Fluid fermionic fragments for optimizing quantum measurements of electronic Hamiltonians in the variational quantum eigensolver

Seonghoon Choi[1,2], Ignacio Loaiza[1,2], and Artur F. Izmaylov[1,2]

[1]Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada
[2]Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada

Measuring the expectation value of the molecular electronic Hamiltonian is one of the challenging parts of the variational quantum eigensolver. A widely used strategy is to express the Hamiltonian as a sum of measurable fragments using fermionic operator algebra. Such fragments have an advantage of conserving molecular symmetries that can be used for error mitigation. The number of measurements required to obtain the Hamiltonian expectation value is proportional to a sum of fragment variances. Here, we introduce a new method for lowering the fragments' variances by exploiting flexibility in the fragments' form. Due to idempotency of the occupation number operators, some parts of two-electron fragments can be turned into one-electron fragments, which then can be partially collected in a purely one-electron fragment. This repartitioning does not affect the expectation value of the Hamiltonian but has non-vanishing contributions to the variance of each fragment. The proposed method finds the optimal repartitioning by employing variances estimated using a classically efficient proxy for the quantum wavefunction. Numerical tests on several molecules show that repartitioning of one-electron terms lowers the number of measurements by more than an order of magnitude.

1 Introduction

The variational quantum eigensolver (VQE) [1–5] is a promising hybrid quantum-classical algorithm for finding the ground-state of the molecular electronic Hamiltonian,

$$\hat{H} = \sum_{pq}^{N} h_{pq}\hat{E}_p^q + \sum_{pqrs}^{N} g_{pqrs}\hat{E}_p^q\hat{E}_r^s, \qquad (1)$$

presented here in so-called chemists' notation, i.e., in terms of one-electron excitation operators, $\hat{E}_p^q = \hat{a}_p^\dagger\hat{a}_q$; while h_{pq} and g_{pqrs} are related to one- and two-electronic integrals, N is the number of spin-orbitals.

VQE circumvents the hardware limitations of today's noisy intermediate-scale quantum (NISQ) devices [6] by exploiting both quantum and classical computers: A quantum computer prepares a parameterized trial state $|\psi_\theta\rangle$ and measures its energy $E_\theta = \langle\psi_\theta|\hat{H}|\psi_\theta\rangle$, while a classical computer suggests new θ parameters to minimize E_θ (for brevity, we omit θ from now). However, measuring the expectation value E is not trivial because digital quantum computers can only measure polynomial functions of Pauli-$\hat{z}$ operators.

A class of widely used approaches based on fermionic operator algebra [7–11] measures E by re-expressing the molecular electronic Hamiltonian in Eq. (1) as

$$\begin{aligned}\hat{H} &= \hat{H}_0 + \sum_{\alpha=1}^{N_f}\hat{H}_\alpha \\ &= \sum_{pq}^{N} h_{pq}\hat{E}_p^q + \sum_{\alpha=1}^{N_f}\sum_{pqrs} g_{pqrs}^{(\alpha)}\hat{E}_p^q\hat{E}_r^s \\ &= \hat{U}_0^\dagger\left(\sum_{p}^{N}\lambda_p\hat{n}_p\right)\hat{U}_0 \\ &\quad + \sum_{\alpha=1}^{N_f}\hat{U}_\alpha^\dagger\left(\sum_{pq}^{N}\lambda_{pq}^{(\alpha)}\hat{n}_p\hat{n}_q\right)\hat{U}_\alpha, \qquad (2)\end{aligned}$$

where $\hat{n}_p = \hat{a}_p^\dagger\hat{a}_p$ are occupation number operators, and $\hat{U}_\alpha = \exp[\sum_{p>q}^{N}\theta_{pq}^{(\alpha)}(\hat{E}_p^q - \hat{E}_q^p)]$ are or-

Artur F. Izmaylov: artur.izmaylov@utoronto.ca

bital rotations. Since $\hat{n}_p$ is mapped to polynomial functions of Pauli-$\hat{z}$ under all standard qubit-fermion transformations [12, 13], each $\hat{U}_\alpha \hat{H}_\alpha \hat{U}_\alpha^\dagger$ is measurable on a quantum computer. Therefore, E can be obtained by measuring each $\hat{H}_\alpha$ term separately using

$$E = \sum_{\alpha=0}^{N_f} \langle\psi|\hat{H}_\alpha|\psi\rangle = \sum_{\alpha=0}^{N_f} \langle\hat{U}_\alpha\psi|\hat{U}_\alpha\hat{H}_\alpha\hat{U}_\alpha^\dagger|\hat{U}_\alpha\psi\rangle. \tag{3}$$

Equation (3) shows that to measure each $\hat{H}_\alpha$, one first has to apply $\hat{U}_\alpha$ on $|\psi\rangle$. Fortunately, implementing $\hat{U}_\alpha$ on a quantum computer is efficient as it only requires $\binom{N}{2}$ two-qubit rotations and a gate depth of N [14].

To find the fragments, $\hat{H}_\alpha$, starting from the Hamiltonian in the second-quantized form, one can employ either the low-rank (LR) decomposition [7–9] or the full-rank (FR) optimization [10, 11]. Both methods are only concerned with the two-electron fragments because $\hat{U}_0$ for the one-electron fragment ($\hat{H}_0$) can be found easily as it simply corresponds to a unitary matrix that diagonalizes the one-electron tensor, h_{pq} [14]. The difference between LR and FR fragments is the rank of the resulting $\lambda_{pq}^{(\alpha)}$ [in Eq. (2)]. In the LR decomposition, $\lambda_{pq}^{(\alpha)}$ is an outer product of some vector $\eta_p^{(\alpha)}$ (i.e., $\lambda_{pq}^{(\alpha)} = \eta_p^{(\alpha)}\eta_q^{(\alpha)}$) and, therefore, has rank 1. In contrast, in FR optimization, $\lambda_{pq}^{(\alpha)}$ can be a full-rank hermitian matrix. Increased flexibility of FR fragments was exploited in Ref. 10 to lower the number of measurements required to obtain E up to error ϵ when each $\hat{H}_\alpha$ is measured independently [15]:

$$M(\epsilon) = \frac{1}{\epsilon^2} \sum_{\alpha=0}^{N_f} \frac{\mathrm{Var}_\psi(\hat{H}_\alpha)}{m_\alpha}, \tag{4}$$

where $\mathrm{Var}_\psi(\hat{H}_\alpha) = \langle\psi|\hat{H}_\alpha^2|\psi\rangle - \langle\psi|\hat{H}_\alpha|\psi\rangle^2$ is the variance of $\hat{H}_\alpha$, and m_α is the fraction of the total measurements allocated to $\hat{H}_\alpha$. Developing techniques for measuring E with a low $M(\epsilon)$ is especially important for VQE because a recent analysis [16] showed that the advantage of VQE over state-of-the-art classical algorithms is limited due to the large $M(\epsilon)$.

