y. O. Nakagawa, and W. Mizukami, Theory of analytical energy derivatives for the variational quantum eigensolver, Physical Review Research **2**, 013129 (2020).

[52] R. M. Parrish, G.-L. R. Anselmetti, and C. Gogolin, Analytical ground-and excited-state gradients for molecular electronic structure theory from hybrid quantum/classical methods, arXiv preprint arXiv:2110.05040 (2021).

[53] W. Kirby, M. Motta, and A. Mezzacapo, Exact and efficient lanczos method on a quantum computer, Quantum **7**, 1018 (2023).

[54] Z. Zhang, A. Wang, X. Xu, and Y. Li, Measurement-efficient quantum krylov subspace diagonalisation, Quantum **8**, 1438 (2024).

[55] P.-O. Löwdin, Quantum theory of cohesive properties of solids, Advances in Physics **5**, 1 (1956).

[56] P.-O. Löwdin, On the nonorthogonality problem, in *Advances in quantum chemistry*, Vol. 5 (Elsevier, 1970) pp. 185–199.

[57] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Derivative studies in hartree-fock and møller-plesset theories, International Journal of Quantum Chemistry **16**, 225 (1979), https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560160825.

[58] E. Koch, 8 the lanczos method, The LDA+ DMFT approach to strongly correlated materials (2011).

[59] Y. Dong, X. Meng, K. B. Whaley, and L. Lin, Efficient phase-factor evaluation in quantum signal processing, Physical Review A **103**, 042419 (2021).

[60] D. Motlagh and N. Wiebe, Generalized quantum signal processing, PRX Quantum **5**, 10.1103/prxquantum.5.020368 (2024).

[61] A. Gilyén, Y. Su, G. H. Low, and N. 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* (2019) pp. 193–204.

[62] O. Oumarou, M. Scheurer, R. M. Parrish, E. G. Hohenstein, and C. Gogolin, Accelerating quantum computations of chemistry through regularized compressed double factorization, Quantum **8**, 1371 (2024).

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

[64] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic partial tomography via classical shadows, Phys. Rev. Lett. **127**, 110504 (2021).

[65] G. H. Low, Classical shadows of fermions with particle number symmetry, arXiv preprint , arXiv:2208.08964 (2022).

[66] K. Wan, W. J. Huggins, J. Lee, and R. Babbush, Matchgate shadows for fermionic quantum simulation, Communications in Mathematical Physics **404**, 629–700 (2023).

[67] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, Nature **603**, 416 (2022).

[68] M. Scheurer, G.-L. R. Anselmetti, O. Oumarou, C. Gogolin, and N. C. Rubin, Tailored and externally corrected coupled cluster with quantum inputs, Journal of Chemical Theory and Computation **20**, 5068–5093 (2024).

[69] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, Pyscf: the python-based simulations of chemistry framework, Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1340 (2018).

[70] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. Akash-Narayanan, A. Asadi, *et al.*, Pennylane: Automatic differentiation of hybrid quantum-classical computations, arXiv preprint arXiv:1811.04968 (2018).

Appendix A: DERIVATIVE OF KRYLOV ENERGY

Let $(E_m, |\xi_m\rangle)$ be a generalized eigenpair w.r.t $(\widetilde{H}, \widetilde{S})$ i.e.

$$\widetilde{H}\,|\xi_m\rangle = E_m \widetilde{S}\,|\xi_m\rangle \tag{34}$$

which subsequently implies that $\langle\xi_m|\,\widetilde{H}\,|\xi_m\rangle = E_m\,\langle\xi_m|\,\widetilde{S}\,|\xi_m\rangle$

$$\frac{d\,\langle\xi_m|\,\widetilde{H}\,|\xi_m\rangle}{dx} = \frac{dE_m}{dx}\,\langle\xi_m|\,\widetilde{S}\,|\xi_m\rangle + E_m \frac{d\,\langle\xi_m|\,\widetilde{S}\,|\xi_m\rangle}{dx} \tag{35}$$

$$\langle\frac{d\xi_m}{dx}|\,\widetilde{H}\,|\xi_m\rangle + \langle\xi_m|\,\frac{d\widetilde{H}}{dx}\,|\xi_m\rangle + \langle\xi_m|\,\widetilde{H}\,|\frac{d\xi_m}{dx}\rangle = \frac{dE_m}{dx}\,\langle\xi_m|\,\widetilde{S}\,|\xi_m\rangle + E_m(\langle\frac{d\xi_m}{dx}|\,\widetilde{S}\,|\xi_m\rangle + \langle\xi_m|\,\frac{d\widetilde{S}}{dx}\,|\xi_m\rangle + \langle\xi_m|\,\widetilde{S}\,|\frac{d\xi_m}{dx}\rangle) \tag{36}$$

using (34) we have:

$$E_m\,\langle\frac{d\xi_m}{dx}|\,\widetilde{S}\,|\xi_m\rangle + \langle\xi_m|\,\frac{d\widetilde{H}}{dx}\,|\xi_m\rangle + E_m\,\langle\xi_m|\,\widetilde{S}\,|\frac{d\xi_m}{dx}\rangle = \frac{dE_m}{dx}\,\langle\xi_m|\,\widetilde{S}\,|\xi_m\rangle + E_m(\langle\frac{d\xi_m}{dx}|\,\widetilde{S}\,|\xi_m\rangle + \langle\xi_m|\,\frac{d\widetilde{S}}{dx}\,|\xi_m\rangle + \langle\xi_m|\,\widetilde{S}\,|\frac{d\xi_m}{dx}\rangle) \tag{37}$$

$$\implies \langle\xi_m|\,\frac{d\widetilde{H}}{dx}\,|\xi_m\rangle = \frac{dE_m}{dx}\,\langle\xi_m|\,\widetilde{S}\,|\xi_m\rangle + E_m\,\langle\xi_m|\,\frac{d\widetilde{S}}{dx}\,|\xi_m\rangle \tag{38}$$

$$\implies \frac{dE_m}{dx} = \frac{\langle\xi_m|\,\frac{d\widetilde{H}}{dx}\,|\xi_m\rangle\,\langle\xi_m|\,\widetilde{S}\,|\xi_m\rangle - \langle\xi_m|\,\frac{d\widetilde{S}}{dx}\,|\xi_m\rangle\,\langle\xi_m|\,\widetilde{H}\,|\xi_m\rangle}{\langle\xi_m|\,\widetilde{S}\,|\xi_m\rangle^2} \tag{39}$$

Appendix B: RDM THROUGH A COHERENT APPROACH

Let us first recall the function f and unitary $\hat{\mathcal{U}}_\Phi$ defined in (26) and (27) respectively. Let $\hat{P}_\nu$ and $\hat{R}_0$ be a Pauli string and the reflection defined in (22). Then picking up from equation (27), we have:

$$\hat{\mathcal{U}}_\Phi\,|0\rangle_a\,|\phi_0\rangle_s = f(\hat{H})\,|0\rangle_a\,|\phi_0\rangle_s + \beta\,|\perp\rangle \tag{40}$$

$$\implies \langle\phi_0|_s\,\langle 0|_a\,\hat{\mathcal{U}}_\Phi^*\hat{P}_\nu\hat{\mathcal{U}}_\Phi\,|0\rangle_a\,|\phi_0\rangle_s = \frac{1}{\eta^2}\,\langle\Psi_0|\,\hat{P}_\nu\,|\Psi_0\rangle + \beta^2\,\langle\perp|\,\hat{P}_\nu\,|\perp\rangle\,, \tag{41}$$

and we have

$$\hat{R}_0\hat{P}_\nu\hat{\mathcal{U}}_\Phi\,|0\rangle_a\,|\phi_0\rangle_s = (2\,|0\rangle\,\langle 0|_a - I)\hat{P}_\nu f(\hat{H})\,|0\rangle_a\,|\phi_0\rangle + \beta(2\,|0\rangle\,\langle 0|_a - I)\hat{P}_\nu\,|\perp\rangle = \hat{P}_\nu f(\hat{H})\,|0\rangle_a\,|\phi_0\rangle - \beta\,|\perp\rangle \tag{42}$$

$$\implies \langle\phi_0|_s\,\langle 0|_a\,\hat{\mathcal{U}}_\Phi^*\hat{R}_0\hat{P}_\nu\hat{\mathcal{U}}_\Phi\,|0\rangle_a\,|\phi_0\rangle_s = \langle\phi_0|\,f(\hat{H})\hat{P}_\nu f(\hat{H})\,|\phi_0\rangle - \beta^2\,\langle\perp|\,\hat{P}_\nu\,|\perp\rangle = \frac{1}{\eta^2}\,\langle\Psi_0|\,\hat{P}_\nu\,|\Psi_0\rangle - \beta^2\,\langle\perp|\,\hat{P}_\nu\,|\perp\rangle \tag{43}$$

Adding (41) and (43) we obtain

$$\langle\phi_0|_s\,\langle 0|_a\,\hat{\mathcal{U}}_\Phi^*\hat{P}_\nu\hat{\mathcal{U}}_\Phi\,|0\rangle_a\,|\phi_0\rangle_s + \langle\phi_0|_s\,\langle 0|_a\,\hat{\mathcal{U}}_\Phi^*\hat{R}_0\hat{P}_\nu\hat{\mathcal{U}}_\Phi\,|0\rangle_a\,|\phi_0\rangle_s = \langle\phi_0|\,f(\hat{H})\hat{P}_\nu f(\hat{H})\,|\phi_0\rangle - \beta^2\,\langle\perp|\,\hat{P}_\nu\,|\perp\rangle = \frac{2}{\eta^2}\,\langle\Psi_0|\,\hat{P}_\nu\,|\Psi_0\rangle\,. \tag{44}$$

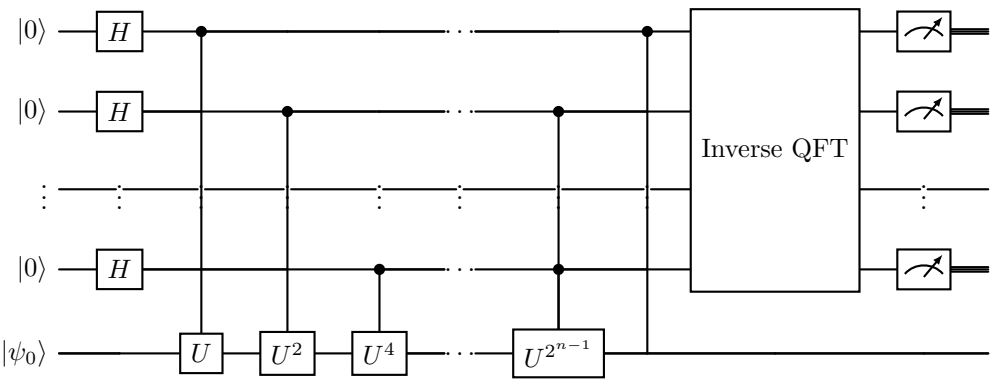


Figure 5: QPE circuit for unitary operator U with input state $|\psi_0\rangle$. The control register (top) is scalable, with inverse QFT and measurement at the end.

5 Linear Depth Clifford+T Decomposition of The Bucket Brigade QRAM

5.1 Preliminaries on Quantum Phase Estimation

The quantum phase estimation algorithm is a prominent and one of the best-known and most studied quantum algorithms in the literature [41]. Therefore, it suffices here to qualitatively describe the procedure. Given a unitary U with spectrum $\{e^{i\phi_i}, i = 0..N-1\}$ and eigenvectors $\{|\psi_i\rangle, i = 0..N-1\}$, the QPE algorithm, provided with input $|\Psi_0\rangle$, outputs in the control registers $2^d\theta_i$ in a total runtime of $O(p_i\epsilon^{-1})$ where $p_i = |\langle\psi_i|\Psi_0\rangle|^2$. The textbook QPE algorithm is depicted in FIG 5 and has three main parts:

- Uniform superposition in the control register using Hadamard gates.
- Exponent of the unitary U controlled on the control qubit register.
- Lastly, an inverse quantum Fourrier transform [41] on the control register and readout

The QPE is hence used within the framework of quantum chemistry, proposed to solve the ESP and fulfills the optimal theoretical bounds, namely the Heisenberg scaling $O(\epsilon^{-1})$. More specifically, the unitary U has been traditionally proposed to be the time-evolution operator $U = e^{-i\hat{\mathcal{H}}t}$ (in which case $N \coloneqq 2^n$) since the phases of the eigenvalues correspond directly to the eigenenergies, i.e. $e^{-iHt}|\psi_i\rangle = e^{-iE_it}|\psi_i\rangle$. However, as we will explain in Section 5.4, in state-of-the-art QPE, the time evolution operator has been replaced with another operator, namely the walk operator. To show the importance of the λ_{LCU} factor in QPE, we start by introducing block-encoding and the qubitization techniques which are essential tools to build the walk operator and lastly present how it is conjugated with QPE.

