Fermionic adaptive sampling theory for variational quantum eigensolvers

Marco Majland [1,2,3] Patrick Ettenhuber [1] and Nikolaj Thomas Zinner [1,2]
[1]*Kvantify Aps, DK-2300 Copenhagen S, Denmark*
[2]*Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark*
[3]*Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark*

(Received 1 August 2023; accepted 23 October 2023; published 21 November 2023)

Quantum chemistry has been identified as one of the most promising areas where quantum computing can have a tremendous impact. For current noisy intermediate-scale quantum (NISQ) devices, one of the best available methods to prepare approximate wave functions on quantum computers is the adaptive derivative-assembled pseudo-Trotter *Ansatz* variational quantum eigensolver (ADAPT-VQE). However, ADAPT-VQE suffers from a significant measurement overhead when estimating the importance of operators in the wave function. In this work, we propose fermionic adaptive sampling theory VQE (FAST-VQE), a method for selecting operators based on importance metrics solely derived from the populations of Slater determinants in the wave function. Thus, our method mitigates measurement overheads for ADAPT-VQE as it is only dependent on the populations of Slater determinants which can simply be determined by measurements in the computational basis. We introduce two heuristic importance metrics, one based on selected configuration interaction with perturbation theory and one based on approximate gradients. In state vector and finite shot simulations, FAST-VQE using the heuristic metric based on approximate gradients converges at the same rate or faster than ADAPT-VQE and requires dramatically fewer shots.

DOI: 10.1103/PhysRevA.108.052422

I. INTRODUCTION

Quantum chemistry has been identified as one of the most promising areas where quantum computing can have great impact on industrial applications [1–4]. However, current quantum hardware is subject to noise and error and thus algorithms such as quantum phase estimation remain intractable for current and near-term devices [5,6]. Therefore, the research community has focused on developing algorithms suitable for an era of noise, error, limited qubits, and limited quantum gates [7,8]. A promising method to approximate electronic wave functions on quantum computers is the adaptive derivative-assembled pseudo-Trotter *Ansatz* variational quantum eigensolver (ADAPT-VQE) algorithm, along with its variants, which has enabled tremendous progress towards this goal [9–14]. Other adaptive algorithms include the qubit coupled cluster method and the iterative qubit coupled cluster method [15,16]. The adaptive approaches for estimating electronic wave functions are in contrast to static approaches such as unitary coupled cluster theory and its variants [17–19].

The adaptive algorithms are proven to converge to chemical accuracy with fewer parameters and more compact wave functions compared to the static algorithms. Thus, the adaptive algorithms may be more feasible for near-term applications. However, one of the primary challenges of ADAPT-VQE is the large measurement overhead incurred by estimating the importance metric for selecting relevant operators for the wave function [9]. Even estimating a single energy evaluation of a wave function through the sampling of expectation values may require significant measurement resources as was demonstrated in recent large-scale benchmarks [3]. For ADAPT-VQE, the importance metric for choosing operators from a predefined pool, $\mathcal{A}$, is the gradient of the energy. Therefore, the number of measurements necessary to rank the operators scales with the size of the pool, i.e., $O(|\mathcal{A}|)$. Since $\mathcal{A}$ typically contains two-body operators, the size of the set of operators $|\mathcal{A}|$ scales as $O(N^4)$, where N is a measure for the size of the chemical system.

In this work, we propose a method for selecting operators based on the populations of Slater determinants in the wave function in order to establish an importance metric for excitation operators. This is in stark contrast to ADAPT-VQE where the importance of operators is established using gradient measurements which requires the sampling of expectation values for each excitation operator. Sampling Slater determinants requires only the sampling of a single operator rather than $O(N^4)$ operators as in ADAPT-VQE. In fact, the required quantities for evaluating the proposed metric can be extracted from a measurement of the energy in variational quantum eigensolver (VQE), a measurement that would in any case have to be performed.