Yet, even the best implementation of the FR optimization was shown to have higher $M(\epsilon)$'s than those in the measurement approaches developed in the qubit space [15, 17–26]. The qubit-space techniques start by applying one of the fermion-qubit mappings [12, 13] to the fermionic Hamiltonian to produce the qubit Hamiltonian,

$$\hat{H}_q = \sum_{j=1} c_j \hat{P}_j, \tag{5}$$

where each $\hat{P}_j$ is an N-qubit Pauli product (i.e., a tensor product of Pauli operators for individual qubits). Thus obtained $\hat{H}_q$ is subsequently partitioned into a linear combination of N_f independently measured fragments, $\hat{H}_\beta$ (i.e., $\hat{H}_q = \sum_{\beta=1}^{N_f} \hat{H}_\beta$), where

$$\hat{H}_\beta = \hat{V}_\beta^\dagger \left(\sum_i a_i^{(\beta)} \hat{z}_i + \sum_{ij} a_{ij}^{(\beta)} \hat{z}_i\hat{z}_j + \dots \right) \hat{V}_\beta. \tag{6}$$

Every $\hat{H}_\beta$ contains only mutually commutative Pauli products and thus allows one to efficiently implement the corresponding $\hat{V}_\beta$ using only one- and two-qubit Clifford gates [15, 24, 27, 28].

The qubit-space methods with the lowest $M(\epsilon)$ take advantage of the flexibility in the fragments offered by the realization that some $\hat{P}_j$ can belong to multiple $\hat{H}_\beta$. The coefficients of $\hat{P}_j$ in different $\hat{H}_\beta$, $c_j^{(\beta)}$, can be varied without changing the total expectation value of $\hat{H}_q$ as long as $c_j^{(\beta)}$ sum to c_j in the qubit Hamiltonian [25]. In addition, even $\hat{P}_j$ not present in $\hat{H}_q$ can be introduced into multiple $\hat{H}_\beta$ provided that corresponding $c_j^{(\beta)}$ sum to zero [26]. A significant reduction in $M(\epsilon)$ was achieved by optimizing $c_j^{(\beta)}$ using approximate variances obtained by employing a classically efficient wavefunction, $|\phi\rangle$, to estimate $\mathrm{Var}_\phi(\hat{H}_\beta)$. The idea of increasing the number of $\hat{P}_j$ measured simultaneously in a single fragment has been successfully employed also in the recently developed classical-shadow-based techniques [18–23] to yield lower $M(\epsilon)$ values. An alternative class of promising qubit-space approaches with $M(\epsilon)$'s competitive with those in some of the Hamiltonian partitioning schemes lowers $M(\epsilon)$ by optimizing positive operator-valued measures (POVMs) [29–31].

In this work, we present an extension to fermionic fragment techniques that further increases their flexibility in reducing the number of measurements. The new approach generalizes the technique of repartitioning of some fragments used in the qubit space. It extends the repartitioning idea from commuting to non-commuting operators. Another motivation for developing

fermionic measurement schemes is their advantage compared to the qubit-space counterparts in conserving molecular symmetries (e.g., electronic number and spin operators). These symmetries can be used for error mitigation techniques, which are essential for advancing quantum computing schemes on near-term devices [32, 33].

2 Theory

2.1 Fluid fermionic fragments

Here, we present a new approach that exploits two properties of fermionic operators to minimize the number of measurements in Eq. (4). First, any linear combination of one-electron hermitian operators can be brought to the factorized form

$$\sum_\alpha c_\alpha \hat{U}_\alpha^\dagger \left(\sum_p \epsilon_p^{(\alpha)} \hat{n}_p \right) \hat{U}_\alpha = \hat{U}^\dagger \left(\sum_p \epsilon_p \hat{n}_p \right) \hat{U}, \tag{7}$$

where $c_\alpha, \epsilon_p^{(\alpha)}$, and ϵ_p are some real coefficients, and $\hat{U}_\alpha, \hat{U}$ are orbital rotations. Second, the occupation number operators are idempotent, i.e., $\hat{n}_p^2 = \hat{n}_p$.

Using the $\hat{n}_p$ idempotency, each $\hat{H}_\alpha$ in Eq. (2) with $\alpha > 0$ can be re-written as a sum of one- and two-electron parts:

$$\hat{H}_\alpha = \hat{U}_\alpha^\dagger \left(\sum_p^N \lambda_{pp}^{(\alpha)} \hat{n}_p + \sum_{p \neq q} \lambda_{pq}^{(\alpha)} \hat{n}_p \hat{n}_q \right) \hat{U}_\alpha. \tag{8}$$

This expression reveals the freedom that one can extract any amount of the one-electron part from every $\hat{H}_\alpha$ and add it to $\hat{H}_0$, thereby repartitioning the one- and two-electron Hamiltonians. Thus, the repartitioned fragments, which we will refer to as fluid fermionic fragments (F^3), are

$$\hat{H}_0' = \hat{H}_0 + \sum_{\alpha=1}^{N_f} \hat{U}_\alpha^\dagger \left(\sum_p^N c_p^{(\alpha)} \hat{n}_p \right) \hat{U}_\alpha, \tag{9}$$

$$\begin{aligned} \hat{H}_\alpha' &= \hat{H}_\alpha - \hat{U}_\alpha^\dagger \left(\sum_p^N c_p^{(\alpha)} \hat{n}_p \right) \hat{U}_\alpha \\ &= \hat{U}_\alpha^\dagger \left[\sum_p^N (\lambda_{pp}^{(\alpha)} - c_p^{(\alpha)}) \hat{n}_p + \sum_{p \neq q} \lambda_{pq}^{(\alpha)} \hat{n}_p \hat{n}_q \right] \hat{U}_\alpha. \end{aligned} \tag{10}$$

Even after the modification, each $\hat{H}_\alpha$ for $\alpha > 0$ remains measurable because $\hat{U}_\alpha \hat{H}_\alpha \hat{U}_\alpha^\dagger$ still maps onto a polynomial function of Pauli-$\hat{z}$ after qubit-fermion transformations. For $\hat{H}_0'$, new $\hat{U}_0'$ and λ_p' can easily be found by simply diagonalizing $h_{pq}' = h_{pq} + \sum_{\alpha=1}^{N_f} \sum_r^N (U_{rp}^{(\alpha)})^* c_r^{(\alpha)} U_{rq}^{(\alpha)}$, where $U_{pq}^{(\alpha)}$ is an $N \times N$ matrix representation of $\hat{U}_\alpha$ [10, 11, 14]. Measuring $\hat{H}_\alpha'$ instead of $\hat{H}_\alpha$ gives the same E because repartitioning does not change the operator sum: $\sum_{\alpha=0}^{N_f} \hat{H}_\alpha' = \sum_{\alpha=0}^{N_f} \hat{H}_\alpha = \hat{H}$. In contrast, $M(\epsilon)$ changes with the choice of $c_p^{(\alpha)}$ because $\mathrm{Var}_\psi(\hat{H}_\alpha')$ has a non-linear dependence on $c_p^{(\alpha)}$. As a consequence, one can reduce $M(\epsilon)$ by optimizing $c_p^{(\alpha)}$. Linearity of fermionic fragments with respect to $c_p^{(\alpha)}$ makes variance optimization particularly efficient.