5.2 Block-Encoding

Informally, the block-encoding of a matrix A is a larger matrix U that is unitary and whose upper-left block is A

$$U \coloneqq \begin{pmatrix} A & * \\ * & * \end{pmatrix}. \tag{86}$$

However, this is only possible if $\|A\| \leq 1$. Hence, to fulfill this condition, A must be normalized by a normalization factor $\lambda \geq \|A\|$. For instance, suppose we have the LCU decomposition of $A = \sum_{k=0}^{K} \alpha_k \hat{P}_k$, $\alpha_k \in \mathbb{C}$, then an upper-bound of $\|A\|$ is

$$\lambda_{LCU} \coloneqq \sum_{k=0}^{K} |\alpha_k|. \tag{87}$$

Block-encoding is a crucial elements for algorithms that apply general linear algebraic operations on the matrix A such as QSP, QSVT, GQS [17, 20, 39]. Formally, an m qubit matrix U is called an (α, a, ϵ) block-encoding of A if [20]

$$\|A - \alpha\langle 0|_a U|0\rangle_a\| \leq \epsilon \tag{88}$$

where $\alpha \geq \|A\|$ is the normalization factor,the subscripts a and n index the ancillary and system qubit registers, respectively. In our case, we focus on the Hamiltonian operator $\hat{\mathcal{H}}$. Using the LCU decomposition, a block-encoding of $\hat{\mathcal{H}}$ can be obtained using two main oracles. These are

$$PREP|0\rangle_a \coloneqq \sum_{k=0}^{K} \sqrt{\frac{c_k}{\lambda_{LCU}}}|k\rangle_a \tag{89}$$

$$SELECT|l\rangle_a|\psi\rangle_s \coloneqq |l\rangle_a \hat{P}_l|\psi\rangle_s. \tag{90}$$

We then have

$$\langle 0|_a PREP^\dagger \cdot SELECT \cdot PREP|0\rangle_a = \frac{1}{\lambda_{LCU}}\hat{\mathcal{H}} \coloneqq \hat{H}. \tag{91}$$

The principal issue with general block-encodings is that it does not necessarily leave the Hilbert space of the system qubits invariant. This problem is then solved by qubitization [35].

5.3 Qubitization

Qubitization [35] has then been proposed to specifically address this issue. Qubitization produces a block-encoding that not only leave the Hilbert space of our system qubits invariant but more specifically it reduces it, the Hilbert space, to a direct sum of 2×2 invariant subspaces, hence the namesake. The walk operator has the following expression

$$\hat{W} \coloneqq \left(2|0\rangle_a\langle 0|_a - \hat{I}_a\right)\left(PREP^\dagger \cdot SELECT \cdot PREP\right). \tag{92}$$

First, it can be seen that $\hat{W}$ is indeed a block-encoding because

$$\langle 0|_a \hat{W}|0\rangle_a = 2\langle 0|PREP \cdot SELECT \cdot PREP|0\rangle - \langle 0|PREP \cdot SELECT \cdot PREP|0\rangle = \hat{H} \tag{93}$$

and

$$\langle 0|_a \hat{W}^2|0\rangle_a = 2\hat{H}^2 - 1 = T_2(\hat{H}). \tag{94}$$

More generally, we have

$$\langle 0|_a \hat{W}^m|0\rangle_a = T_m(\hat{H}). \tag{95}$$

To show eq (95), observe that $\hat{W}$ is the product of two reflections, namely $\hat{R}$ and $\left(PREP^\dagger \cdot SELECT \cdot PREP\right)$. Using Jordan theorem [55], the two reflections can be simultaneously block-diagonalized in the same basis. Let us define $\mathcal{B}_j \coloneqq \{|\lambda_i\rangle, \frac{1}{\sqrt{1-\lambda_i^2}}(\hat{W} - \lambda_i \hat{I})|\lambda_i\rangle)\}$ and the basis $\mathcal{B} \coloneqq \bigcup \mathcal{B}_j$. In $\mathcal{B}_j$, the walk operator reduces to

$$\hat{W}_{|\mathcal{B}_j} = \begin{pmatrix} \lambda_j & \sqrt{1-\lambda_j^2} \\ -\sqrt{1-\lambda_j^2} & \lambda_j \end{pmatrix} = e^{-i\hat{Y}\arccos(\lambda_j)} \tag{96}$$

and in $\mathcal{B}$:

$$\hat{W}_{|\mathcal{B}} = \begin{pmatrix} e^{-i\hat{Y}\arccos(\lambda_0)} & & & \\ & e^{-i\hat{Y}\arccos(\lambda_1)} & & \\ & & \ddots & \\ & & & e^{-i\hat{Y}\arccos(\lambda_{N-1})} \end{pmatrix} \tag{97}$$

therefore

$$\hat{W}^m_{|\mathcal{B}} = \begin{pmatrix} e^{-i\hat{Y}m\arccos(\lambda_0)} & & & \\ & e^{-i\hat{Y}m\arccos(\lambda_1)} & & \\ & & \ddots & \\ & & & e^{-i\hat{Y}m\arccos(\lambda_{N-1})} \end{pmatrix} \tag{98}$$

and

$$\langle 0|\hat{W}^m_{|\mathcal{B}}|0\rangle = \begin{pmatrix} \cos(m\arccos(\lambda_0)) & & & \\ & \cos(m\arccos(\lambda_1)) & & \\ & & \ddots & \\ & & & \cos(m\arccos(\lambda_{N-1})) \end{pmatrix} \tag{99}$$

$$= \begin{pmatrix} T_m(\lambda_0) & & & \\ & T_m(\lambda_1) & & \\ & & \ddots & \\ & & & T_m(\lambda_{N-1}) \end{pmatrix} \tag{100}$$

where we used the definition of the Chebychev polynomial in Eq (74). Hence, T_m can be block-encoded exactly with no approximation error as previously mentioned in Sec 4 giving it consequently an advantage over other real functions that may suffer from approximation errors at finite depth.

5.4 QPE with Walk Operators

As we previously presented, the QPE algorithm remained for a long time conjugated with the time-evoltuion operator and the output phase henceforth equals to the ground state energy of the Hamiltonian up to a given accuracy. However, in [56, 8], they made the critical remark that one can still infer the ground state energy by using the Walk operator Eq (92) bypassing hence the time-evolution operator: "in order to get a good simulation of nature, it is not necessary to imitate nature" [56]. This of course, comes with the benefit of eliminating systematic errors that would have otherwise existed with time-evolution approximation methods such as Trotter-Suzuki product formulas or even the QSP approach. This is possible because, as we have in Eq (98), the walk operator reduces to disjoint block-diagonal rotations $\hat{Y}$. Therefore, the corresponding spectrum is $\bigcup\{\pm\arccos(\lambda_j)\}$ with respective eigenvectors $\bigcup\{\frac{1}{\sqrt{2}}(|\lambda_j\rangle \pm \frac{1}{\sqrt{1-\lambda_j^2}}(\hat{W}-\lambda_j)|\lambda_j\rangle)\}$. Plus, the overlap of the eigenvectors of the walk operator $\hat{W}$ with the starting state ψ_0 is conserved, i.e. equals to the original overlaps with the Hamiltonian operator

$$\langle\psi_0|\frac{1}{\sqrt{2}}(|\lambda_j\rangle \pm \frac{1}{\sqrt{1-\lambda_j^2}}(\hat{W}-\lambda_j)|\lambda_j\rangle) = \frac{1}{\sqrt{2}}\langle\psi_0|\lambda_j\rangle. \tag{101}$$

Finally, the output phases estimate the spectrum of the normalized $\hat{H}$. To infer that of $\hat{\mathcal{H}}$, a multiplication with the normalization factor λ_{LCU} is in order. Therefore, the final depth of the QPE circuit to estimate the ground state energy is $O(\lambda_{LCU}2^{\epsilon^{-1}})$. We can then see that reducing the λ_{LCU} factor is an important step to reduce the overall depth and run time of QPE. Our procedure RC-DF achieves this goal and reduces the λ_{LCU} values approximately $2\times$ compared to THC as we have obtained in our benchmarks [47]. Since then, quantum walk operators have been the standard for resource estimates of QPE [5, 31, 62, 18]. However, recently a new approach [37, 28] has seen the light which in Layman's terms uses the operator $\sqrt{\hat{H}}$ and hence reduces the dependence from λ to $\sqrt{\lambda}$.

5.5 Efficient Clifford+T Bucket-Brigade QRAM Decomposition

State-preparation quantum routines are used in a plethora of quantum algorithms including but not limited to the set of algorithms where a state with good overlap with the ground-state is needed such as QPE [46], SPE [33, 75] KQSD and the tailored and externally corrected couple-cluster with quantum inputs [65] all the way to being used to prepare a block-encoding of an LCU operator [5, 62, 35, 31]. As previously presented, in the context of our work, an input state $|\psi_0\rangle$ with non-vanishing overlap p_0 with the ground-state of the molecular system must be prepared, as well as the flagging state prepared by the PREPARE oracle for the block-encoding $\sum_i \sqrt{\frac{c_i}{\lambda_{LCU}}}|i\rangle$ for both QPE and QKSD. An efficient state-preparation routine is now more than ever a crucial task, as the set of EFTQC methods [75, 33, 29, 17] invoke shallow circuits which in turn makes the state preparation routine no-longer negligible in comparison to the rest of the circuit. The state-preparation research literature is already rich and vast, and there is a host of methods that have been proposed, some of which can be proven to be optimal for a given state structure. To list a few, we have the QNP fabrics [3] for preparing states within the symmetry subspaces of $\hat{N}_\alpha$, $\hat{N}_\beta$ and $\hat{S}^2$. We also have the entangler circuits in [59, 38, 69] that prepare matrix product states (MPS) or sparse states in general using QRAMs [11, 51, 73, 72]. In our work [50], we propose an efficient Clifford+T decomposition of the bucket-brigade QRAM. Said decomposition has a constant, i.e. independent of the system size, and without introducing any additional ancillae. We remark, when it comes to the layout of the Toffoli gates in the bucket-brigade QRAM circuit, that exactly two scenarios

are possible. Namely, the first scenario is in the fanout/fanin where the neighboring Toffoli gates share one single common control qubit and different target qubits. We leverage the flexibility in switching both control qubits when doing the decomposition and achieve $O(1)$ depth Clifford+T decomposition of parallel Toffolis with shared control and different target qubits. The second scenario is in the query section of the QRAM where the Toffoli gates share the target qubit and have different control qubits. We make the remark that these essentially reduce to a series of CCZ gates with different control qubits and sharing the target qubit sandwiched between two Hadamard gates. We use the fact that the target qubit of a CCZ can be exchanged with any of its control qubits, hence recreating the first scenario where we have a control qubit in common and different targets achieving therefore an $O(1)$ depth Clifford+T decomposition of the query section of the QRAM. The final result is a QRAM with linear Clifford+T depth without introducing any ancilla qubits.

5.6 Parallelizing the Queries in a Bucket-brigade Quantum Random Access Memory

Bilbiographic Information

Paler, A., **Oumarou, O.**, & Basmadjian, R. (2020). Parallelizing the queries in a bucket-brigade quantum random access memory. Physical Review A, 102(3). https://doi.org/10.1103/physreva.102.032608

Summary

The bucket-brigade layout is a prominent proposal for quantum random access memories (QRAM). In this work, we optimize the query time of one of the bucket-brigade QRAM by essentially reducing the T-depth (and subsequently) the depth of the circuit to a constant (O(1)). Furthermore, the proposed Clifford+T quantum circuit decomposition of the constant T-depth bucket-brigade does not introduce additional ancilla qubit registers. Therefore, it is very efficient, both in terms of qubit count and circuit depth.

Contribution

O.O wrote the code for implementing and testing the bucket-brigade QRAM and its Clifford+T QRAM decomposition using Cirq, implemented a library of various Clifford+T decomposition of the Toffoli gate, wrote the constant depth Clifford+T decomposition of the bucket-brigade QRAM, generated the benchmark data for estimating the resource count, and co-wrote the manuscript.