For selecting operators, we are considering two metrics, one that is related to the approximate gradient used in ADAPT-VQE and a second one that is inspired by classical selected configuration interaction (SCI) [20]. In classical SCI, the determinants used to diagonalize the Hamiltonian are chosen using an importance metric typically based on a perturbation method [21–25]. Here we consider selecting operators based on second-order Epstein-Nesbet (EN) perturbation theory [26,27]. The methods are compared to ADAPT-VQE by calculating the ground state energies of two small molecules which are typically used in benchmarks, namely H_4 and LiH. The ground state energies are calculated using state vector (infinite shot) and finite shot simulations to investigate the

performance of the methods both cases. An interesting application of SCI methods in quantum computing is presented in Ref. [28]. Using an approximate wave-function encoded in a quantum computer, the authors present a method to sample from the wave function to obtain a set of important Slater determinants which is used to perform a classical diagonalization of the Hamiltonian. Our method may be used to prepare the approximate wave function in the initial step of their algorithm. The sampling of Slater determinants may be further improved by applying methods such as those presented in [29].

The paper is organized as follows. In Sec. II, we provide the theoretical background of ADAPT-VQE and SCI. In Sec. III we provide the background for the scaling reduction in FAST-VQE and derive the gradient-based and SCI-based metrics. In Sec. IV, we provide a pseudo-algorithm for FAST-VQE and provide the computational details of our calculations which we will present and discuss in Sec. V. Finally, we conclude with a summary and present some future research avenues in Sec. VI.

II. BACKGROUND

In this section, we will provide the background necessary for understanding the construction of our method in Sec. III, starting with ADAPT-VQE and followed by SCI.

A. ADAPT-VQE

In ADAPT-VQE, an *Ansatz* is built by successively adding parametrized unitary operators acting on a reference state $|\Phi_0\rangle$, which is often taken as the Hartree-Fock (HF) ground state determinant. Thus, the ADAPT-VQE wave function in iteration k of the algorithm can be expressed as

$$|\Psi^{(k)}\rangle = \prod_{\mu\in\mathcal{A}^{(k)}} e^{-\theta_\mu \hat{A}_\mu} |\Phi_0\rangle\,, \tag{1}$$

where $\mathcal{A}^{(k)}$ is the set of operators in the wave function at iteration k, $\hat{A}_\mu = \hat{\tau}_\mu - \hat{\tau}_\mu^\dagger$, with $\hat{\tau}_\mu$ being an excitation operator, and μ enumerates the excitation. The excitation operators are chosen from a pool of operators, $\mathcal{A} = \{A_\mu\}$, based on an importance metric, $w(\hat{A}_\mu, |\Psi^{(k)}\rangle)$. In standard ADAPT-VQE, the importance metric is the gradient of the energy with respect to the parameter of the operator. The energy of the $(k+1)$th iteration may be written as

$$E^{(k+1)} = \langle\Psi^{(k)}|e^{\theta_\mu \hat{A}_\mu}\hat{H}e^{-\theta_\mu \hat{A}_\mu}|\Psi^{(k)}\rangle \tag{2}$$

such that

$$g_\mu = \left.\frac{\partial E^{(k+1)}}{\partial\theta_\mu}\right|_{\theta_\mu=0} = \langle\Psi^{(k)}|[\hat{A}_\mu, \hat{H}]|\Psi^{(k)}\rangle. \tag{3}$$

To evaluate this expression, ADAPT-VQE relies on measuring operators of the type $[\hat{A}_\mu, \hat{H}]$, yielding a significant overhead in measurements to be performed.

B. Selected CI

In SCI, determinants are selected iteratively by an importance metric in order to adaptively increase the subspace in which the CI eigenvalue problem is solved. One possibility for selecting determinants is based on perturbation theory [20]. In this paper, we consider EN perturbation theory [26,27]. EN theory weights the importance of a Slater determinant $|D\rangle$ for extending a wave function $|\Psi^{(k)}\rangle$ in iteration k as

$$E_D^{(k)} = \frac{|\langle D|\hat{H}|\Psi^{(k)}\rangle|^2}{E^{(k)} - \langle D|\hat{H}|D\rangle} = \sum_{ij}\frac{c_i c_j^* \langle D_j|\hat{H}|D\rangle\langle D|\hat{H}|D_i\rangle}{E^{(k)} - \langle D|\hat{H}|D\rangle}, \tag{4}$$

where the states $|D_i\rangle$ are Slater determinants and $c_i = \langle D_i|\Psi^{(k)}\rangle$ CI coefficients.