2.2 Optimization of the number of measurements

In the following, we will present the approach for optimally repartitioning the one- and two-electron fragments to lower $M(\epsilon)$ (initial fragments are obtained as described in Appendix A). Since $M(\epsilon)$ depends on the fragment variances evaluated with the quantum wavefunction $|\psi\rangle$, which is classically difficult, we minimize the approximation to $M(\epsilon)$ computed with $\mathrm{Var}_\phi(\hat{H}_\alpha')$:

$$M_\phi(\epsilon) = \frac{1}{\epsilon^2} \sum_{\alpha=0}^{N_f} \frac{\mathrm{Var}_\phi(\hat{H}_\alpha')}{m_\alpha}; \tag{11}$$

in this work, the configuration interaction singles and doubles (CISD) wavefunction was used as the classically efficient proxy for the quantum wavefunction ($|\phi\rangle$). The variances of the fragments after the repartition [$\hat{H}_\alpha'$ in Eqs. (9) and (10)] are obtained as

$$\begin{aligned} \mathrm{Var}_\phi(\hat{H}_0') =& \mathrm{Var}_\phi(\hat{H}_0) \\ &+ \sum_{\alpha,\beta}^{N_f} \sum_{p,q}^N c_p^{(\alpha)} c_q^{(\beta)} \mathrm{Cov}_\phi(\hat{O}_p^{(\alpha)}, \hat{O}_q^{(\beta)}) \\ &+ \sum_\alpha^{N_f} \sum_p^N c_p^{(\alpha)} \overline{\mathrm{Cov}}_\phi(\hat{H}_0, \hat{O}_p^{(\alpha)}), \end{aligned} \tag{12}$$

$$\begin{aligned} \mathrm{Var}_\phi(\hat{H}_\alpha') =& \mathrm{Var}_\phi(\hat{H}_\alpha) \\ &+ \sum_{p,q}^N c_p^{(\alpha)} c_q^{(\alpha)} \mathrm{Cov}_\phi(\hat{O}_p^{(\alpha)}, \hat{O}_q^{(\alpha)}) \\ &- \sum_p^N c_p^{(\alpha)} \overline{\mathrm{Cov}}_\phi(\hat{H}_\alpha, \hat{O}_p^{(\alpha)}), \end{aligned} \tag{13}$$

where we introduced $\overline{\mathrm{Cov}}_\phi(\hat{A},\hat{B}) = \mathrm{Cov}_\phi(\hat{A},\hat{B}) + \mathrm{Cov}_\phi(\hat{B},\hat{A})$ and $\hat{O}_p^{(\alpha)} = \hat{U}_\alpha^\dagger \hat{n}_p \hat{U}_\alpha$ for notational simplicity. To minimize Eq. (11) with respect to $c_p^{(\alpha)}$ and m_α, we perform two-step iterative optimization following Refs. 25 and 26: 1) $c_p^{(\alpha)}$ are optimized with fixed m_α and 2) m_α are updated to the optimal allocation according to Eq. (22) with fixed $c_p^{(\alpha)}$. As this iterative procedure is a linearization heuristic for the non-linear problem of simultaneously optimizing $c_p^{(\alpha)}$ and m_α, the obtained solution does not necessarily correspond to the minimum. Yet, in practice, we could typically reach convergence within ~ 5 iterations to find solutions with low gradients. To optimize the $c_p^{(\alpha)}$ variables with fixed m_α, one can solve the system of linear equations:

$$\epsilon^2 \frac{\partial M_\phi(\epsilon)}{\partial c_p^{(\alpha)}} = \frac{1}{m_0}\Big[\sum_\beta^{N_f}\sum_q^{N} c_q^{(\beta)} \overline{\mathrm{Cov}}_\phi(\hat{O}_p^{(\alpha)}, \hat{O}_q^{(\beta)}) + \overline{\mathrm{Cov}}_\phi(\hat{H}_0, \hat{O}_p^{(\alpha)})\Big] + \frac{1}{m_\alpha}\Big[\sum_q^{N} c_q^{(\alpha)} \overline{\mathrm{Cov}}_\phi(\hat{O}_p^{(\alpha)}, \hat{O}_q^{(\alpha)}) - \overline{\mathrm{Cov}}_\phi(\hat{H}_\alpha, \hat{O}_p^{(\alpha)})\Big] = 0. \quad (14)$$

The final m_α obtained at the end of the iterative procedure suggests the optimal allocation of the total budget for each $\hat{H}_\alpha$. The suggested Mm_α measurements for each fragment $\hat{H}_\alpha$ is not an integer, but rounding Mm_α to the nearest integer should only have a negligible effect on the measurement error because $M \gtrsim 10^6$ in practice. We will refer to the algorithm proposed in this work as the fluid fermionic fragments (F^3) method.

2.3 Reducing the number of optimization variables

The computational cost for optimizing $c_q^{(\alpha)}$ in the F^3 method increases with the number of $c_q^{(\alpha)}$ (N_c). The two main contributors to the computational time are the evaluation of covariances, $\overline{\mathrm{Cov}}_\phi(\hat{O}_p^{(\alpha)}, \hat{O}_q^{(\beta)})$, and solving the system of linear Eqs. (14). Because the evaluation of $\overline{\mathrm{Cov}}_\phi(\hat{O}_p^{(\alpha)}, \hat{O}_q^{(\beta)})$ scales quadratically with N_c and the computational time required for solving Eq. (14) has an approximately cubic scaling with N_c, the cost of F^3 can be lowered significantly by reducing N_c. Therefore, we propose several restrictions on $c_q^{(\alpha)}$ to lower their number.

Using spin symmetry of the electronic Hamiltonian written in a spin-restricted spin-orbital basis, we achieve a twofold reduction in the number of $c_q^{(\alpha)}$. Note that $\lambda^{(\alpha)}_{2i-1,2i-1} = \lambda^{(\alpha)}_{2i,2i}$ for $i = 1,\ldots,N/2$ in the initial $\hat{H}_\alpha$ fragments obtained by considering the smaller $\tilde{g}_{ijkl}$ tensor over spatial orbitals (see Appendix A for definitions). Because $\lambda^{(\alpha)}_{2i-1,2i-1} = \lambda^{(\alpha)}_{2i,2i}$, we impose that the same amount of $\lambda^{(\alpha)}_{2i-1,2i-1}$ and $\lambda^{(\alpha)}_{2i,2i}$ is extracted from $\hat{H}_\alpha$, i.e., we impose that $c^{(\alpha)}_{2i-1} = c^{(\alpha)}_{2i} = \tilde{c}^{(\alpha)}_i$, thereby reducing the number of optimization variables by half. In the resulting F^3-Full method, $N_c = N_f N/2$, and the system of equations is simplified to

$$\epsilon^2 \frac{\partial M_\phi(\epsilon)}{\partial \tilde{c}_i^{(\alpha)}} = \frac{1}{m_0}\Big[\sum_\beta^{N_f}\sum_j^{N/2} \tilde{c}_j^{(\beta)} \overline{\mathrm{Cov}}_\phi(\hat{P}_i^{(\alpha)}, \hat{P}_j^{(\beta)}) + \overline{\mathrm{Cov}}_\phi(\hat{H}_0, \hat{P}_i^{(\alpha)})\Big] + \frac{1}{m_\alpha}\Big[\sum_j^{N/2} \tilde{c}_j^{(\alpha)} \overline{\mathrm{Cov}}_\phi(\hat{P}_i^{(\alpha)}, \hat{P}_j^{(\alpha)}) - \overline{\mathrm{Cov}}_\phi(\hat{H}_\alpha, \hat{P}_i^{(\alpha)})\Big] = 0, \quad (15)$$

where $\hat{P}_i^{(\alpha)} = \hat{O}_{2i-1}^{(\alpha)} + \hat{O}_{2i}^{(\alpha)} = \sum_\sigma \hat{U}_\alpha^\dagger \hat{n}_{i\sigma} \hat{U}_\alpha$.