Parallelising the Queries in Bucket Brigade Quantum RAM

Alexandru Paler[+,*], Oumarou Oumarou[π], Robert Basmadjian[π]
[+] *University of Transilvania, B-dul Eroilor 29, 500036, Brașov, România*
[*] *Johannes Kepler University, Altenberger Str. 69, 4040, Linz, Austria and*
[π] *Department of Informatics, Clausthal University of Technology, 38678 Clausthal-Zellerfeld, Germany* [*]

Quantum algorithms often use quantum RAMs (QRAM) for accessing information stored in a database-like manner. QRAMs have to be fast, resource efficient and fault-tolerant. The latter is often influenced by access speeds, because shorter times introduce less exposure of the stored information to noise. The total execution time of an algorithm depends on the QRAM access time which includes: 1) address translation time, and 2) effective query time. The bucket brigade QRAMs were proposed to achieve faster addressing at the cost of exponentially many ancillae. We illustrate a systematic method to significantly reduce the effective query time by using Clifford+T gate parallelism. The method does not introduce any ancillae qubits. Our parallelisation method is compatible with the surface code quantum error correction. We show that parallelisation is a result of advantageous Toffoli gate decomposition in terms of Clifford+T gates, and after addresses have been translated, we achieve theoretical $\mathcal{O}(1)$ parallelism for the effective queries. We conclude that, in theory: 1) fault-tolerant bucket brigade quantum RAM queries can be performed approximately with the speed of classical RAM; 2) the exponentially many ancillae from the bucket brigade addressing scheme are a trade-off cost for achieving exponential query speedup compared to quantum read-only memories whose queries are sequential by design. The methods to compile, parallelise and analyse the presented QRAM circuits were implemented in software which is available online.

I. Introduction

Quantum random access memory (QRAM) was proposed for querying (e.g. reading and writing) information during quantum information processing. However, their practical utility is still under study due to the exponential costs associated to explicitly storing and retrieving information [1]. Nevertheless, QRAM and their read-only-variant QROM (quantum read-only memories) [2] are still capturing the attention of researchers due to their conceptual similarity to classical RAMs, and due to the fact that they can be used in subroutines of relevant quantum algorithms, such as the one for linear systems of equations [3].

Whenever QRAM queries are frequent, the speed of the QRAMs influences the total execution time of the quantum algorithms. Moreover, the total execution time of an algorithm impacts the amount of resources necessary for achieving fault-tolerance [4]. Fast QRAMs may prove useful in implementing large scale fault-tolerant quantum computations. Implementations of QRAM circuits were presented in [5] and available online[1].

The access time of a QRAM includes: 1) the time necessary to determine, based on an input set of addresses, the memory locations to query, and 2) the time required to perform the effective queries (e.g. retrieve or store information). The bucket brigade QRAM variant [6] was proposed to reduce the address translation time necessary for computing the memory pointers where the queries should be executed.

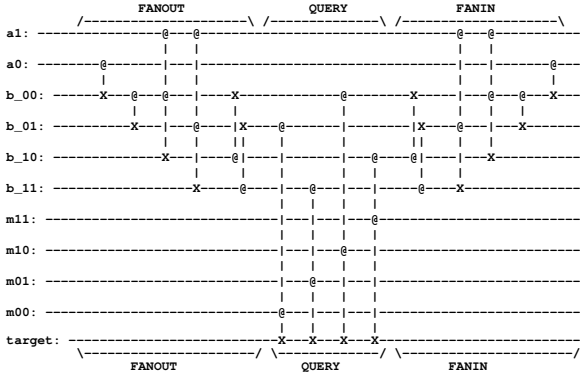


FIG. 1. A bucket brigade circuit consisting of CNOTs and Toffoli gates. The controlled gates from the diagram use @ for controls and the X for target wires. The address wires are named a_i, the memory cells holding single bits are m_i. The classical bits stored on the wires b_i determine which memory wires to read/write.

QRAMs can be described in the form of quantum circuits. Information is stored to and retrieved from the wires/qubits (e.g. analogue to classical memory cell), while quantum gates are used to implement address translation and the queries (e.g. effective read and write operations). In this work, we focus on retrieving information from the bucket brigade QRAM because, due to their structure, reading is more expensive than writing. Once the addresses have been translated, the bucket brigade requires Toffoli (double-controlled-X, or CCX) gates for reading, but only CNOT (controlled-NOT, or controlled-X) gates for writing. We assume the information is already stored in the QRAM database, and that the quantum algorithm queries frequently the QRAM.

For a QRAM with 2^q distinct memory wires (and ex-

* alexandrupaler@gmail.com
[1] https://github.com/quantumresource/quantify

ponential number of ancillae wires, each representing a memory cell that stores a single bit, such as in [7]), there are q bits necessary to address any of the wires. A QRAM circuit takes as input a superposition $|adr_j\rangle$ addresses to be queried (α_js are complex amplitudes), and outputs the contents of the memory cell m_j addressed by $|adr_j\rangle$.

$$\sum_{j=0}^{2^q-1} \alpha_j |adr_j\rangle|0\rangle \xrightarrow{QRAM} \sum_{j=0}^{2^q-1} \alpha_j |adr_j\rangle|m_j\rangle$$

A bucket brigade QRAM circuit consists of three sub-circuits (regions): 1) FANOUT where an exponential number of ancilla bits b_i is used to store control bits for pointing to the corresponding memory cells m_i; 2) QUERY where the memory bits m_i are queried and the results are stored on the ancilla wire $|target\rangle$; 3) FANIN (the uncomputation of FANOUT) in order to maintain reversibility, where the b_i bits are returned to their initial $|0\rangle$ state. The three sub-circuits can be easily recognised in Fig. 1. The QUERY region includes only Toffoli gates.

From the quantum circuit perspective: 1) address-translation duration is equivalent to the circuit depth of the FANOUT/FANIN regions; 2) the effective query speed is analogous to the depth of the QUERY region. The trade-off cost for speeding-up of the address translation is the exponentially large number of b_i ancillae, which translates to additional resource requirement and hence increase in circuit's width. In the standard bucket brigade QRAM model, the queries are considered strictly sequential. Thus, the depth of the QUERY region is dictated by the number of queries. In the worst case, where the entire QRAM is queried, the depth of QUERY is exponentially long [7, 8].

The state-of-the-art Clifford+T formulation of the bucket brigade quantum circuits [7] was obtained from the original formulation [6] and decomposing the Toffoli gates into known Clifford+T gate sequences. The authors of [7] mentioned that important resource savings could be possible if the CCZ (double-controlled-Z gate)/Toffoli transformation is employed in an advantageous manner for the compilation of QRAM. The CCZ gate is obtained from a Toffoli and two Hadamard gates.

This work continues on that line of thought and illustrates some of the possible improvements. First, the depth of the QUERY region is exponentially reduced by parallelising the Toffolis. Second, the herein presented method achieves a linear width improvement compared to the state-of-the-art from [7], where parallelisation was obtained by introducing four ancillae for each Toffoli gate. We show that our method is ancillae free, and reduces the width by a factor of four. Third, the presented QRAMs have a T-count which is almost double to QROM [2], and approximately half compared to the QRAMs from [7] (c.f. T_{bbs}, T_{rom}, T_{bbp} in Sec. II).

The Toffoli construction used in this paper can be interpreted as the enabler of trading depth for width between QROM and QRAM: it exponentially increases the width of the QRAM circuit, while it exponentially decreases the depth of the circuit. Fig. 2 illustrates the improvements obtained by parallelising the QRAM.

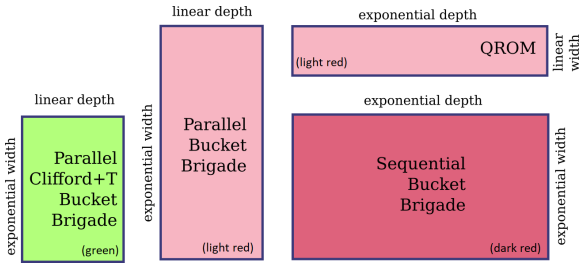


FIG. 2. Comparison between different versions of QRAM. The proposed QRAM is marked green. The parallel bucket brigade (light red) [7] introduces an exponential number of ancillae to achieve parallelism. This is not the case with our decomposition. The original (sequential) bucket brigade circuit (dark red) is sequential and has the same number of wires like the one proposed herein, but requires an exponential number of control signals. The QROM [2] circuit was specifically designed with sequential queries, but does not require the control signals.

In the following, the Results section presents exact resource counts for the proposed bucket brigade QRAMs, and includes a detailed comparison to other QRAM variant layouts. The Methods section discusses how quantum gate level parallelism is achieved, and illustrates how Clifford+T decompositions of Toffoli gates can be used in an advantageous manner for each QRAM sub-circuit. Finally, we present empiric evidence to conjecture that the presented optimisation technique could be applicable to many types of circuits based on multi-controlled-operations [9], which are the common building blocks of quantum adders and multipliers.

In order to achieve a fair comparison between QRAM circuit implementations, we consider $n \leq q$, with 2^n being the number of memory queries, and 2^q denoting the number of memory cells m_i, such that $0 \leq i < 2^q$. The worst case for the number of queries, as used in [7], is for $n=q$.

II. Results

The advantage of bucket brigade QRAM queries is that these can be parallelised. The presented method is implemented in an open-source software[2] [5].

The gate level parallelism used in this work is conceptually very similar to classical instruction level parallelism (see Sec. III A). In practice, the QRAM gate parallelism was lost when translating the circuits to Clifford+T – an often necessary step for estimating the computational resources required for quantum error-correction [4]. The width (number of wires, not including the exponential number of memory cells m_i), T-count,

[2] https://github.com/quantumresource/quantify

and the depth of the QRAM circuits [7] are:

$$Q_{bbs}=q+2^q+5$$
$$T_{bbs}=21\cdot 2^q-28$$
$$D_{bbs}=21\cdot 2^q+2q-26$$

Since their initial formulation, a disadvantage of bucket brigade QRAMs is the fact that an exponentially high number of ancilla b_i bits has to be computed in order to achieve FANOUT parallelism. The exponential overhead is far from ideal, when considering that quantum hardware resources are and will remain very scarce. For this reason, the QROM was proposed in [2]. It does not require the b_i bits, but queries are strictly sequentialised (one after the other). By using optimised Toffoli decompositions, the T-count of the QROM is linear in the number of queries 2^n (the authors of [2] used L for the number of queries). The formulas for width (Q_{rom}) and depth (D_{rom}) do not refer to the memory cells m_i.

$$Q_{rom}=q+1$$
$$T_{rom}=4\cdot 2^n-4$$
$$D_{rom}=10\cdot 2^n+c\cdot\mathcal{O}(2^n)$$

We present a bucket brigade construction that has a reduced T-count, and a depth exponentially shallower than the circuits from [7]. Our construction is achieved using an advantageous parallelisable CCZ/Toffoli decomposition.

$$Q_{bbp}=q+2^q+1 \quad (1)$$
$$T_{bbp}=T_{fanout}+T_{query}=(4\cdot 2^q)+(6\cdot 2^n) \quad (2)$$
$$D_{bbp}=D_{fanout}+D_{query}+D_{fanin}=(10\cdot q)+10+(4\cdot q) \quad (3)$$

Our construction has a Q_{bbp} which is the same as the one from [6, 7]. We do not modify the width of the QRAM, but we optimise the depth of the QRAM. The equivalent quantum volume of the circuit being protected by the surface code is not evaluated in this work, although the parallelism is achieved by using a surface-code compliant parallel CNOT construction. The hardware footprint of bucket brigade QRAM is prohibitively high – this is dictated mostly by the width of the circuit. Having exponentially reduced the QRAM depth and the T-count, maintains the exponentially large width (compared to QROM, $Q_{rom}\lll Q_{bpp}$). For future work, it will be reasonable to estimate the resources necessary for surface code protection, especially when these QRAMs are included in practical quantum algorithms.

III. Methods

This work assumes that, in general, two gates G_1 and G_2 can be parallelised if their commutator $[G_1,G_2]=0$. From a functional perspective, it does not matter in which order the gates are applied: in an ideal setting, the gates can be executed without one depending on the output of the other. This approach to quantum gate parallelism is very similar to classical computing instruction-level parallelism.

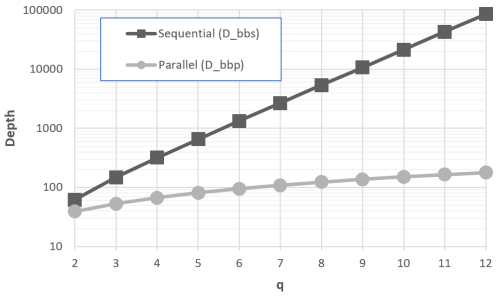


FIG. 3. Comparison between D_{bbs} and D_{bbp}. This plot assumes the worst case where the number of queries 2^n equals the number of memory cells 2^q. The exponent q is plotted along the horizontal axis.