III. FAST-VQE

In this section, we present a method for selecting operators solely based on the population of Slater determinants in the wave function by establishing importance metrics for excitation operators. This is in stark contrast to ADAPT-VQE where the importance of operators is established by measuring the expectation value of the non-diagonal gradient operators of Eq. (3). We start this section with a discussion of sampling populations of Slater determinants and diagonal Hamiltonian measurements in Sec. III A and then build the two metrics in Secs. III C and III B.

A. Sampling populations of Slater determinants

A population of Slater determinants may, for example, be obtained from the energy evaluations in the VQE optimization or as a separate measurement. For separate measurements, given $|\Psi^{(k)}\rangle$, one may repeatedly perform measurements in the computational basis to obtain a bit string representation of determinants from $|\Psi^{(k)}\rangle$ in the HF basis. These measurements may be collected in a multiset of determinants. The multiset may be written as

$$S^{(k)} = \{|D_i\rangle, \langle D_i|\Psi^{(k)}\rangle \neq 0\}, \tag{5}$$

where the frequency of each determinant $|D_i\rangle$ is proportional to $|c_i|^2$ and where the restriction is fulfilled by construction. With this set of determinants, we can build metrics suitable to assign importance weights to operators from an operator pool $\mathcal{A}$ based on the expected contribution to the wave function. In the following sections we will introduce two such metrics.

For energy measurements, the population of Slater determinants may be obtained through sampling the diagonal elements of the Hamiltonian. In VQE, the Hamiltonian is mapped to a qubit Hamiltonian,

$$\hat{H} = \sum_a h_a \hat{P}_a, \tag{6}$$

where

$$\hat{P}_a = \bigotimes_b \hat{\sigma}_b^\alpha, \quad \alpha \in \{x, y, z\}, \tag{7}$$

denotes a product of Pauli operators. Consider a partitioning of the Hamiltonian $\hat{H} = \hat{H}^z + \hat{H}^c$ where $\hat{H}^z$ is diagonal, then we can express $\hat{H}^z$ as $\hat{H}^z = \sum_a h_a \hat{P}_a^z$, where $\hat{P}_a^z$ are products of Pauli-z operators. We can then write an energy functional depending on the wave function parameters $\boldsymbol{\theta}$ in terms of this

partitioned Hamiltonian as

$$\begin{aligned} E^{(k)}(\boldsymbol{\theta}) &= \langle\Psi^{(k)}|\hat{H}^z + \hat{H}^c|\Psi^{(k)}\rangle \\ &= \sum_a h_a \langle\Psi^{(k)}|\hat{P}_a^z|\Psi^{(k)}\rangle + \langle\Psi^{(k)}|\hat{H}^c|\Psi^{(k)}\rangle \\ &= \sum_{ai} h_a |c_i|^2 \langle D_i|\hat{P}_a^z|D_i\rangle + \langle\Psi^{(k)}|\hat{H}^c|\Psi^{(k)}\rangle \\ &= \sum_i h_{ii}|c_i|^2 + \langle\Psi^{(k)}|\hat{H}^c|\Psi^{(k)}\rangle. \end{aligned} \tag{8}$$

Thus, we can perform measurements of diagonal Hamiltonian terms in the computational basis in order to sample Slater determinants $|D_i\rangle$ in $|\Psi^{(k)}\rangle$ with a probability that is proportional to $|c_i|^2$. Note that Eq. (8) is evaluated repeatedly in order to optimize the wave function parameters $\boldsymbol{\theta}$, e.g., using VQE, such that no additional cost is introduced to calculate Slater determinant populations.