Even more drastic reduction in N_c can be achieved if we restrict $c_p^{(\alpha)}$ to be p-independent. This restricts us to repartitioning of only a fraction of the entire one-electron part of a two-electron $\hat{H}_\alpha$ fragment [i.e., a fraction of $\hat{U}_\alpha^\dagger(\sum_p \lambda_{pp}^{(\alpha)} \hat{n}_p)\hat{U}_\alpha$]. To this end, we restrict $c_p^{(\alpha)}$ as a scalar multiple of $\lambda_{pp}^{(\alpha)}$: $c_p^{(\alpha)} = c^{(\alpha)} \lambda_{pp}^{(\alpha)}$. As a result, $N_c = N_f$, and the system (14) simplifies down to

$$\epsilon^2 \frac{\partial M_\phi(\epsilon)}{\partial c^{(\alpha)}} = \frac{1}{m_0}\Big[\sum_\beta^{N_f} c^{(\beta)} \overline{\mathrm{Cov}}_\phi(\hat{O}^{(\alpha)}, \hat{O}^{(\beta)}) + \overline{\mathrm{Cov}}_\phi(\hat{H}_0, \hat{O}^{(\alpha)})\Big] + \frac{1}{m_\alpha}\Big[2c^{(\alpha)} \mathrm{Var}_\phi(\hat{O}^{(\alpha)}) - \overline{\mathrm{Cov}}_\phi(\hat{H}_\alpha, \hat{O}^{(\alpha)})\Big] = 0, \quad (16)$$

where $\hat{O}^{(\alpha)} = \sum_p^N \lambda_{pp}^{(\alpha)} \hat{O}_p^{(\alpha)}$. We will refer to this reduced version of F^3 as F^3-R1.

Yet another reduction of variables can be done and is motivated by the relationship between $[\mathrm{Var}_\psi(\hat{H}_\alpha)]^{1/2}$ appearing in the expression for the total measurement number with optimal m_α [$M_{\mathrm{opt}}(\epsilon)$ in Appendix A] and the L_1 norm of a

coefficient vector for a linear combination of unitaries (LCU) decomposition: $\hat{H}_\alpha = \sum_j a_j^{(\alpha)} \hat{V}_j + d_\alpha \hat{1}$, where $\hat{V}_j$ are some unitaries [34]. Maximum $[\mathrm{Var}_\psi(\hat{H}_\alpha)]^{1/2}$ for any $|\psi\rangle$ occurs when $|\psi\rangle = (|\mathrm{max}\rangle_\alpha + |\mathrm{min}\rangle_\alpha)/\sqrt{2}$, where $|\mathrm{max}\rangle_\alpha$ ($|\mathrm{min}\rangle_\alpha$) is the eigenstate of $\hat{H}_\alpha$ with the highest (lowest) eigenvalue, $E_{\mathrm{max}}^{(\alpha)}$ ($E_{\mathrm{min}}^{(\alpha)}$); the corresponding maximum is

$$\max_\psi \sqrt{\mathrm{Var}_\psi(\hat{H}_\alpha)} = \Delta E_\alpha/2 \equiv (E_{\mathrm{max}}^{(\alpha)} - E_{\mathrm{min}}^{(\alpha)})/2. \tag{17}$$

Using Theorem 1 of Ref. 34, which shows that the LCU L_1 norm, $\sum_j |a_j^{(\alpha)}|$, is in turn an upper bound for $\Delta E_\alpha/2$, we find that

$$\sqrt{\mathrm{Var}_\psi(\hat{H}_\alpha)} \leq \sum_j |a_j^{(\alpha)}|. \tag{18}$$

Thus, one can use $\hat{H}_\alpha$ with a low LCU L_1 norm as a heuristic approach to lowering $M(\epsilon)$.

One way to reduce the L_1 norm for a collection of $\hat{H}_\alpha$ while maintaining their measurability is substituting every $\hat{n}_p$ operator in the two-electron $\hat{H}_\alpha$ fragments with reflections: $\hat{r}_p = 1 - 2\hat{n}_p$ (satisfying $\hat{r}_p^2 = 1$ and $\hat{r}_p^\dagger = \hat{r}_p$) [7, 34–36]. Because $\hat{r}_p$ maps onto an all-$\hat{z}$ Pauli product under all standard transformations (e.g., it maps to $\hat{z}_p$ under the Jordan–Wigner transformation) [13], the fragment remains measurable even after the substitution:

$$\hat{H}'_\alpha = \hat{U}_\alpha^\dagger \left(\sum_{pq}^{N} \frac{\lambda_{pq}^{(\alpha)}}{4} \hat{r}_p \hat{r}_q \right) \hat{U}_\alpha. \tag{19}$$

To ensure that $\sum_{\alpha=0}^{N_f} \hat{H}'_\alpha = \hat{H}$, the one-electron $\hat{H}_0$ term must also be modified as

$$\hat{H}'_0 = \hat{H}_0 + \sum_{\alpha=1}^{N_f} \sum_{pq}^{N} \left(\lambda_{pq}^{(\alpha)} \hat{O}_p^{(\alpha)} - \frac{\lambda_{pq}^{(\alpha)}}{4} \right). \tag{20}$$

Concerning measurements, the constant term in Eq. (20) affects neither $M(\epsilon)$ nor the measurability of the fragments. Moving the constant terms, $-\sum_{pq} \lambda_{pq}^{(\alpha)}/4$, from $\hat{H}'_0$ back to each $\hat{H}'_\alpha$ reveals the connection to F^3: the substitution of $\hat{n}_p$ with $\hat{r}_p$ simply corresponds to the F^3 approach with $c_p^{(\alpha)}$ fixed as $c_p^{(\alpha)} = \sum_q \lambda_{pq}^{(\alpha)}$.

Inspired by this connection to F^3 and the reduction in the upper bound for the fragment variances achieved by the substitution of $\hat{r}_p$, we heuristically propose F^3-R2 that restricts the optimization variables as $c_p^{(\alpha)} = c^{(\alpha)} \sum_q \lambda_{pq}^{(\alpha)}$. Note that this choice is equivalent to substituting $\hat{r}_p$ only for a fraction of the two-electron fragment ($c^{(\alpha)} \hat{H}_\alpha$) while leaving the rest [$(1 - c^{(\alpha)}) \hat{H}_\alpha$] as functions of $\hat{n}_p$. The $c^{(\alpha)}$ variables in F^3-R2 are optimized by solving the system of equations identical to Eq. (16) except that $\hat{O}^{(\alpha)} = \sum_{pq}^{N} \lambda_{pq}^{(\alpha)} \hat{O}_p^{(\alpha)}$ (instead of $\hat{O}^{(\alpha)} = \sum_p^N \lambda_{pp}^{(\alpha)} \hat{O}_p^{(\alpha)}$ in F^3-R1).

3 Results and discussions

We compare $M(\epsilon)$ in the different versions of F^3 applied to either LR or FR fragments with $M(\epsilon)$ in the initial LR and GFRO fragments (see Appendix A for definitions). In addition, the performance of F^3 is compared with the best qubit-space techniques that have the lowest $M(\epsilon)$: the iterative coefficient splitting (ICS) [25] and shared Pauli products (SPP) [26] methods. In particular, it was shown in Ref. 25 that for measuring the expectation values of molecular electronic Hamiltonians, the $M(\epsilon)$ values in ICS and SPP are severalfold lower than $M(\epsilon)$ in some recently developed classical-shadow-based techniques [21, 22].