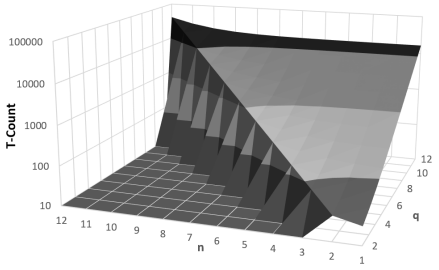


FIG. 4. The T-count of T_{bbp}. The number of queries (2^n) is varied by the parameter n, and the number of memory cells 2^q by the parameter q, such that $n\leq q$. The T-count increases exponentially with q, and the depth (cf. Fig. 3) is linear in q.

A. Parallel CNOTs

For the purpose of this work, two CNOTs are parallel if they either share the controls, or no wires at all (see Fig. 6). CNOTs can also share the target, but this fact will not be used herein. Most of the Clifford+T decompositions are compiled and optimised due to the inherent necessity for quantum error-correction, and this kind of CNOT parallelism is supported by the surface code [10] (braided and lattice surgery variants), for example.

B. Parallelisable CCZ

The three qubit Toffoli gate is in general derived from the CCZ gate, which is a controlled application of the two-qubit CZ gate. The CCZ flips the sign of phase of a state if there is a $|1\rangle$ on all of the gate's three input wires (qubit and wire are used interchangeably). Thus, the phase flip can be expressed as $(-1)^{xyz}$, where x,y,z are the bit values of the inputs: the phase is multiplied by -1 if all three bits are 1, and no phase flip is applied otherwise ($(-1)^0=1$). It has been shown by [9], and more recently in [11], that the following Boolean formula is useful for expressing the CCZ through CNOTs and T

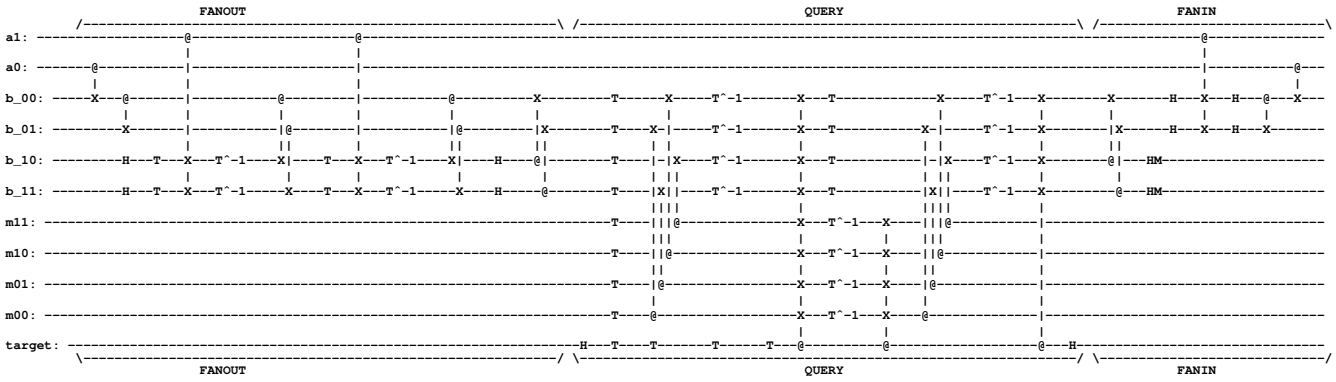


FIG. 5. The equivalent Clifford+T representation of the circuit from Fig. 1. The depth of the circuit is constant for a fixed q by FANIN and FANOUT, while QUERY has a constant depth irrespective of q or n. The circuit operates on 2^q memory cells and executes 2^n queries with $n{\leq}q$. The sequence of four T gates on the target qubit will be cancelled.

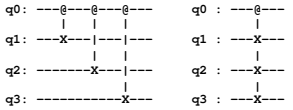


FIG. 6. Parallel CNOTs are allowed to share the control wire.

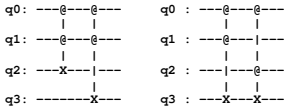


FIG. 7. Toffoli gate parallelism. Two parallel Toffoli gates can either be applied to; left) the same qubits acting as controls, or right) the target qubit and a shared control. Note: In the left diagram, the second Toffoli is not necessary and can be replaced with two CNOT gates, one before and another one after the first Toffoli gate.

gates.

$$4xyz{=}x{+}y{+}z{-}(x{\oplus}y){-}(y{\oplus}z){-}(x{\oplus}z){+}(x{\oplus}y{\oplus}z) \quad (4)$$

The T gate rotates the phase of a state by $\frac{\pi}{4}$, and $-1{=}\omega^4{=}e^{i\frac{\pi}{4}}$, such that $(-1)^{xyz}{=}\omega^{4xyz}$. Thus, seven T gates are necessary, each conditioned on one of the parity sums from Eq. 4.

Eq. 4 is a recipe for generating valid CCZ gate decompositions. Seven parity sums are computed by using CNOTs and T gates, while ensuring that two conditions are met. The first condition is for the T gates to be applied at the right moment, when the necessary bit parities are stored on any of the wires. Second, the parities on each of the three wires have to be uncomputed in order to reflect a correct functionality. Ancillae used for parity computations have to be uncomputed, too. Due to the form of Eq. 4, containing seven parities, the number of ancillae seems to be bounded by seven, but in practice the maximum is four, because three of the parities are formed by single bit values. Consequently, the literature includes a large number of decompositions of the CCZ in terms of Clifford+T gates.

In the following, we use a CCZ decomposition that maintains the Toffoli gate parallelism when decomposed into Clifford+T. The decomposition is obtained after making the observation that two Toffoli gates are parallel whenever they are arranged like in Fig. 7. There are two non-trivial situations: a) one wire is shared; b) two wires are shared. Whenever three wires are shared, the gates cancel each other.

Another practical observation is that, whenever two Toffoli gates are parallel, these can be formulated as CCZ gates which share one or two controls out of the three.

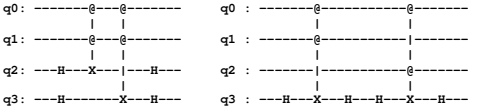


FIG. 8. CCZ gate parallelism is whenever at least one control is shared. In this figure two wires are shared. The two Hadamards in the center of the right-most circuit will cancel.

In order to implement parallel Clifford+T decompositions of Toffoli gates, there has to exist a Clifford+T parallelisable decomposition of CCZ. The decomposition in Fig. 9 can be used whenever two CCZ/Toffoli gates share a single wire. It can be noticed that the wire $|q1\rangle$ acts only as a control for the CNOTs, which are applied to the wires $|q0\rangle$ and $|q2\rangle$, and the T gate commutes with the control of the CNOTs.

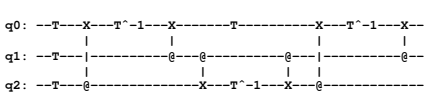


FIG. 9. CCZ gate Clifford+T decomposition that maintains Toffoli gate parallelism when a single wire is shared. The $|q1\rangle$ wire can be shared by two parallel Toffoli gates.

The wire ordering in Fig. 9 does not play any role, because the output will still reflect the $(-1)^{xyz}$ phase flip. Effectively, a Toffoli gate can be obtained by surrounding the CCZ with two Hadamards on any of the wires. Advantageous configurations are whenever the Hadamards are placed such that the wire corresponding to $|q1\rangle$ is

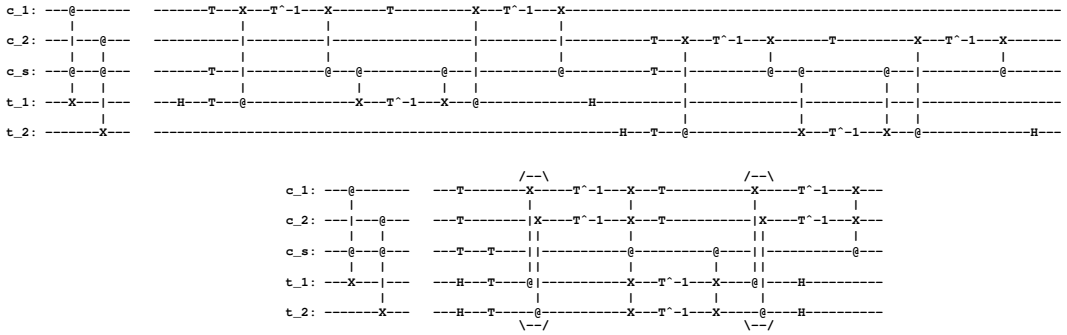


FIG. 10. Two Toffoli gates sharing a control wire. Parallelisation with canonical Toffoli decomposition.

shared. The presented technique is similar to template based quantum circuit optimisation [12], in the sense that the most advantageous Toffoli rewrite rule is chosen from the three possible decompositions based on the expected gate parallelism. During the writing of the manuscript, we found out that the decomposition from Fig. 9 has been also presented in [13], but its effect on the parallelisation of Toffoli/CCZ gates has not been described in the subsequent literature.

C. FANOUT: Shared Control

Whenever two Toffoli gates share a control, the CCZ gate is decomposed such that the shared wire is the one that enables parallelism (i.e. $|q1\rangle$ in Fig. 9). This scenario appears in the FANOUT region of the QRAM.

D. QUERY: Shared Target

The QUERY is formed by a sequence of Toffoli gates conditioned by distinct pairs of (b_i,m_i) wires. The only wire shared is the target. Therefore, the decomposition from Fig. 9 can be used, but by making $|q1\rangle$ correspond to the target. For two consecutive Toffolis the H gates on the target wire will cancel, as illustrated in Fig. 10. T-count optimisation is a side-effect of this kind of parallelism because along the shared target wire the T gates can pairwise be transformed into S gates.

E. FANOUT: Compute Logical AND

Approximate gates have been discussed since [9], but their relevance for quantum circuit design was recognised once T-count optimisation became urgent. This is the case for the approximate CCZ/Toffoli which uses four T gates like in Fig. 12. The approximate Toffoli is also called a logical AND, when the target is an ancilla initialised to $|0\rangle$. However, after a logical AND the state is left with a phase shift, which has to be reversed, once the computed AND bit is not necessary anymore in the circuit (see following section).

Due to their lower T-count it is preferable, whenever possible, to use logical AND gates instead of more general CCZ/Toffoli gates. Two logical ANDs sharing a control can be decomposed like in Fig. 13.

F. FANIN: Uncompute Logical AND

The inverse of the logical AND is its uncomputation. Because the b_i qubits are ancillary, their usage is not necessary after the queries have been executed. Thus, these can be measured and a correctional CZ gate can be applied conditionally on the measurement result.

G. The Parallel Bucket Brigade QRAM

The parallelised QRAM circuit is obtained by concatenating the parallelised circuits for FANOUT, QUERY and FANIN (e.g. Fig. 5). One of the surprising results is that the depth of the FANOUT is constant with respect to q (cf. Eq. 3): the depth is reduced from $\mathcal{O}(Q\log Q)$ where $Q{=}2^q$ to just $\mathcal{O}(\log Q)$. This indicates that there is a significant speedup in computing Q sums of the form $b_j{=}\sum_{i=0}^{q}(a_i{\cdot}2^i)$ where a_i are the bits of the address state vectors used as input to the QRAM (even in superposition).

Another interesting observation is that the T-count T_{bbp} can be reduced with additional 2^n T gates if, after applying the QUERY, the memory is not entangled to the rest of the computation. In that case, the T gates on the m_i wires will affect only the global phase of the m_i states. This can be seen if the decomposition from Fig. 9 is used in reversed gate order, such that, for example, in Fig. 11 half of the T gates applied to c_i in the leftmost column appear at the end of the circuit. Thus, by using parallel Toffoli gate decompositions one could obtain a T-count comparable to logical AND formulations. However, the major disadvantage of logical ANDs is their sequential nature, and being uncomputed through measurements.