B. Heuristic gradient

To introduce the first heuristic importance metric, we start from g_μ in Eq.(3) which may be expressed as

$$\begin{aligned} g_\mu &= \langle\Psi^{(k)}|\hat{A}_\mu\hat{H} - \hat{H}\hat{A}_\mu|\Psi^{(k)}\rangle \\ &= -\langle\Psi^{(k)}|\hat{A}_\mu^\dagger\hat{H} + \hat{H}\hat{A}_\mu|\Psi^{(k)}\rangle \\ &= -2\,\mathrm{Re}(\langle\Psi^{(k)}|\hat{A}_\mu^\dagger\hat{H}|\Psi^{(k)}\rangle) \\ &= -2\sum_{ij}\mathrm{Re}(c_i^* c_j \langle D_i|\hat{A}_\mu^\dagger\hat{H}|D_j\rangle). \end{aligned} \tag{9}$$

Then, dropping the off-diagonal part of the sum in Eq. (9) yields

$$\begin{aligned} \mathrm{diag}(g_\mu) &= 2\sum_i \mathrm{Re}(|c_i|^2\langle D_i|\hat{A}_\mu^\dagger\hat{H}|D_i\rangle) \\ &= 2\sum_i |c_i|^2\,\mathrm{Re}(\langle D_i|\hat{A}_\mu^\dagger\hat{H}|D_i\rangle). \end{aligned} \tag{10}$$

The manifold into which $\hat{A}_\mu^\dagger$ excites, $\{\langle D_j|\hat{A}_\mu^\dagger, D_j \in S^{(k)}\}$, may be classically constructed. Such a manifold contains information on how the diagonal is connected to off-diagonal elements. To include that information in the final metric, a second sum over the determinants will therefore be introduced. In this regard, $S^{(k)}$ of Eq. (5) will be used directly since the number of occurrences of a determinant $|D_i\rangle$ in this multiset is proportional to $|c_i|^2$. Additionally, the second summation is introduced and all prefactors are removed, as the final ranking will not be dependent on constant factors. Thus, one obtains

$$\alpha_\mu = \sum_{D_i\in S^{(k)}}\sum_{D_j\in S^{(k)}}\mathrm{Re}(\langle D_i|\hat{A}_\mu^\dagger\hat{H}|D_j\rangle), \tag{11}$$

which concludes the construction of the first importance metric. This metric roughly corresponds to dropping the phases and prefactors from Eq. (9). Note that this expression can be evaluated classically once $S^{(k)}$ has been obtained. This importance metric will be denoted heuristic gradient (HG) in the following.

In contrast to ADAPT-VQE, it is necessary to remove operators already used in the *Ansatz*, $\mathcal{A}^k$, from the operator pool, $\mathcal{A}$, in order to avoid using the same operator twice. However, to converge to the ull configuration interaction (FCI) energy, it may be necessary to repeat operators in the *Ansatz*. Thus, whenever $\max_\mu(\alpha_\mu) < \epsilon$, the operators $\mathcal{A}^k$ are added to the pool again.

C. Heuristic selected CI

In order to introduce a second heuristically motivated metric, SCI theory will be leveraged. In contrast to SCI theory, which works with the determinants directly, it is required to build a metric that relates determinants and their frequencies in the sampling procedure to operators in order to gauge the effect of adding them to the *Ansatz*. In this section, such a metric will be constructed based on the EN criterion from Eq. (4).

First, consider the ADAPT-VQE *Ansatz* in Eq. (1). The addition of a new operator corresponds to the multiplication of a new exponential which operates on all previous exponentials and the reference wave function. Thus, the contribution must be evaluated for all determinants already in $|\Psi^{(k)}\rangle$ and appropriately weighted. The construction of the heuristic operator metric based on determinants begins by noting that $\langle D_i|\hat{A}_\mu^\dagger = \langle D_k|$ is just another determinant or zero, establishing a connection between operators and determinants. From this, it would be possible to evaluate the contribution of $\hat{A}_\mu^\dagger$ by applying the EN criterion in Eq. (4) directly using D_j as the contribution to be evaluated. However, naively sampling the operator $H\,|D_k\rangle\,\langle D_k|\,H$ comes at a significant cost with a scaling of $O(N^8)$. In order to make this manageable and to be able to evaluate this on a classical computer, the off-diagonal elements of the sum over i and j from Eq. (4) may be neglected. Furthermore, one must evaluate and sum such a metric for all the determinants that an operator $\hat{A}_\mu^\dagger$ is able to create from the determinants in $|\Psi^{(k)}\rangle$, i.e., for practical implementations, all the determinants of the multiset $S^{(k)}$. For representing the wave function in Eq. (4), the same approach as used to arrive at Eq. (11) will be used, i.e., a finite shot representation given the determinants collected in $S^{(k)}$ and using only the diagonal contributions.