The algorithms were used to compute $\epsilon^2 M(\epsilon)$ for electronic Hamiltonians of several molecules in the STO-3G basis and the following nuclear geometries: $R(\mathrm{H}-\mathrm{H}) = 1$Å with $\angle\mathrm{HHH} = 180°$ (for H_3^+, H_4, and H_6), $R(\mathrm{H}-\mathrm{X}) = 1$Å (for X=F, Li), $R(\mathrm{C}-\mathrm{H}) = 1$Å with $\angle\mathrm{HCH} = 101.9°$ (for CH_2), $R(\mathrm{Be}-\mathrm{H}) = 1$Å with $\angle\mathrm{HBeH} = 180°$ (for BeH_2), $R(\mathrm{O}-\mathrm{H}) = 1$Å with $\angle\mathrm{HOH} = 107.6°$ (for H_2O), and $R(\mathrm{N}-\mathrm{H}) = 1$Å with $\angle\mathrm{HNH} = 107°$ (for NH_3). The data for the electronic Hamiltonians [Eq. (1)] and the corresponding initial LR and GFRO fragments [Eq. (2)] can be found in Ref. 37.

The presented $\epsilon^2 M(\epsilon)$ value is equivalent to the number of measurements in millions required to obtain E with 10^{-3} a.u. accuracy (see Appendix B). While the optimization in F^3 is performed using the covariances approximated with the CISD wavefunction, $|\phi\rangle$, once the optimal fragments and measurement allocation are obtained, the $\epsilon^2 M(\epsilon)$ values are then evaluated according to Eq. (4) using the exact covariances computed with the quantum wavefunction, $|\psi\rangle$. [Note that the use of approximate covariances instead of the exact covariances in F^3 had an insignificant influence on $\epsilon^2 M(\epsilon)$. In all systems,

Table 1: Required measurement numbers [$\epsilon^2 M(\epsilon)$] in the different versions of F^3 applied to either LR or FR fragments are compared with $\epsilon^2 M(\epsilon)$ in LR, GFRO, ICS [25], and SPP [26] for Hamiltonians of several molecules (N is the number of spin-orbitals and is equal to the number of qubits).

						F^3-LR			F^3-FR		
Sys	N	LR	GFRO	ICS	SPP	Full	R1	R2	Full	R1	R2
H_3^+	6	0.458	0.261	0.179	0.142	0.148	0.156	0.162	0.0803	0.0815	0.0805
H_4	8	1.50	0.643	0.754	0.446	0.538	0.578	0.554	0.316	0.318	0.317
H_6	12	3.06	1.21	2.27	1.23	1.08	1.42	1.13	0.519	0.587	0.554
HF	12	43.5	41.9	0.670	0.436	0.278	7.10	0.454	0.327	15.5	0.593
LiH	12	3.16	2.73	0.295	0.148	0.127	0.338	0.196	0.122	0.197	0.165
CH_2	14	22.0	15.6	2.75	0.796	0.985	4.80	1.23	0.689	12.2	0.830
BeH_2	14	1.86	1.61	0.543	0.341	0.543	0.848	0.680	0.430	0.948	0.583
H_2O	14	58.5	49.4	2.05	1.16	0.892	9.83	1.10	0.709	25.7	0.911
NH_3	16	58.1	46.1	4.83	2.62	1.49	9.22	1.70	0.990	31.3	1.18

Table 2: Number of optimization variables (N_c) in the different versions of F^3 applied to either LR or FR fragments is compared with N_c in ICS [25] and SPP [26] for the systems presented in Table 1.

				F^3-LR		F^3-FR	
Sys	N	ICS	SPP	Full	R1/R2	Full	R1/R2
H_3^+	6	93	97	18	6	24	8
H_4	8	236	644	40	10	92	23
H_6	12	1702	8002	108	18	354	59
HF	12	1220	3409	126	21	384	64
LiH	12	1346	2718	126	21	354	59
CH_2	14	1765	7534	196	28	539	77
BeH_2	14	1265	3048	196	28	602	86
H_2O	14	1863	7987	196	28	938	134
NH_3	16	7203	35875	288	36	904	113

$\epsilon^2 M(\epsilon)$ in F^3 optimized using the exact covariances were lower only by 9% or less compared to the presented $\epsilon^2 M(\epsilon)$ values in F^3 optimized with approximate covariances.]

Table 1 shows that between the different versions of F^3, the most flexible full version yields the lowest $M(\epsilon)$ in all examples: on average, $M(\epsilon)$ in F^3-Full is a factor of 11 lower than that in F^3-R1 and a factor of 1.2 lower than that in F^3-R2. However, as shown in Table 2, the increased flexibility also results in F^3-Full having $N/2$-fold more optimization variables than the other versions. While F^3-R1 and F^3-R2 have the same N_c, F^3-R2 has a much lower $M(\epsilon)$ for many of the molecules; for the others, it has a similar $M(\epsilon)$ to that in F^3-R1. The success of F^3-R2 can be heuristically justified since it is designed to lower the LCU L_1 norm, which is an upper bound for fragment variances, as discussed in Sec. 2.3. Because F^3-R2 can achieve a lower $M(\epsilon)$ with the identical computational cost as F^3-R1, there is no reason to employ F^3-R1 instead of F^3-R2. Therefore, we omit F^3-R1 in the following discussion.

The comparison of $M(\epsilon)$ in F^3 with that in the initial set of LR or FR fragments demonstrates the success of the proposed method. On average, $M(\epsilon)$ in F^3-Full is a factor of 35 lower than that in the initial fragments, and $M(\epsilon)$ in F^3-R2 is a factor of 24 lower than that in the initial fragments. Since $M(\epsilon)$ in GFRO is always lower than that in LR, $M(\epsilon)$ in F^3 is also typically lower when FR fragments are used as the initial fragments. However, finding the initial GFRO fragments involves an iterative nonlinear optimization procedure which is computationally more expensive than the LR decomposition. Moreover, while N_f in LR is upper bounded by $(N^2/8 + N/4)$, N_f in GFRO is usually higher (for the presented molecules, GFRO has, on average, three times more fragments). Since the number of optimization variables is directly proportional to N_f ($N_c = N_f N/2$ in F^3-Full and $N_c = N_f$ in F^3-R2), employing F^3 on the FR fragments requires more computational effort than employing it on the LR fragments.

For systems in Table 2, N_c scales as N^x, where $x = 2.8$ (for F^3-LR-Full), $x = 1.8$ (for F^3-LR-R2), $x = 3.7$ (for F^3-FR-Full), and $x = 2.7$ (for F^3-FR-R2). Because the bottleneck in the F^3 method is solving the system of linear equations [i.e., solving Eq. (15) or (16)], which has $\sim N_c^3$ scaling, the classical optimization costs in the different versions of F^3 have approximately $N^{8.5}, N^{5.5}, N^{11}$, and $N^{8.2}$ scaling. Therefore, F^3-LR-R2, with its classical computational cost scaling as $\sim N^{5.5}$

with system size (N), would be particularly useful for larger systems.