IV. Discussion: Practical Speedups

The presented exponential speedups in Sec. II are theoretical. In practice, quantum circuits need to be mapped to hardware, or be compiled to error-corrected structures. The underlying physical architecture determines

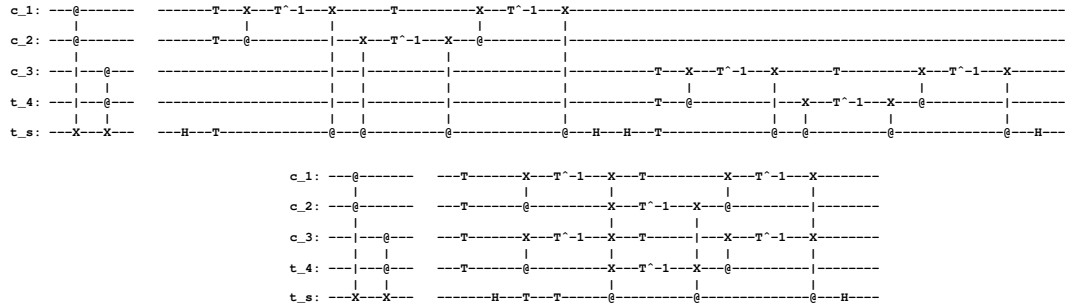


FIG. 11. Two Toffoli gates sharing a target wire. Two of the T gates on the target wire can be reduced to an S gate. Simplification is not illustrated. Furthermore, half of the column of T gates at the beginning of the circuit could be eliminated in some circumstances (see Sec. III G).

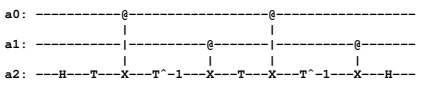


FIG. 12. The parallelisable logical AND has two wires acting always as control. This circuit is similar to the one from [2]. The target wire cannot be shared for parallelisation. If the qubit $|a2\rangle$ is initialised into $|0\rangle$, at the end of this computation will be in the state $(-i)^{a_0 a_1}|a_0 a_1\rangle$ which represents up to a phase shift the correct value of the Boolean AND operation between a_0 and a_1.

the achievable speedup which can be less than the ideally observed (e.g. theoretical) quantum gate level parallelism. Consequently, the available level of parallelism in a circuit may not result in exactly the same execution speedup.

A concern could be that the physical realisations of QRAM will not be able to achieve an exponential speedup using the CNOT parallelisation scheme we presented. The question is if it is practically feasible to parallelise a single control multiple target CNOT, called parallel CNOT in Sec. III A, while maintaining the speedup in practice.

Similar to the theoretical speedup, the practical one is exponential as well. However, the speedup is scaled by some factor (e.g. overhead). Independent of the design choices of the QRAM circuit per se (main focus of this paper), the physical realisation (e.g. implementation which is not the main focus of this paper) of the QRAM in practice requires the usage of additional ancillae. Although we focus on the theoretical speedup, we demonstrate that CNOT parallelism is practically feasible by considering two perspectives: 1) fault-tolerant QRAM protected by the surface code. 2) un-error corrected QRAM;

A. Fault-tolerant QRAM Speedup

This work and the herein presented QUERY parallelisation was originally formulated with the surface code in mind. The surface code (in all its known variants and implementations) supports parallel CNOTs (Sec. III A). The achieved QUERY speedups are $\mathcal{O}(1)$ in the case of surface code protected quantum circuits. Moreover, CNOT logical gates are transversal in the surface code, and the classical post-processing time of these logical gates is constant (see [14]).

The physical realisation of the surface code requires nearest neighbour physical connectivity. Mapping surface code circuits to a two-dimensional lattice or 3D cluster state [15] is an efficient procedure both in terms of resulting number of gates (from logical to physical implementation - the number of qubits and gates to implement a logical operation is approximated by a polynomial function), as well as algorithmic (the mapping procedure has polynomial complexity).

The remaining concern could be related to the arrangements of the surface code protected qubits such that the exponential parallel CNOT complexity is achieved. However, this can be solved without introducing any overhead by ordering the logical qubits in pairs, similarly to the one presented in [6]: a_0,...,a_q,...b_i,m_i,b_{i+1},m_{i+1}...*target*

With respect to the surface code implementation of the bucket brigade QRAM, it should be noted that QRAM query time is exponentially shorter after exponentially reducing the depth of the circuits. The estimations in [7] mention 0.35 ms for querying a 4KB QRAM (15 bit addresses) that is protected by the surface code. The presented optimisation *would* reduce the query time of the surface code error corrected QRAM by a factor of 2^{15} to approximately 11 ns.

B. Un-Error Corrected QRAM Speedup

If the QRAM circuit would be executed un-error-corrected directly on a quantum chip, the connectivity of the chip determines the exact achievable speedup. However, the practical speedup would still be exponential, by making the strong assumption that *sufficient ancillae are available.*

The theoretical exponential speedups are scaled only by a factor on the order of q, which is the number of address qubits (see Sec. I). In particular, for a given address length q, the value of q is a *constant*: once the memory size is known and configured for an algorithm, this cannot be changed anymore.

A parallel CNOT can be implemented using a binary

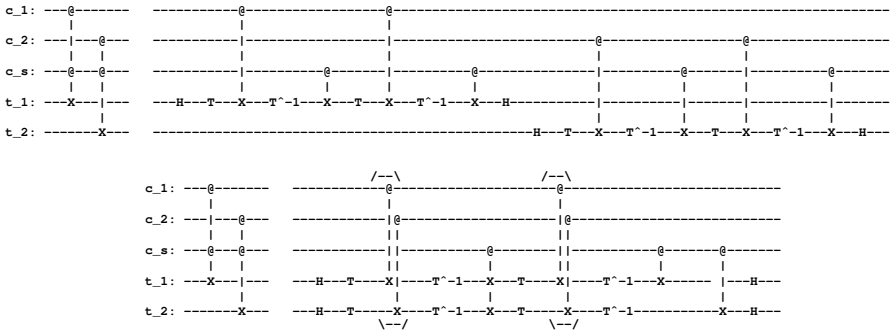


FIG. 13. Two Toffoli gates sharing a control wire. Parallelisation with logical AND versions of the Toffoli gate.

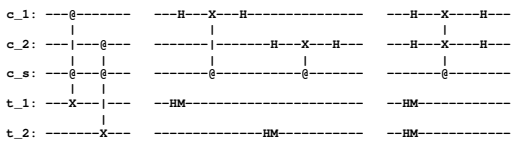


FIG. 14. Logical AND uncompute. HM represents the measurement in the X basis (a Hadamard gate followed by a measurement in the computational basis). Depending on the measurement result a CZ gate correction is applied (here shown always and not conditioned on measurement result). The CZ gate is decomposed with the target being on the wire that is not shared. All measurements can be parallelised and the correction can be applied depending on the individual measurement results. In this figure, a parallel CNOT is applied for the case when both corrections would be required.

tree where the CNOT control is the root, and the CNOT targets are the leaves. The new ancillae are initialised in $|+\rangle$ and used to construct in $\log(2^q)=q$ steps a large GHZ (Greenberger–Horne–Zeilinger) state to encode the state of the control qubit. Afterwards, the parallel CNOT is a transversal application of CNOTs between the GHZ state and the original targets implement. This naive construction illustrates that theoretical $O(1)$ speedups: 1) are delayed by a factor of q which is the time necessary to construct the trees; 2) are not impacted by the number of additional ancillae, although the resource efficiency of this scheme is reduced.

It should be noted that the GHZ construction doubles the number of qubits in the QRAM, from Q_{bbp} to $2{\cdot}Q_{bpp}$. It does not introduce an overhead of the form $2^{Q_{bpp}}$.

The naive tree construction is structurally very similar to the FANOUT region of the bucket brigade QRAM. The difference is that the FANOUT region is performing logical-AND operations (Toffoli gates) which are more complex than the logical-XOR operations necessary for the parallel CNOT.

It is possible to implement the parallel CNOT using less ancillae, compared to the naive construction, by using non-fault-tolerant cluster states [14]. The cluster state and measurement-based formalism are another proof that the exponential speedup is scaled by a constant factor, because "gates from the Clifford group do not contribute to the complexity of a quantum algorithm" [14]. The parallel CNOT is a Clifford gate.

V. Conclusion

Quite a few quantum algorithms require access to information stored in a database like manner. To this end, QRAMs were proposed in general, and the bucket brigade QRAM model in particular.

A bucket brigade QRAM includes three stages: 1) FANOUT, where the input addresses are used to compute exponentially many memory pointers to all possible memory cells; 2) QUERY, where the memory cells are queried using Toffoli gates, and in this process the memory pointers are used to control the queries; 3) FANIN, where the memory pointers are uncomputed, leaving the ancillae wires in their original state before FANIN. The exponentially many memory pointer wires increase the speed of the addressing, but, as shown in this paper, have also another advantage: can be used to massively parallelise the queries when considering the Clifford+T gate set decomposition of the Toffoli gates.

We construct bucket brigade QRAM circuits having a constant QUERY depth. We achieve this by using advantageous Toffoli gate decompositions, and do not introduce any additional ancilla qubits into the QRAM circuit (except the already available memory pointers/control signal wires). The depth, when formulated with Clifford+T gates, is independent of the number of queries, and it depends solely on the number of bits necessary to address the memory. Compared to state of the art bucket brigade implementations, we reduced the depth of QUERY from $\mathcal{O}(2^q)$ to $\mathcal{O}(q)$ in the worst case when any of the 2^q memory cells QRAM are being queried.

Incidentally, the presented construction has a T-count more than half smaller, compared to the existing Clifford+T formulations. Our construction reduces significantly the depth of the topological assembly [10] representing the bucket brigade circuit and, thus, also the distance of the surface code to protect the assembly.

The parallel QUERY construction shows that, if quantum hardware would not be a scarce resource, exponential query speedups are possible compared to state-of-the art QROM designs (Sec. IV). The QROM uses exponen-

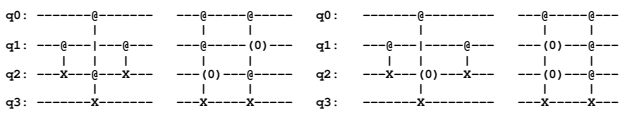


FIG. 15. Parallelising operations by introducing Toffolis ((0) represents a negative control) reduces computational depth. In the worst case, where no other optimisations are available, it increases T-count. These circuit identities are inverse to the ones from [17].

tially less wires than bucket brigade, but has exponentially slower querying in the worst case. The QROM circuits execute queries sequentially, and to the best of our knowledge no parallel construction seems possible.

The parallelisation of bucket brigade was achieved without introducing any additional ancillae, and future work will parallelise quantum circuits using templates like the one presented in Fig. 15. It may seem counter intuitive, that it may be more resource efficient to include T gates while not introducing ancilla: in Fig. 15 a single Toffoli is transformed into two Toffoli gates, thus the number of T gates increases from 7 to 14. However, the introduced T gates may cancel, due to efficient Toffoli gate parallelism in the overall circuit. We argue, that the cost of large scale quantum error-corrected quantum circuits is not necessarily related to T-counts and state distillations, but to the error-corrected associated with the error-corrected Clifford operations [16]. This includes identity operations on unused wires and ancillae, as for example in carry save adders.

Acknowledgments

A.P. was supported by a Google Faculty Research Award, and the NUQAT project funded by the University Transilvania Brasov. We thank Olivia Di Matteo for her very valuable feedback during the preparation of the circuits and the writing of the manuscript, and Simon Devitt for feedback on the final manuscript.

[1] M. Schuld and F. Petruccione, *Supervised learning with quantum computers*, Vol. 17 (Springer, 2018).
[2] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding electronic spectra in quantum circuits with linear t complexity, Physical Review X **8**, 041015 (2018).
[3] P. J. Coles, S. Eidenbenz, S. Pakin, A. Adedoyin, J. Ambrosiano, P. Anisimov, W. Casper, G. Chennupati, C. Coffrin, H. Djidjev, *et al.*, Quantum algorithm implementations for beginners, arXiv preprint arXiv:1804.03719 (2018).
[4] A. Paler, D. Herr, and S. J. Devitt, Really small shoe boxes: On realistic quantum resource estimation, Computer **52**, 27 (2019).
[5] O. Oumarou, A. Paler, and R. Basmadjian, QUANTIFY: A framework for resource analysis and design verification of quantum circuits, in *Computer Society Annual Symposium on VLSI*, Vol. 1 (IEEE, 2020) pp. 1–13.
[6] S. Arunachalam, V. Gheorghiu, T. Jochym-O'Connor, M. Mosca, and P. V. Srinivasan, On the robustness of bucket brigade quantum ram, New Journal of Physics **17**, 123010 (2015).
[7] O. D. Matteo, V. Gheorghiu, and M. Mosca, Fault-tolerant resource estimation of quantum random-access memories, IEEE Transactions on Quantum Engineering **1**, 1 (2020).
[8] J. Bang, A. Dutta, S.-W. Lee, and J. Kim, Optimal usage of quantum random access memory in quantum machine learning, Physical Review A **99**, 012326 (2019).
[9] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Elementary gates for quantum computation, Physical review A **52**, 3457 (1995).
[10] A. Paler, A. G. Fowler, and R. Wille, Synthesis of arbitrary quantum circuits to topological assembly: Systematic, online and compact, Scientific reports **7**, 1 (2017).
[11] C. Gidney, Halving the cost of quantum addition, Quantum **2**, 74 (2018).
[12] D. Maslov, C. Young, D. M. Miller, and G. W. Dueck, Quantum circuit simplification using templates, in *Design, Automation and Test in Europe* (IEEE, 2005) pp. 1208–1213.
[13] M. Amy, D. Maslov, and M. Mosca, Polynomial-time t-depth optimization of clifford+ t circuits via matroid partitioning, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **33**, 1476 (2014).
[14] R. Raussendorf, D. E. Browne, and H. J. Briegel, Measurement-based quantum computation on cluster states, Physical review A **68**, 022312 (2003).
[15] A. Paler, S. J. Devitt, K. Nemoto, and I. Polian, Mapping of topological quantum circuits to physical hardware, Scientific reports **4**, 1 (2014).
[16] A. Paler and R. Basmadjian, Clifford gate optimisation and t gate scheduling: Using queueing models for topological assemblies, in *2019 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)* (IEEE, 2019) pp. 1–5.
[17] M. Z. Rahman and J. E. Rice, Templates for positive and negative control toffoli networks, in *International Conference on Reversible Computation* (Springer, 2014) pp. 125–136.