This concludes the construction of the heuristic importance metric β_μ, which may be written as

$$\beta_\mu \coloneqq \sum_{D_i\in\mathcal{S}^{(k)}}\sum_{D_j\in\mathcal{S}^{(k)}}\frac{|\langle D_i|\hat{A}_\mu^\dagger\hat{H}|D_j\rangle|^2}{E^{(k)} - \langle D_i|\hat{A}_\mu^\dagger\hat{H}\hat{A}_\mu|D_i\rangle}. \tag{12}$$

This importance metric will be denoted as heuristic selected CI (HSCI). Note that also for this metric we need to remove used operators from $\mathcal{A}$, as explained in Sec. III B.

The importance metrics in Eqs. (11) and (12) both use the operator $\hat{A}^\dagger\hat{H}$ for the evaluation of the importance of an operator $\hat{A}_\mu$ when improving the wave function in the next iteration. From a set of determinants, it is trivial to evaluate the expectation values for this operator on a classical computational resource with polynomial scaling in the number of electrons and orbitals.

Algorithm 1. General pseudo-code for the adaptive algorithms. Note that for ADAPT-VQE, we skip lines 7-10.

Input: $\hat{H}$, $|\Phi_0\rangle$, E_{thr}, $n_{\max}$, $\mathcal{A}$, O_{thr}
Output: $|\Psi^{(k)}(\boldsymbol{\theta})\rangle$
1 $k = 0$
2 $|\Psi^{(k)}\rangle = |\Phi_0\rangle$
3 $E_k = \langle\Psi^{(k)}|\hat{H}|\Psi^{(k)}\rangle$
4 $\mathcal{A}^{(k)} = \{\emptyset\}$
5 $\bar{\mathcal{A}} = \mathcal{A}$
6 **while** $k < n_{max}$ *and* $\Delta E > E_{thr}$ **do**
7 $\quad \hat{A}_\mu = \arg\max_{\hat{A}_\nu \in \bar{\mathcal{A}}} w(\hat{A}_\mu, |\Psi^{(k)}\rangle)$
8 $\quad$ **if** $\sum_{\hat{A}_\nu \in \bar{\mathcal{A}}} w(\hat{A}_\nu, |\Psi^{(k)}\rangle) < O_{thr}$ **then**
9 $\quad\quad \bar{\mathcal{A}} = \mathcal{A}$
10 $\quad$ **else**
11 $\quad\quad \mathcal{A}^{(k+1)} = \mathcal{A}^{(k)} \cup \{\hat{A}_\mu\}$
12 $\quad\quad \bar{\mathcal{A}} = \bar{\mathcal{A}} \cup \{\hat{A}_\mu\}$
13 $\quad |\Psi^{(k+1)}\rangle = e^{-i\theta_\mu \hat{A}_\mu}|\Psi^{(k)}\rangle$
14 $\quad \min_\theta E_{k+1}(\boldsymbol{\theta}) = \langle\Psi^{(k+1)}|\hat{H}|\Psi^{(k+1)}\rangle$
15 $\quad \Delta E = E_k - E_{k+1}$
16 $\quad k+ = 1$

IV. COMPUTATIONAL DETAILS

In this section, the algorithms and computational details of the calculations will be reviewed, starting with a description of the algorithm in Sec. IV A, a description of the choice of operator pool in Sec. IV B, and finally with a description of the numerical experiments in Sec. IV C

A. Review of algorithms

The general algorithm for ADAPT-VQE and FAST-VQE is presented in Algorithm 1. Note that the major difference between these methods is the skipping of lines 7–10 for ADAPT-VQE. For ADAPT-VQE, the importance metric reads $w(\hat{A}_\mu, |\Psi^k\rangle)_{\text{ADAPT-VQE}} = g_\mu$, while for FAST-VQE we are using the importance metrics introduced earlier, i.e., $w(\hat{A}_\mu, |\Psi^{(k)}\rangle)_{\text{HG}} = \alpha_\mu$ and $w(\hat{A}_\mu, |\Psi^{(k)}\rangle)_{\text{HSCI}} = \beta_\mu$.