Comparison of F^3 with state-of-the-art qubit-space techniques (ICS and SPP) shows that for all molecules except BeH_2, F^3 has lower $M(\epsilon)$. In particular, even the computationally most efficient combination of applying F^3-R2 to the LR fragments yields a lower $M(\epsilon)$ than that in ICS, whereas this most efficient combination yields a similar $M(\epsilon)$ to that in SPP. Furthermore, $M(\epsilon)$ in the best fermionic-space technique (F^3-FR-Full) is, on average, a factor of 1.6 smaller than $M(\epsilon)$ in the best qubit-space technique (SPP). The bottleneck in the optimization procedures in all three methods (F^3, ICS, and SPP) is solving a system of linear equations. Therefore, one can assess the required computational cost by examining the number of optimization variables. Table 2 shows that even the most computationally costly combination, F^3-FR-Full has a much lower N_c than that in either ICS or SPP. The classical computational costs of ICS and SPP have approximately N^{12} and N^{15} scaling, which are worse even than that in F^3-FR-Full, with the worst scaling among different versions of F^3.

4 Conclusion

This work proposes a new method that achieves a significant reduction in the number of measurements by taking advantage of the possibility of extracting fractions of one-electron parts from the two-electron fermionic fragments and combining them with the purely one-electron fragment. This repartitioning keeps the fragments measurable and conserves the Hamiltonian expectation value. On the other hand, the number of measurements due to its dependence on variances of fragments can be lowered by this repartitioning. The proposed algorithm finds the repartitioning that minimizes the number of measurements and achieves a severalfold reduction in the number of measurements compared to those for initially generated fragments.

Even though the number of measurements in the previously proposed methods suggested that the qubit-space techniques are superior to their fermionic counterparts, by employing the fluid fermionic fragments method, we were able to achieve the number of measurements lower even than those in the best qubit-space techniques (ICS [25] and SPP [26]). Furthermore, compared to these techniques, the method presented here was shown to have much fewer optimization variables. In particular, the number of optimization variables in the computationally most efficient version of the proposed algorithm scales sub-quadratically with the number of spatial orbitals, thereby making the algorithm applicable for larger systems. In addition, while the errors due to the non-unit quantum gate fidelities were not considered in this work, fermionic fragments conserve molecular symmetries and their measurements can be corrected by error mitigation methods employing these symmetries [38–40]. Therefore, the advantage of the proposed fluid fermionic fragments method over the qubit-space techniques should be even greater when quantum hardware errors are taken into account.

While this work focused on the measurements in VQE for ground-state energy estimation, some promising extensions of VQE, including those for excited-state calculations [41–43], require measuring expectation values of additional effective Hamiltonians containing three- and four-electron terms. A common approach for lowering the number of measurements required to obtain these additional expectation values relies on improving the efficiency of simultaneously estimating all three- and four-body reduced density matrices [44, 45]. Alternatively, as a potential extension of this work, one could develop a more targeted algorithm that finds optimal fragments for only a few necessary effective Hamiltonians. It would be informative to examine whether this more targeted approach can further reduce the required number of measurements.

Lastly, although we employed the fluid fermionic fragments method to optimize quantum measurements, it can also be used to optimize Hamiltonian fragments for other purposes. In the fault-tolerant paradigm, one can solve the electronic structure problem by time evolving the initial guess state (with a good overlap with the ground state of $\hat{H}$), then applying the quantum phase estimation algorithm [46]. A commonly employed technique for the time evolution is Trotterization [47, 48]. Because the Hamiltonian fragments required for Trotterization are equivalent to the measurable fragments considered in this work, the fluid fermionic fragments method can also be applied in the context of Trotterization.

The only necessary modification is that instead of minimizing the number of measurements required in a VQE step, one would employ the fluid fermionic fragments method to minimize the Trotter error [48, 49].

Acknowledgments

A.F.I. is grateful to Tom O'Brien for insightful discussions and acknowledges financial support from the Google Quantum Research Program, Zapata Computing Inc., and the National Science Foundation under Grant No. NSF PHY-1748958. S.C. thanks Tzu-Ching Yen for helpful discussions and acknowledges financial support from the Swiss National Science Foundation through the Postdoc Mobility Fellowship (Grant No. P500PN-206649). This research was enabled in part by support provided by Compute Ontario and Compute Canada.

Appendix A: Initial Hamiltonian fragments

The LR decomposition is less ambiguous compared to its FR counterpart, and we use an LR decomposition procedure described in Ref. 9. Among different implementations of the FR decomposition, we employ the "greedy" FR optimization (GFRO), since it has the lowest $M(\epsilon)$ [10]. Greedy algorithms typically have low $M(\epsilon)$, and their success can be attributed to the sum of square roots appearing in

$$M_{\text{opt}}(\epsilon) = \frac{1}{\epsilon^2}\left[\sum_{\alpha=0}^{N_f}\sqrt{\text{Var}_\psi(\hat{H}_\alpha)}\right]^2 \tag{21}$$

obtained by choosing the optimal measurement allocation,

$$m_\alpha = \frac{\sqrt{\text{Var}_\psi(\hat{H}_\alpha)}}{\sum_{\beta=0}^{N_f}\sqrt{\text{Var}_\psi(\hat{H}_\beta)}}, \tag{22}$$

that minimizes $M(\epsilon)$. For a fixed sum of $\text{Var}_\psi(\hat{H}_\alpha)$, the sum of square roots in Eq. (21) is lower if the variances are distributed unevenly, and greedy approaches tend to yield an uneven distribution of $\text{Var}_\psi(\hat{H}_\alpha)$.

The computational effort of both LR and FR decompositions is reduced significantly by working with a smaller two-electron tensor over spatial orbitals, $\tilde{g}_{ijkl}$, instead of the tensor over spin-orbitals, g_{pqrs}. One can rewrite the electronic Hamiltonian as

$$\hat{H} = \sum_\sigma \sum_{ij}^{N/2} \tilde{h}_{ij}\hat{E}_{i\sigma}^{j\sigma} + \sum_{\sigma\sigma'}\sum_{ijkl}^{N/2}\tilde{g}_{ijkl}\hat{E}_{i\sigma}^{j\sigma}\hat{E}_{k\sigma'}^{l\sigma'} \tag{23}$$

where σ and σ' specify the spin-z projection, while $\tilde{h}_{ij}$ and $\tilde{g}_{ijkl}$ are one- and two-electronic integrals:

$$\tilde{h}_{ij} = \int d\vec{r}\phi_i^*(\vec{r})\left(-\frac{\nabla^2}{2} - \sum_I \frac{Z_I}{|\vec{r}-\vec{r}_I|}\right)\phi_j(\vec{r}) - \sum_k^{N/2}\tilde{g}_{ikkj} \tag{24}$$

and

$$\tilde{g}_{ijkl} = \frac{1}{2}\int\int d\vec{r}_1 d\vec{r}_2 \frac{\phi_i^*(\vec{r}_1)\phi_j(\vec{r}_1)\phi_k(\vec{r}_2)\phi_l^*(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}, \tag{25}$$

where $\phi_i(\vec{r})$ is the ith one-particle electronic basis function in the position representation, and the charge and position of the Ith nucleus are denoted by Z_I and $\vec{r}_I$.

To optimize the computational cost for LR and FR decompositions, we work with $\tilde{g}_{ijkl}$ and then subsequently convert the resulting $\tilde{\lambda}_{ij}^{(\alpha)}$ and $\tilde{U}_{ij}^{(\alpha)}$ into $\lambda_{pq}^{(\alpha)}$ and $U_{pq}^{(\alpha)}$ according to

$$\lambda_{2i-1,2j-1}^{(\alpha)} = \lambda_{2i-1,2j}^{(\alpha)} = \lambda_{2i,2j-1}^{(\alpha)} = \lambda_{2i,2j}^{(\alpha)} = \tilde{\lambda}_{ij}^{(\alpha)}, \tag{26}$$

$$U_{2i-1,2j-1}^{(\alpha)} = U_{2i,2j}^{(\alpha)} = \tilde{U}_{ij}^{(\alpha)}, \quad U_{2i,2j-1}^{(\alpha)} = U_{2i-1,2j}^{(\alpha)} = 0 \tag{27}$$

for $i,j = 1,\ldots,N/2$.