6 Discussion

In this dissertation, we first sought to tackle the significant overhead of energy measurement that -as we argue later in this section- poses one of the major challenges in scaling the VQE set of methods to large and industrially relevant systems. This overhead is directly reflected in the number of distinct observables that need to be measured and their corresponding number of samples. To address the problem, we proposed RC-DF, in Section 1.4, which successfully overcomes both problems. Conceptually, RC-DF is a purely classical preprocessing method that optimizes and compresses the number of terms of the molecular Hamiltonian double-factorized LCU while constraining the coefficients (of the LCU in question) to be small.

As shown in the benchmarks, RC-DF produces the best variances compared to then-state-of-the-art double-factorization and Pauli grouping schemes with a reduction factor ranging from $\times 3$ to $\times 6$. This in turn, yields a smaller number of samples for a given accuracy of energy evaluations. Such improvement is a direct result of the relatively small prefactors of the LCU produced by the RC-DF optimization. In addition, RC-DF compresses the number of observables to be measured from quartic to a quasi-linear quantity. These observables consist exclusively of the diagonal operators $\hat{Z}_k$ and $\hat{Z}_k\hat{Z}_l$, therefore they are jointly measurable (thanks to their commutativity). The measurements in these bases can be readily used to invoke symmetry-preserving error-mitigation techniques and without additional circuit overhead.

However, such benefits of RC-DF, both in number of observables and number of shots, come at the cost of appending a circuits to implement the resulting unitaries of the RC-DF LCU ($\{U^t, t \in \{0, \ldots, T-1\}\}$). These circuits can be implemented with a quadratic number of Givens gates laid out in a linear depth.

It should be noted that other prominent optimization-based factorization methods have since seen light, namely [1, 36].

More generally, RC-DF benefits transcend VQE to affect the QPE family of algorithms as we will present later in this section when we get to the FTQC-related results of the thesis.

We next set to develop a quantum algorithm to determine the gradient of energies estimated with RC-DF as well as other double-factorized methods, namely XDF and CDF. As explained in Section , energy gradients are fundamentally crucial to various quantum chemistry methods and applications. Coupled with the prominence of double-factorization methods such as XDF, CDF and our RC-DF, it becomes evidently important to work a quantum algorithm to efficiently calculate the gradient of the energy estimators of these methods. Therefore, in Sections 2.1 and 2.2, we successfully proposed an efficient quantum algorithm that determines the analytic gradient of the VQE energy measured with respect to double-factorized Hamiltonians. The central task is to derive an expression for the relaxed one- and two-particles RDMs and the gradient is then subsequently obtained through simple chain-rule that is effectively a tensor contraction involving the relaxed RDMs. We employed Lagrangian formalism to obtain said RDMs where the Lagrangians we devised, by construction, fulfill the constraint required by the XDF and RC-DF LCU parameters (the unitaries and their respective coefficients). We tested the correctness by comparing the molecular nuclear gradient with respect to the complete and truncated Hamiltonian to that obtained through finite-difference, where the results showed a high level of accuracy ranging from 10^-6 to 10^{-7}. Furthermore, we performed two other benchmarks, namely, QM/MM simulations of 327 quantum and 18470 total atoms and a geometry optimization of reactant, product and transition state geometries for the propylene oxide to allyl alcohol isomerization reaction.

Our algorithm depend however on the gradient of the energy with respect to the angles parameters used to implement the unitaries of the LCU. We propose to use parameter-shift rules to obtain such gradients. This can be potentially a bottle-neck if the final result is highly sensitive to it.

A follow-up work addressing this particular issue, aimed at analyzing and quantifying the algorithm's sensitivity –and eventually exploring strategies to enhace its robustness– would be fundamental for deploying the algorithm on hardware.

Next, given the importance of error-mitigation in the NISQ era, we sought to experimentally evaluate the performance of the two most prominent purification error-mitigation techniques presented in Section 3.1, namely echo-verification and virtual distillation. We presented the results of these experiments in Section 3.2. We performed record (at the time) 20-qubit VQE experiments. In the first experiment, we calculated, with and without error-mitigation, the energies of a pre-optimized UpCCD ansatz of the molecule of cyclobutene along the coordinates of the ring opening reaction. In the second experiment, we similarly determined the energy of the pre-optimized UpCCD ansatz of the Richardson-Gaudin model. We observed that error-mitigated energies, in both experiments, have one to two orders of magnitude improvement compared to both raw, noisy and symmetry-based post-selected energies. To conclude, we extrapolated the overhead needed in terms of sampling complexity for classically intractable systems, which we will comment on its impact for achieving quantum advantage in the next paragraph.

This concludes the results that are related to VQE and more generally to NISQ, and before continuing with the rest of the results (which are more on the FTQC side of the spectrum), let us discuss the viability of achieving quantum advantage with VQE in quantum chemistry.

To gauge the capacity of VQE algorithms to achieve quantum advantage –that is, to successfully produce

ground state energies of classically intractable systems– we need to examine where the bottleneck lies among its composing modules (ansatz, measurement and optimization).

Regarding the measurement module, we note again that RC-DF produces a quasi-linear number of observables and a low energy variance that can even be potentially optimized further with state-of-the-art shadow protocols such as [34, 80]. As to choice of ansatz, we have a set of sensible choices such as the proven highly expressive and symmetry-preserving ansätze [3].

It can then be affirmed, with a fair level of certainty, that the major roadblocks to achieving quantum advantage with VQE in quantum chemistry specifically, are mostly the optimization and error-circumvention approaches.

The gradient-based optimization methods face a set of challenges, not the least of which is the evaluation of the energy gradient with respect to the ansatz parameters. This figures to be a very hard and challenging task. In particular, using the prominent parameter-shift rules [77, 67, 6], adds on the complexity of the number of distinct circuits for energy evaluation by a multiplicative factor that scales, at least, with the number of parameters of the ansatz.

In addition to that, when scaling up the sizes of molecular system and the ansätze, the gradient-based optimization over high-dimensional space is prone to the barren plateaus problem caused by exponentially vanishing gradients, which in turn requires likewise accuracies of the gradient values. Hence, it adds a prohibitive sampling budget to the already costly number of distinct circuits generated by the parameter-shift rules.

This has, consequently, led to the realization that perhaps the way we approach optimization in this setup must be fundamentally different from the classical gradient-based optimization. This resulted in quantum gradient-free optimization methods such as [12], albeit at the cost of implementing time-evolution. Some other approaches assume the convexity of the energy landscape in a given voisinage of the initial parameters and subsequently fit a model cost function with the result of energy measurements of a mesh of points in the parameter space in said voisinage. Meaning free and exact gradient calculations on the fitted model cost function [23, 30, 71] otherwise known as surrogate models.

All of these are interesting and very important milestones, even. However, to my knowledge, there has been no definitive optimization approach with cheap or reasonably scaling overhead that, even assuming access to a perfect quantum computer, will unlock quantum advantage through VQE for quantum chemistry.

The second and equally major challenge is how to deal with errors that result from faulty hardware, in general, regardless of their nature or origin. The purification-based error-mitigation techniques, while they delivered as proven in our experiments [45], they do suffer from a prohibitive projected overhead in number of samples –partly due to the chemical accuracy requirements– and circuit complexity. This raises the question of whether there could be a practical use for VQE with error-mitigation only and without error-correction. If not, would it be wise to use VQE with error-correction in the FTQC regime. So far, the response, in my humble opinion, is probably "no". Because there is already a set of (E)FTQC methods, such as the phase estimation –in both its deep or statistical and shallow versions– and the QKSD approaches that do not employ optimization loops and come with analytical guaranties. We further note that in this FTQC regime, our RC-DF [47] has a direct impact, thus transcending NISQ and VQE. Expressly, as previously mentioned, RC-DF, through the small set of prefactors thanks to regularization, directly reduces the normalization factor to which the QPE circuit depth is propotional. Thus, RC-DF reduces the circuit depth (in terms of calls to the Walk operator $\hat{W}$) of QPE).

However, recent work by Rocca et al [61], shows that RC-DF will effectively incur a bigger qubit count compared to the exact DF. This is due to the full-rank property of the prefactors tensors produced by RC-DF which in turn directly yields a higher qubit count compared to the exact DF that produces single rank prefactor tensors (See [61] for a full qubit count breakdown). To circumvent this, an additional constraint is imposed in [61] –on top of our regularization constraint in the RC-DF optimization [47]– in the cost function, which guaranties a low rank of the prefactors' tensors.

We now continue with the second part of our results that is FTQC-centered. Given the withering faith in NISQ and the gradual yet stark realization of the necessity of error-correction for quantum algorithms to be of any utility to industrial scale use-cases, the field has already started anticipating the FTQC era with a remarkable focus over the last few years on the EFTQC. Both families of the statistical phase estimation and quantum Krylov methods are classified as EFTQC (ruling out VQE for the reasons we have listed earlier). Compared to FTQC algorithms such as QPE, they share a common trait, that is, they trade depth for number of repetitions.

In this context, we identified in Section 4 the problem of determining the gradient of quantum Krylov energies, that is the quadratic scale (in Krylov subspace dimension) of relaxed RDM calculation in terms of observables compared to the linear scale of the cost of energy estimation. To overcome this issue, we have developed in Section 4.3, to our knowledge, the first method to determine the gradient of quantum Krylov energies [49]. The core idea is the observation that the QKSD ground-state can be expressed as polynomial of the Hamiltonian applied to an initial state. Hence, we propose to use QSP to efficiently prepare this state, without systematic error. Our approach is analytic and efficient in both qubit count and circuit depth. In fact,

the QSP circuit has exactly the same depth (in terms of the number of Walk operators) of the largest circuit used in the QKSD and similar order of Toffoli gates. Subsequently, the relaxed 1- and 2-particles RDMs are obtained by directly measuring the prepared the prepared states with respect to the corresponding operators. Therefore, we reduced the cost to O(1) observable (per RDM) and an overall linear scale for the quantum Krylov energy gradient.

Furthermore, we also proposed a coherent method for the evaluation of these quantum Krylov energy gradients that bypasses post-selecting on the right subspace of the block-encoding. This consequently resulted in an improvement of around 2× in the number of samples for a target accuracy.

Although QKSD is an EFTQC algorithm, by virtue of its shallow circuit, a proper and concrete estimation of the required resources is still critical to put it in perspective vis-à-vis QPE and SPE. Furthermore, the predominantly part of the existing work on QKSD yielded loose asymptotic scalings and upper-bounds. Therefore, a numerically informed resource estimate would be practically insightful.

The shared similarities between QKSD and SPE naturally leads to the speculation that there is perhaps latent link between both approaches. Meaning that both methods potentially can be ran using the same set of inputs. This will help to better compare and understand the characteristics of both methods. We investigate these insights in a separate follow-up work.

Lastly, QRAMs are an integral part of several quantum routines such as state preparation used in various quantum algorithms, not the least of which the QPE algorithm. The bucket-brigade is a prominent QRAM circuit layout that originally has a linear depth in Toffoli gates. However, the problem arise when the QRAM circuit is decomposed into elementary Clifford+T gates as the resulting circuit depth becomes exponential. This constitutes a major obstacle that hinders its practical utility. We have successfully overcome this challenge with a proposal of an efficient bucket-brigade QRAM Clifford+T decomposition that maintains the depth linearity without introducing additional ancilla qubits. The key observation that enabled such a result is the commutation of the control qubits of a Toffoli gate and the fact that in the bucket-brigade layout two neighboring Toffoli gates share either a single control qubit or the target qubit.