Note that modifications for ADAPT-VQE, for example TETRIS-ADAPT-VQE [13], which adds more than one operator per iteration, are also applicable to FAST-VQE. However, we do not expect the relative performance of the algorithms to differ when using these types of improvements since the importance metrics are identical for the operators despite adding more than one operator per iteration. Thus, standard implementations for ADAPT-VQE and FAST-VQE are used.

B. Choice of operator pools

In general, any type of operator pool may be utilized. However, one-body and two-body excitation operators are enough to parametrize an FCI wave function [30]. Since the quantum gates required for implementing N-body excitation operators increase rapidly with N, operator pools are typically restricted to one-body and two-body excitation operators. According to Ref. [30] all possible many-body operators may be decomposed as one-body and two-body excitation operators, specifically as infinite sequences of one- and two-body particle-hole operators. Particle-hole excitation operators are excitation operators which annihilate electrons in occupied spin orbitals in the HF reference state and create electrons in virtual spin-orbitals of the HF reference. In the original formulation of ADAPT-VQE, the operator pool consisted of general excitations (particle-hole excitations and excitations within the pure virtual-virtual or occupied-occupied blocks) in the Jordan-Wigner (JW) encoding [31]. The resulting operator pools determine the scaling and convergence of the procedures. Additionally, rather than using these physically motivated operator pools, one can build operator pools that are computationally motivated. For example, several approximations have been suggested such as qubit excitation based ADAPT-VQE (QEB-ADAPT-VQE) [11] and spin-adapted ADAPT-VQE [32]. Recently, operator pools which consider qubit-space operators were suggested [10]. In this article, we will use particle-hole excitation operators in the QEB-ADAPT-VQE encoding since the primary task of this paper is to investigate importance metrics and not the operators themselves.

C. Systems and details

Benchmarks of the algorithms are performed by calculating the ground state energies for H_4 and LiH, which are typically used to benchmark ADAPT-VQE algorithms [9,16,17]. In these calculations, the STO-3G basis set was used. The molecular integrals were obtained using PYSCF. The optimization of the wave-function parameters in VQE is calculated with the L-BFGS-B method as implemented in QISKIT [33]. For all molecules, four types of calculations were performed: one state vector simulation and three simulations with finite sampling (100, 500, and 1000 shots per expectation value estimation). The optimization of the wave function in the VQE was performed using state-vector simulations since we are restricting our study to the evaluation of importance measures for operators. Thus, the method for reusing VQE optimization measurements for FAST-VQE was not used such that finite shot simulations were performed to estimate population of Slater determinants. Since the identical number of operators must be sampled in the VQE optimization for each algorithm, we do not expect the relative comparison between the ADAPT-VQE and FAST-VQE to differ in terms of VQE optimization. The state vector and finite shot calculations were performed in QISKIT. The state vector simulation serves as a benchmark for infinite shots. All quantum simulations are compared to an FCI calculation for the same molecule–basis-set combination in PYSCF. These results are presented in Sec. V.

V. RESULTS

In this section, the results from the setup described in Sec. IV are presented. We will conclude this section with a discussion of the results.