The LR decomposition is particularly efficient because it finds the rank-1 matrices $\tilde{\lambda}_{ij}^{(\alpha)} = \tilde{\eta}_i^{(\alpha)}\tilde{\eta}_j^{(\alpha)}$ by diagonalizing the two-electron tensor $\tilde{g}_{ij,kl}$ considered as a matrix with each dimension spanned by a pair of basis indices. This diagonalization gives a theoretical limit on $N_f \leq N_o(N_o+1)/2$, where $N_o = N/2$, and the less-than sign originates from a truncation of the expansion by removing terms for low magnitude eigenvalues (see Ref. 8 for further details on LR decomposition). Having a low N_f is beneficial for the fluid fermionic fragments method because the number of optimization variables ($c_p^{(\alpha)}$) is directly proportional to N_f.

The classical computational cost of LR has $\sim N_o^6$ scaling due to the cost of diagonalization, whereas the precise scaling for GFRO cannot be determined since it is a heuristic algorithm. In the examples considered in this work, the classical cost of GFRO was comparable to that of the F^3 optimization. Typically, the LR decomposition requires less computational effort than GFRO, but $M(\epsilon)$ in GFRO is lower owing to more flexible $\tilde{\lambda}_{ij}^{(\alpha)}$ [10, 11].

The αth GFRO fragment, $\hat{H}_\alpha$, is found by minimizing the L_1 norm of the $\tilde{\mathbf{G}}^{(\alpha+1)}$ tensor in

$$\sum_{ijkl\sigma\sigma'} \tilde{G}_{ijkl}^{(\alpha)} \hat{E}_{i\sigma}^{j\sigma} \hat{E}_{k\sigma'}^{l\sigma'} - \hat{H}_\alpha = \sum_{ijkl\sigma\sigma'} \tilde{G}_{ijkl}^{(\alpha+1)} \hat{E}_{i\sigma}^{j\sigma} \hat{E}_{k\sigma'}^{l\sigma'}, \tag{28}$$

where $\tilde{G}_{ijkl}^{(1)} = \tilde{g}_{ijkl}$. In each iteration, the L_1 norm of $\tilde{\mathbf{G}}^{(\alpha+1)}$ is minimized over the space of $\{\tilde{\lambda}_{ij}^{(\alpha)}, \tilde{\theta}^{(\alpha)}\}$ variables parameterizing the αth fragment,

$$\hat{H}_\alpha = \hat{U}(\tilde{\theta}^{(\alpha)})^\dagger \left(\sum_{ij\sigma\sigma'} \tilde{\lambda}_{ij}^{(\alpha)} \hat{n}_{i\sigma} \hat{n}_{j\sigma'} \right) \hat{U}(\tilde{\theta}^{(\alpha)}), \tag{29}$$

where $\hat{U}(\tilde{\theta}^{(\alpha)}) = \exp[\sum_{i>j}^{N_o} \tilde{\theta}_{ij}^{(\alpha)} \sum_\sigma (\hat{E}_{i\sigma}^{j\sigma} - \hat{E}_{j\sigma}^{i\sigma})]$. The iteration terminates when the L_1 norm of $\tilde{\mathbf{G}}^{\alpha+1}$ is below a given threshold ($1 \cdot 10^{-5}$ is used in this work).

Appendix B: Number of measurements required to obtain the Hamiltonian expectation value

In every method considered in Sec. 3, the expectation value of the Hamiltonian, $E = \langle\psi|\hat{H}|\psi\rangle$, is obtained as the sum of expectation values of the separately measured $\hat{H}_\alpha$ fragments [see Eq. (3)]. Therefore, the most straightforward estimator for E (denoted by $\bar{H}$) is the sum of estimators for individual fragments, i.e., $\bar{H} = \sum_\alpha \bar{H}_\alpha$. Each $\bar{H}_\alpha$ is the average result of M_α repeated quantum measurements of $\hat{H}_\alpha$:

$$\bar{H}_\alpha = \frac{1}{M_\alpha} \sum_{i=1}^{M_\alpha} H_{\alpha,i}, \tag{30}$$

where $H_{\alpha,i}$ is the ith outcome of measuring $\hat{H}_\alpha$.

Because each fragment is measured independently, the covariances between the fragments, $\mathrm{Cov}(\bar{H}_\alpha, \bar{H}_\beta)$, are zero, and therefore the variance of $\bar{H}$ is simply the sum of variances of $\bar{H}_\alpha$:

$$\mathrm{Var}(\bar{H}) = \sum_\alpha \mathrm{Var}(\bar{H}_\alpha). \tag{31}$$

Assuming that each M_α is large, we can invoke the central limit theorem to evaluate $\mathrm{Var}(\bar{H}_\alpha)$ using the quantum operator variances, i.e., $\mathrm{Var}(\bar{H}_\alpha) = \mathrm{Var}_\psi(\hat{H}_\alpha)/M_\alpha$. Therefore, the Hamiltonian estimator variance can be evaluated as

$$\mathrm{Var}(\bar{H}) = \sum_\alpha \frac{\mathrm{Var}_\psi(\hat{H}_\alpha)}{M_\alpha}. \tag{32}$$

To obtain the expression for $M(\epsilon)$, i.e., the total number of required measurements to obtain E with $\epsilon \equiv [\mathrm{Var}(\bar{H})]^{1/2}$ accuracy, we introduce fractional measurement allocation, $m_\alpha = M_\alpha/M$ (with $\sum_\alpha m_\alpha = 1$), then rearrange Eq. (32) as

$$M(\epsilon) = \frac{1}{\epsilon^2} \sum_\alpha \frac{\mathrm{Var}_\psi(\hat{H}_\alpha)}{m_\alpha}. \tag{33}$$

References

[1] 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". Nat. Commun. **5**, 1–7 (2014).

[2] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". New J. Phys. **18**, 023023 (2016).

[3] Ilya G. Ryabinkin, Robert A. Lang, Scott N. Genin, and Artur F. Izmaylov. "Iterative qubit coupled cluster approach with efficient screening of generators". J. Chem. Theory Comput. **16**, 1055–1063 (2020).

[4] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J Coles. "Variational quantum algorithms". Nat. Rev. Phys. **3**, 625–644 (2021).

[5] Abhinav Anand, Philipp Schleich, Sumner Alperin-Lea, Phillip W. K. Jensen, Sukin Sim, Manuel Díaz-Tinoco, Jakob S. Kottmann, Matthias Degroote, Artur F. Izmaylov, and Alán Aspuru-Guzik. "A quantum computing view on unitary coupled

cluster theory". Chem. Soc. Rev. **51**, 1659–1684 (2022).

[6] John Preskill. "Quantum computing in the NISQ era and beyond". Quantum **2**, 79 (2018).

[7] 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).

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

[9] 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 Inf. **7**, 1–9 (2021).

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

[11] Jeffrey Cohn, Mario Motta, and Robert M. Parrish. "Quantum filter diagonalization with compressed double-factorized hamiltonians". PRX Quantum **2**, 040352 (2021).