Availability of primary data

For transparency and reproducibility, we provide open access to the majority of the data and software developed in this thesis. All simulation data are available or can be accessed upon request, and several of the software tools have been released as open-source. The benchmark data for RC-DF presented in Section 1.4 are publicly available on Zenodo [48] at `https://doi.org/10.5281/zenodo.7866658` , along with instructions for loading the data in Python. Similarly, the data supporting the purification-based error-mitigation experiments discussed in Section 3.2 can be accessed on Zenodo [43] at `https://doi.org/10.5281/zenodo.7225821` . Additionally, the implementation of the QSP iterative solver has been integrated into the open-source quantum software toolkit Pennylane, with documentation available at `https://docs.pennylane.ai/en/stable/code/api/pennylane.poly_to_angles.html`. Some of the software tools developed to validate theoretical results and run certain simulations, remains proprietary due to the industrial collaboration.

References

[1] DOI: 10.1021/acs.jctc.4c01390.s001. URL: http://dx.doi.org/10.1021/acs.jctc.4c01390.s001.

[2] Eva Pavarini et al., eds. *The LDA+DMFT approach to strongly correlated materials*. Vol. 1. Schriften des Forschungszentrums Jülich. Reihe modeling and simulation. Record converted from VDB: 12.11.2012. Jülich: Forschungszenrum Jülich GmbH Zentralbibliothek, Verlag, 2011, getr. Zählung. URL: https://juser.fz-juelich.de/record/17645.

[3] Gian-Luca R Anselmetti et al. "Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems". In: *New Journal of Physics* 23.11 (Nov. 2021), p. 113010. ISSN: 1367-2630. DOI: 10.1088/1367-2630/ac2cb3. URL: http://dx.doi.org/10.1088/1367-2630/ac2cb3.

[4] Brian M. Austin, Dmitry Yu. Zubarev, and William A. Jr. Lester. "Quantum Monte Carlo and Related Approaches". In: *Chemical Reviews* 112.1 (2012). PMID: 22196085, pp. 263–288. DOI: 10.1021/cr2001564. eprint: https://doi.org/10.1021/cr2001564. URL: https://doi.org/10.1021/cr2001564.

[5] Ryan Babbush et al. "Encoding Electronic Spectra in Quantum Circuits with Linear T Complexity". In: *Physical Review X* 8.4 (Oct. 2018). ISSN: 2160-3308. DOI: 10.1103/physrevx.8.041015. URL: http://dx.doi.org/10.1103/physrevx.8.041015.

[6] Leonardo Banchi and Gavin E. Crooks. "Measuring Analytic Gradients of General Quantum Evolution with the Stochastic Parameter Shift Rule". In: *Quantum* 5 (Jan. 2021), p. 386. ISSN: 2521-327X. DOI: 10.22331/q-2021-01-25-386. URL: http://dx.doi.org/10.22331/q-2021-01-25-386.

[7] Ewout van den Berg et al. "Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors". In: *Nature Physics* 19.8 (May 2023), pp. 1116–1121. ISSN: 1745-2481. DOI: 10.1038/s41567-023-02042-2. URL: http://dx.doi.org/10.1038/s41567-023-02042-2.

[8] Dominic W. Berry et al. "Improved techniques for preparing eigenstates of fermionic Hamiltonians". In: *npj Quantum Information* 4.1 (May 2018). ISSN: 2056-6387. DOI: 10.1038/s41534-018-0071-5. URL: http://dx.doi.org/10.1038/s41534-018-0071-5.

[9] X. Bonet-Monroig et al. "Low-cost error mitigation by symmetry verification". In: *Physical Review A* 98.6 (Dec. 2018). ISSN: 2469-9934. DOI: 10.1103/physreva.98.062339. URL: http://dx.doi.org/10.1103/physreva.98.062339.

[10] Sergey B. Bravyi and Alexei Yu. Kitaev. "Fermionic Quantum Computation". In: *Annals of Physics* 298.1 (May 2002), pp. 210–226. ISSN: 0003-4916. DOI: 10.1006/aphy.2002.6254. URL: http://dx.doi.org/10.1006/aphy.2002.6254.

[11] Pablo Antonio Moreno Casares. "Circuit implementation of bucket brigade qRAM for quantum state preparation". In: *arXiv preprint arXiv:2006.11761* (2020).

[12] Ahmet Burak Catli, Sophia Simon, and Nathan Wiebe. "Exponentially Better Bounds for Quantum Optimization via Dynamical Simulation". In: *arXiv preprint arXiv:2502.04285* (2025).

[13] Garnet Kin-Lic Chan et al. "An Introduction to the Density Matrix Renormalization Group Ansatz in Quantum Chemistry". In: *Frontiers in Quantum Systems in Chemistry and Physics*. Springer Netherlands, pp. 49–65. ISBN: 9781402087073. DOI: 10.1007/978-1-4020-8707-3_4. URL: http://dx.doi.org/10.1007/978-1-4020-8707-3_4.

[14] Jiří Čížek. "On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods". In: *The Journal of Chemical Physics* 45.11 (Dec. 1966), pp. 4256–4266. ISSN: 0021-9606. DOI: 10.1063/1.1727484. eprint: https://pubs.aip.org/aip/jcp/article-pdf/45/11/4256/18845884/4256_1_online.pdf. URL: https://doi.org/10.1063/1.1727484.

[15] William R. Clements et al. "Optimal design for universal multiport interferometers". In: *Optica* 3.12 (Dec. 2016), p. 1460. ISSN: 2334-2536. DOI: 10.1364/optica.3.001460. URL: http://dx.doi.org/10.1364/optica.3.001460.

[16] Jeffrey Cohn, Mario Motta, and Robert M. Parrish. "Quantum Filter Diagonalization with Compressed Double-Factorized Hamiltonians". In: *PRX Quantum* 2.4 (Dec. 2021). ISSN: 2691-3399. DOI: 10.1103/prxquantum.2.040352. URL: http://dx.doi.org/10.1103/prxquantum.2.040352.

[17] Yulong Dong et al. "Efficient phase-factor evaluation in quantum signal processing". In: *Physical Review A* 103.4 (Apr. 2021). ISSN: 2469-9934. DOI: 10.1103/physreva.103.042419. URL: http://dx.doi.org/10.1103/physreva.103.042419.

[18] Alicja Dutkiewicz et al. *Error mitigation and circuit division for early fault-tolerant quantum phase estimation*. 2024. arXiv: 2410.05369 [quant-ph]. URL: https://arxiv.org/abs/2410.05369.

[19] Ethan N Epperly, Lin Lin, and Yuji Nakatsukasa. "A theory of quantum subspace diagonalization". In: *SIAM Journal on Matrix Analysis and Applications* 43.3 (2022), pp. 1263–1290.

[20] András Gilyén et al. "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*. STOC '19. ACM, June 2019, pp. 193–204. DOI: 10.1145/3313276.3316366. URL: http://dx.doi.org/10.1145/3313276.3316366.

[21] D.R. HARTREE. "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion". In: *Self-Consistent Fields in Atoms*. Elsevier, 1975, pp. 167–193. ISBN: 9780080178196. DOI: 10.1016/b978-0-08-017819-6.50028-9. URL: http://dx.doi.org/10.1016/b978-0-08-017819-6.50028-9.

[22] Andre He et al. "Zero-noise extrapolation for quantum-gate error mitigation with identity insertions". In: *Physical Review A* 102.1 (July 2020). ISSN: 2469-9934. DOI: 10.1103/physreva.102.012426. URL: http://dx.doi.org/10.1103/physreva.102.012426.

[23] Magnus R. Hestenes and Eduard Stiefel. "Methods of Conjugate Gradients for Solving Linear systems". In: *Journal of Research of the National Bureau of Standards* 49.6 (Dec. 1952), pp. 409–436. DOI: 10.6028\%2Fjres.049.044. URL: https://doi.org/10.6028%5C%2Fjres.049.044.

[24] Edward G. Hohenstein et al. "Efficient quantum analytic nuclear gradients with double factorization". In: *The Journal of Chemical Physics* 158.11 (Mar. 2023). ISSN: 1089-7690. DOI: 10.1063/5.0137167. URL: http://dx.doi.org/10.1063/5.0137167.

[25] William J. Huggins et al. "Unbiasing fermionic quantum Monte Carlo with a quantum computer". In: *Nature* 603.7901 (Mar. 2022), pp. 416–420. ISSN: 1476-4687. DOI: 10.1038/s41586-021-04351-z. URL: http://dx.doi.org/10.1038/s41586-021-04351-z.

[26] William J. Huggins et al. "Virtual Distillation for Quantum Error Mitigation". In: *Phys. Rev. X* 11 (4 Nov. 2021), p. 041036. DOI: 10.1103/PhysRevX.11.041036. URL: https://link.aps.org/doi/10.1103/PhysRevX.11.041036.

[27] Abhinav Kandala et al. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets". In: *Nature* 549.7671 (Sept. 2017), pp. 242–246. ISSN: 1476-4687. DOI: 10.1038/nature23879. URL: http://dx.doi.org/10.1038/nature23879.

[28] Robbie King et al. "Quantum simulation with sum-of-squares spectral amplification". In: *arXiv preprint arXiv:2505.01528* (2025).

[29] William Kirby, Mario Motta, and Antonio Mezzacapo. "Exact and efficient Lanczos method on a quantum computer". In: *Quantum* 7 (May 2023), p. 1018. ISSN: 2521-327X. DOI: 10.22331/q-2023-05-23-1018. URL: http://dx.doi.org/10.22331/q-2023-05-23-1018.

[30] Andrew V. Knyazev and Ilya Lashuk. "Steepest Descent and Conjugate Gradient Methods with Variable Preconditioning". In: *SIAM Journal on Matrix Analysis and Applications* 29.4 (2008), pp. 1267–1280. DOI: 10.1137/060675290. eprint: https://doi.org/10.1137/060675290. URL: https://doi.org/10.1137/060675290.

[31] Joonho Lee et al. "Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction". In: *PRX Quantum* 2.3 (July 2021). ISSN: 2691-3399. DOI: 10.1103/prxquantum.2.030305. URL: http://dx.doi.org/10.1103/prxquantum.2.030305.

[32] Jörg Liesen and Petr Tichý. "Convergence analysis of Krylov subspace methods". In: *GAMM-Mitteilungen* 27.2 (Dec. 2004), pp. 153–173. ISSN: 1522-2608. DOI: 10.1002/gamm.201490008. URL: http://dx.doi.org/10.1002/gamm.201490008.

[33] Lin Lin and Yu Tong. "Heisenberg-Limited Ground-State Energy Estimation for Early Fault-Tolerant Quantum Computers". In: *PRX Quantum* 3.1 (Feb. 2022). ISSN: 2691-3399. DOI: 10.1103/prxquantum.3.010318. URL: http://dx.doi.org/10.1103/prxquantum.3.010318.

[34] Guang Hao Low. "Classical shadows of fermions with particle number symmetry". In: *arXiv preprint arXiv:2208.08964* (2022).

[35] Guang Hao Low and Isaac L. Chuang. "Hamiltonian Simulation by Qubitization". In: *Quantum* 3 (July 2019), p. 163. ISSN: 2521-327X. DOI: 10.22331/q-2019-07-12-163. URL: http://dx.doi.org/10.22331/q-2019-07-12-163.

[36] Guang Hao Low et al. "Fast Quantum Simulation of Electronic Structure by Spectral Amplification". In: *Physical Review X* 15.4 (Oct. 2025). ISSN: 2160-3308. DOI: 10.1103/pb2g-j9cw. URL: http://dx.doi.org/10.1103/pb2g-j9cw.

[37] Guang Hao Low et al. "Fast quantum simulation of electronic structure by spectrum amplification". In: *arXiv preprint arXiv:2502.15882* (2025).

[38] Daniel Malz et al. "Preparation of Matrix Product States with Log-Depth Quantum Circuits". In: *Physical Review Letters* 132.4 (Jan. 2024). ISSN: 1079-7114. DOI: 10.1103/physrevlett.132.040404. URL: http://dx.doi.org/10.1103/physrevlett.132.040404.

[39] Danial Motlagh and Nathan Wiebe. "Generalized Quantum Signal Processing". In: *PRX Quantum* 5.2 (June 2024). ISSN: 2691-3399. DOI: 10.1103/prxquantum.5.020368. URL: http://dx.doi.org/10.1103/prxquantum.5.020368.

[40] Mario Motta et al. "Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution". In: *Nature Physics* 16.2 (Nov. 2019), pp. 205–210. ISSN: 1745-2481. DOI: 10.1038/s41567-019-0704-4. URL: http://dx.doi.org/10.1038/s41567-019-0704-4.

[41] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[42] T. E. O'Brien et al. "Purification-based quantum error mitigation of pair-correlated electron simulations". In: *Nature Physics* 19.12 (Dec. 2023), pp. 1787–1792. ISSN: 1745-2481. DOI: 10.1038/s41567-023-02240-y. URL: https://doi.org/10.1038/s41567-023-02240-y.

[43] Thomas E. O'Brien et al. *Raw and processed data for 'Purification-based quantum error mitigation of pair-correlated electron simulations'*. Version 1.0.0. Zenodo, Oct. 2022. DOI: 10.5281/zenodo.7225821. URL: https://doi.org/10.5281/zenodo.7225821.

[44] Thomas E. O'Brien et al. "Efficient quantum computation of molecular forces and other energy gradients". In: *Physical Review Research* 4.4 (Dec. 2022). ISSN: 2643-1564. DOI: 10.1103/physrevresearch.4.043210. URL: http://dx.doi.org/10.1103/physrevresearch.4.043210.

[45] Thomas E. O'Brien et al. "Error Mitigation via Verified Phase Estimation". In: *PRX Quantum* 2.2 (May 2021). ISSN: 2691-3399. DOI: 10.1103/prxquantum.2.020317. URL: http://dx.doi.org/10.1103/prxquantum.2.020317.

[46] Pauline J. Ollitrault et al. "Enhancing Initial State Overlap through Orbital Optimization for Faster Molecular Electronic Ground-State Energy Estimation". In: *Physical Review Letters* 133.25 (Dec. 2024). ISSN: 1079-7114. DOI: 10.1103/physrevlett.133.250601. URL: http://dx.doi.org/10.1103/physrevlett.133.250601.

[47] Oumarou Oumarou et al. "Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization". In: *Quantum* 8 (June 2024), p. 1371. ISSN: 2521-327X. DOI: 10.22331/q-2024-06-13-1371. URL: http://dx.doi.org/10.22331/q-2024-06-13-1371.

[48] Oumarou Oumarou et al. *Data for Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization*. Zenodo, Apr. 2023. DOI: 10.5281/zenodo.7866658. URL: https://doi.org/10.5281/zenodo.7866658.

[49] Oumarou Oumarou et al. "Molecular Properties from Quantum Krylov Subspace Diagonalization". In: *Journal of Chemical Theory and Computation* 21.9 (Apr. 2025), pp. 4543–4552. ISSN: 1549-9626. DOI: 10.1021/acs.jctc.5c00194. URL: http://dx.doi.org/10.1021/acs.jctc.5c00194.

[50] Alexandru Paler, Oumarou Oumarou, and Robert Basmadjian. "Parallelizing the queries in a bucket-brigade quantum random access memory". In: *Physical Review A* 102.3 (Sept. 2020). ISSN: 2469-9934. DOI: 10.1103/physreva.102.032608. URL: http://dx.doi.org/10.1103/physreva.102.032608.

[51] Daniel K. Park, Francesco Petruccione, and June-Koo Kevin Rhee. "Circuit-Based Quantum Random Access Memory for Classical Data". In: *Scientific Reports* 9.1 (Mar. 2019). ISSN: 2045-2322. DOI: 10.1038/s41598-019-40439-3. URL: http://dx.doi.org/10.1038/s41598-019-40439-3.

[52] 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". In: *arXiv preprint arXiv:2110.05040* (2021).

[53] Robert M. Parrish et al. "Quantum Computation of Electronic Transitions Using a Variational Quantum Eigensolver". In: *Physical Review Letters* 122.23 (June 2019). ISSN: 1079-7114. DOI: 10.1103/physrevlett.122.230401. URL: http://dx.doi.org/10.1103/physrevlett.122.230401.

[54] Alberto Peruzzo et al. "A variational eigenvalue solver on a photonic quantum processor". In: *Nature Communications* 5.1 (July 2014). ISSN: 2041-1723. DOI: 10.1038/ncomms5213. URL: http://dx.doi.org/10.1038/ncomms5213.

[55] Robert J. Plemmons. "Matrix Analysis (Roger A. Horn and Charles R. Johnson)". In: *SIAM Review* 30.1 (Mar. 1988), pp. 153–154. ISSN: 1095-7200. DOI: 10.1137/1030034. URL: http://dx.doi.org/10.1137/1030034.

[56] David Poulin et al. "Quantum Algorithm for Spectral Measurement with a Lower Gate Count". In: *Physical Review Letters* 121.1 (July 2018). ISSN: 1079-7114. DOI: 10.1103/physrevlett.121.010501. URL: http://dx.doi.org/10.1103/physrevlett.121.010501.

[57] George D. Purvis and Rodney J. Bartlett. "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples". In: *The Journal of Chemical Physics* 76.4 (Feb. 1982), pp. 1910–1918. ISSN: 1089-7690. DOI: 10.1063/1.443164. URL: http://dx.doi.org/10.1063/1.443164.

[58] Google AI Quantum et al. "Hartree-Fock on a superconducting qubit quantum computer". In: *Science* 369.6507 (2020), pp. 1084–1089. DOI: 10.1126/science.abb9811. eprint: https://www.science.org/doi/pdf/10.1126/science.abb9811. URL: https://www.science.org/doi/abs/10.1126/science.abb9811.

[59] Shi-Ju Ran. "Encoding of matrix product states into quantum circuits of one- and two-qubit gates". In: *Physical Review A* 101.3 (Mar. 2020). ISSN: 2469-9934. DOI: 10.1103/physreva.101.032310. URL: http://dx.doi.org/10.1103/physreva.101.032310.

[60] Ben Reggio et al. "Fast partitioning of Pauli strings into commuting families for optimal expectation value measurements of dense operators". In: *Physical Review A* 110.2 (Aug. 2024). ISSN: 2469-9934. DOI: 10.1103/physreva.110.022606. URL: http://dx.doi.org/10.1103/physreva.110.022606.

[61] Dario Rocca et al. "Reducing the Runtime of Fault-Tolerant Quantum Simulations in Chemistry through Symmetry-Compressed Double Factorization". In: *Journal of Chemical Theory and Computation* 20.11 (May 2024), pp. 4639–4653. ISSN: 1549-9626. DOI: 10.1021/acs.jctc.4c00352. URL: http://dx.doi.org/10.1021/acs.jctc.4c00352.

[62] Nicholas C. Rubin et al. "Fault-Tolerant Quantum Simulation of Materials Using Bloch Orbitals". In: *PRX Quantum* 4.4 (Oct. 2023). ISSN: 2691-3399. DOI: 10.1103/prxquantum.4.040303. URL: http://dx.doi.org/10.1103/prxquantum.4.040303.

[63] Yousef Saad. "On the rates of convergence of the Lanczos and the block-Lanczos methods". In: *SIAM Journal on Numerical Analysis* 17.5 (1980), pp. 687–706.

[64] Raffaele Santagati et al. "Efficient calculation of energy derivatives on a Fault-Tolerant Quantum Computer". In: *APS March Meeting Abstracts*. Vol. 2023. 2023, A64–007.

[65] Maximilian Scheurer et al. "Tailored and Externally Corrected Coupled Cluster with Quantum Inputs". In: *Journal of Chemical Theory and Computation* 20.12 (June 2024), pp. 5068–5093. ISSN: 1549-9626. DOI: 10.1021/acs.jctc.4c00037. URL: http://dx.doi.org/10.1021/acs.jctc.4c00037.

[66] Norbert Schuch and Frank Verstraete. "Computational complexity of interacting electrons and fundamental limitations of density functional theory". In: *Nature Physics* 5.10 (Aug. 2009), pp. 732–735. ISSN: 1745-2481. DOI: 10.1038/nphys1370. URL: http://dx.doi.org/10.1038/nphys1370.

[67] Maria Schuld et al. "Evaluating analytic gradients on quantum hardware". In: *Physical Review A* 99.3 (Mar. 2019). ISSN: 2469-9934. DOI: 10.1103/physreva.99.032331. URL: http://dx.doi.org/10.1103/physreva.99.032331.

[68] J. C. Slater. "A Simplification of the Hartree-Fock Method". In: *Phys. Rev.* 81 (3 Feb. 1951), pp. 385–390. DOI: 10.1103/PhysRev.81.385. URL: https://link.aps.org/doi/10.1103/PhysRev.81.385.

[69] Kevin C. Smith et al. "Constant-Depth Preparation of Matrix Product States with Adaptive Quantum Circuits". In: *PRX Quantum* 5.3 (Sept. 2024). ISSN: 2691-3399. DOI: 10.1103/prxquantum.5.030344. URL: http://dx.doi.org/10.1103/prxquantum.5.030344.

[70] Mark Steudtner et al. "Fault-tolerant quantum computation of molecular observables". In: *Quantum* 7 (Nov. 2023), p. 1164. ISSN: 2521-327X. DOI: 10.22331/q-2023-11-06-1164. URL: http://dx.doi.org/10.22331/q-2023-11-06-1164.

[71] Ryan Sweke et al. "Stochastic gradient descent for hybrid quantum-classical optimization". In: *Quantum* 4 (Aug. 2020), p. 314. ISSN: 2521-327X. DOI: 10.22331/q-2020-08-31-314. URL: https://doi.org/10.22331/q-2020-08-31-314.

[72] Tiago M. L. de Veras, Leon D. da Silva, and Adenilton J. da Silva. "Double sparse quantum state preparation". In: *Quantum Information Processing* 21.6 (June 2022). ISSN: 1573-1332. DOI: 10.1007/s11128-022-03549-y. URL: http://dx.doi.org/10.1007/s11128-022-03549-y.

[73] Tiago M. L. de Veras et al. "Circuit-Based Quantum Random Access Memory for Classical Data With Continuous Amplitudes". In: *IEEE Transactions on Computers* 70.12 (Dec. 2021), pp. 2125–2135. ISSN: 2326-3814. DOI: 10.1109/tc.2020.3037932. URL: http://dx.doi.org/10.1109/tc.2020.3037932.

[74] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F. Izmaylov. "Measurement optimization in the variational quantum eigensolver using a minimum clique cover". In: *The Journal of Chemical Physics* 152.12 (Mar. 2020). ISSN: 1089-7690. DOI: 10.1063/1.5141458. URL: http://dx.doi.org/10.1063/1.5141458.

[75] Kianna Wan, Mario Berta, and Earl T. Campbell. "Randomized Quantum Algorithm for Statistical Phase Estimation". In: *Physical Review Letters* 129.3 (July 2022). ISSN: 1079-7114. DOI: 10.1103/physrevlett.129.030503. URL: http://dx.doi.org/10.1103/physrevlett.129.030503.

[76] Steven R. White and Richard L. Martin. "Ab initio quantum chemistry using the density matrix renormalization group". In: *The Journal of Chemical Physics* 110.9 (Mar. 1999), pp. 4127–4130. ISSN: 1089-7690. DOI: 10.1063/1.478295. URL: http://dx.doi.org/10.1063/1.478295.

[77] David Wierichs et al. "General parameter-shift rules for quantum gradients". In: *Quantum* 6 (Mar. 2022), p. 677. ISSN: 2521-327X. DOI: 10.22331/q-2022-03-30-677. URL: http://dx.doi.org/10.22331/q-2022-03-30-677.

[78] Sebastian Wouters and Dimitri Van Neck. "The density matrix renormalization group for ab initio quantum chemistry". In: *The European Physical Journal D* 68.9 (Sept. 2014). ISSN: 1434-6079. DOI: 10.1140/epjd/e2014-50500-1. URL: http://dx.doi.org/10.1140/epjd/e2014-50500-1.

[79] Zongkang Zhang et al. "Measurement-efficient quantum Krylov subspace diagonalisation". In: *Quantum* 8 (Aug. 2024), p. 1438. ISSN: 2521-327X. DOI: 10.22331/q-2024-08-13-1438. URL: http://dx.doi.org/10.22331/q-2024-08-13-1438.

[80] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. "Fermionic Partial Tomography via Classical Shadows". In: *Physical Review Letters* 127.11 (Sept. 2021). ISSN: 1079-7114. DOI: 10.1103/physrevlett.127.110504. URL: http://dx.doi.org/10.1103/physrevlett.127.110504.