[12] Sergey B. Bravyi and Alexei Yu. Kitaev. "Fermionic quantum computation". Ann. Phys. **298**, 210–226 (2002).

[13] Jacob T. Seeley, Martin J. Richard, and Peter J. Love. "The Bravyi-Kitaev transformation for quantum computation of electronic structure". J. Chem. Phys. **137**, 224109 (2012).

[14] Ian D. Kivlichan, Jarrod McClean, Nathan Wiebe, Craig Gidney, Alán Aspuru-Guzik, Garnet Kin-Lic Chan, and Ryan Babbush. "Quantum simulation of electronic structure with linear depth and connectivity". Phys. Rev. Lett. **120**, 110501 (2018).

[15] Ophelia Crawford, Barnaby van Straaten, Daochen Wang, Thomas Parks, Earl Campbell, and Stephen Brierley. "Efficient quantum measurement of Pauli operators in the presence of finite sampling error". Quantum **5**, 385 (2021).

[16] Jérôme F. Gonthier, Maxwell D. Radin, Corneliu Buda, Eric J. Doskocil, Clena M. Abuan, and Jhonathan Romero. "Measurements as a roadblock to near-term practical quantum advantage in chemistry: Resource analysis". Phys. Rev. Research **4**, 033154 (2022).

[17] Andrew Jena, Scott Genin, and Michele Mosca. "Pauli partitioning with respect to gate sets" (2019). arXiv:1907.07859.

[18] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". Nat. Phys. **16**, 1050–1057 (2020).

[19] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. "Measurements of quantum hamiltonians with locally-biased classical shadows". Commun. Math. Phys. **391**, 951–967 (2022).

[20] Stefan Hillmich, Charles Hadfield, Rudy Raymond, Antonio Mezzacapo, and Robert Wille. "Decision diagrams for quantum measurements with shallow circuits". In 2021 IEEE International Conference on Quantum Computing and Engineering (QCE). Pages 24–34. (2021).

[21] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Efficient estimation of Pauli observables by derandomization". Phys. Rev. Lett. **127**, 030503 (2021).

[22] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. "Overlapped grouping measurement: A unified framework for measuring quantum states" (2021). arXiv:2105.13091.

[23] Charles Hadfield. "Adaptive Pauli shadows for energy estimation" (2021). arXiv:2105.12207.

[24] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F. Izmaylov. "Measuring all compatible operators in one series of single-qubit measurements using unitary transformations". J. Chem. Theory Comput. **16**, 2400–2409 (2020).

[25] Tzu-Ching Yen, Aadithya Ganeshram, and Artur F Izmaylov. "Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates" (2022). arXiv:2201.01471.

[26] Seonghoon Choi, Tzu-Ching Yen, and Artur F. Izmaylov. "Improving quantum mea-

surements by introducing "ghost" pauli products". J. Chem. Theory Comput. **18**, 7394–7402 (2022).

[27] Scott Aaronson and Daniel Gottesman. "Improved simulation of stabilizer circuits". Phys. Rev. A **70**, 052328 (2004).

[28] Zachary Pierce Bansingh, Tzu-Ching Yen, Peter D. Johnson, and Artur F. Izmaylov. "Fidelity overhead for nonlocal measurements in variational quantum algorithms". J. Phys. Chem. A **126**, 7007–7012 (2022).

[29] Guillermo García-Pérez, Matteo A.C. Rossi, Boris Sokolov, Francesco Tacchino, Panagiotis Kl. Barkoutsos, Guglielmo Mazzola, Ivano Tavernelli, and Sabrina Maniscalco. "Learning to measure: Adaptive informationally complete generalized measurements for quantum algorithms". PRX Quantum **2**, 040342 (2021).

[30] 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" (2022). arXiv:2203.07369.

[31] Adam Glos, Anton Nykänen, Elsi-Mari Borrelli, Sabrina Maniscalco, Matteo A. C. Rossi, Zoltán Zimborás, and Guillermo García-Pérez. "Adaptive POVM implementations and measurement error mitigation strategies for near-term quantum devices" (2022). arXiv:2208.07817.

[32] Ilya G. Ryabinkin, Scott N. Genin, and Artur F. Izmaylov. "Constrained variational quantum eigensolver: Quantum computer search engine in the fock space". J. Chem. Theory Comput. **15**, 249–255 (2019).

[33] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Edward Farhi, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Matthew P. Harrigan, Alan Ho, Sabrina Hong, Trent Huang, William J. Huggins, Lev Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Erik Lucero, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Hartmut Neven, Murphy Yuezhen Niu, Thomas E. O'Brien, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Doug Strain, Kevin J. Sung, Marco Szalay, Tyler Y. Takeshita, Amit Vainsencher, Theodore White, Nathan Wiebe, Z. Jamie Yao, Ping Yeh, and Adam Zalcman. "Hartree-fock on a superconducting qubit quantum computer". Science **369**, 1084–1089 (2020).

[34] Ignacio Loaiza, Alireza Marefat Khah, Nathan Wiebe, and Artur F. Izmaylov. "Reducing molecular electronic hamiltonian simulation cost for linear combination of unitaries approaches" (2022). arXiv:2208.08272.

[35] 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. Research **3**, 033055 (2021).

[36] 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).

[37] Seonghoon Choi, Ignacio Loaiza, and Artur F. Izmaylov. "Data for: Fluid fermionic fragments for optimizing quantum measurements of electronic Hamiltonians in the variational quantum eigensolver". url: doi.org/10.5281/zenodo.7335451.

[38] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien. "Low-cost error mitigation by symmetry verification". Phys. Rev. A **98**, 062339 (2018).

[39] Suguru Endo, Zhenyu Cai, Simon C. Benjamin, and Xiao Yuan. "Hybrid quantum-classical algorithms and quantum error

mitigation". J. Phys. Soc. Japan **90**, 032001 (2021).

[40] 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). arXiv:2210.00921.

[41] Jarrod R. McClean, Mollie E. Kimchi-Schwartz, Jonathan Carter, and Wibe A. de Jong. "Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states". Phys. Rev. A **95**, 042308 (2017).

[42] William J Huggins, Joonho Lee, Unpil Baek, Bryan O'Gorman, and K Birgitta Whaley. "A non-orthogonal variational quantum eigensolver". New J. Phys. **22**, 073009 (2020).

[43] Nicholas H. Stair, Renke Huang, and Francesco A. Evangelista. "A multireference quantum krylov algorithm for strongly correlated electrons". J. Chem. Theory Comput. **16**, 2236–2245 (2020).

[44] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O'Brien. "Nearly optimal measurement scheduling for partial tomography of quantum states". Phys. Rev. X **10**, 031064 (2020).

[45] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. "Fermionic partial tomography via classical shadows". Phys. Rev. Lett. **127**, 110504 (2021).

[46] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. "Simulated quantum computation of molecular energies". Science **309**, 1704–1707 (2005).

[47] Seth Lloyd. "Universal quantum simulators". Science **273**, 1073–1078 (1996).

[48] Luis A. Martínez-Martínez, Tzu-Ching Yen, and Artur F. Izmaylov. "Assessment of various hamiltonian partitionings for the electronic structure problem on a quantum computer using the trotter approximation" (2022). arXiv:2210.10189.

[49] Masuo Suzuki. "Fractal decomposition of exponential operators with applications to many-body theories and monte carlo simulations". Phys. Lett. A **146**, 319–323 (1990).