A. H_4

In Fig. 1, we present the ground state calculations using ADAPT-VQE and FAST-VQE with both importance

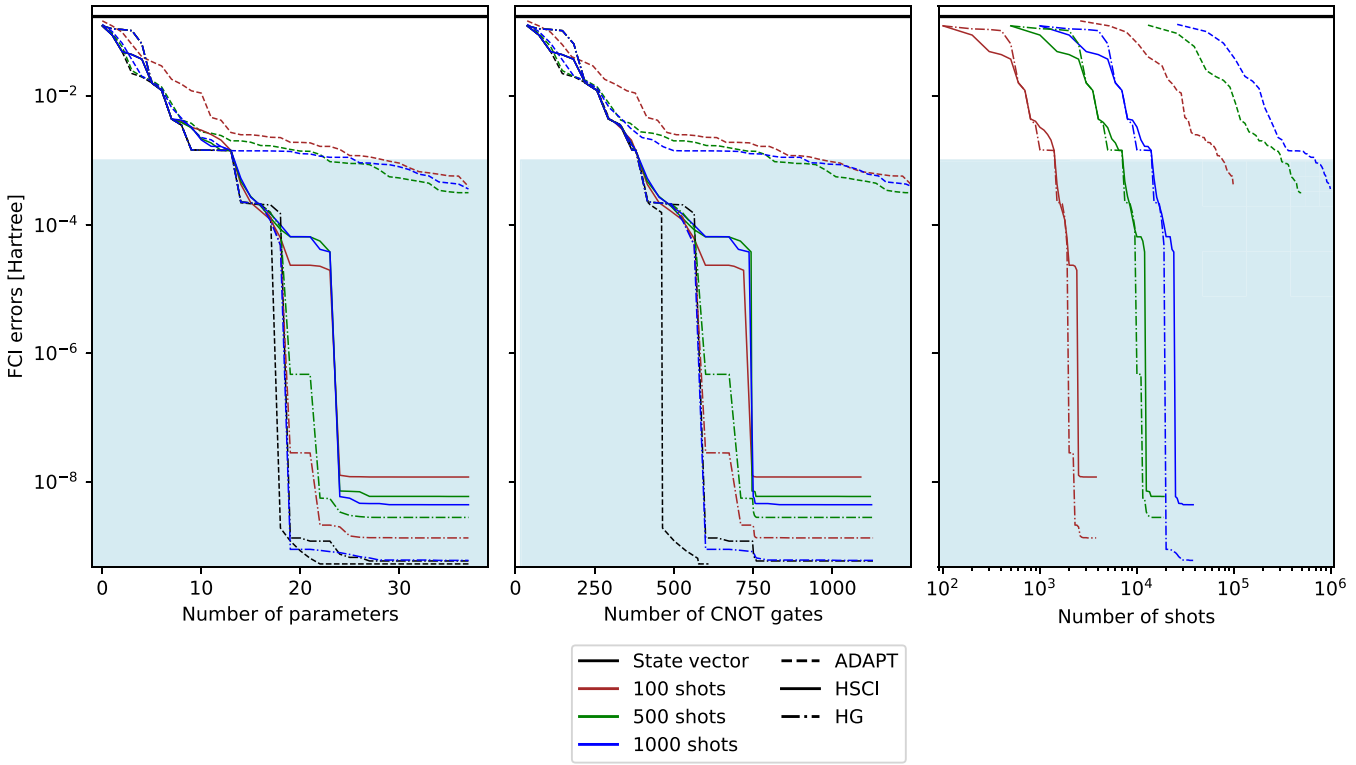

FIG. 1. Convergence of ADAPT-VQE (dashed), HSCI (full), and HG (dotted) with respect to the number of parameters (left), CNOT gates (middle), and the total number of shots (right) at a given precision of the obtained wave function for a linear H_4 chain at 1.5 Å separation between subsequent atoms. The upper horizontal line corresponds to the HF energy. The blue shaded region corresponds to chemical precision (1 kcal/mol $\approx$ 1.59 mhartree).

metrics, HSCI and HG, for a linear chain of H_4 in terms of the error relative to the FCI ground state energy. To an error of above 10^{-3} hartree with respect to FCI, the convergence in terms of the number of operators (parameters) added to the *Ansatz* is very similar for all methods and numbers of shots per operator evaluation. Beyond that point, the fastest convergence is observed for the state vector simulation for ADAPT-VQE closely followed by the state vector for HSCI and HG and finite shot simulations for HG. HSCI converges slower for finite shot simulations. The slowest convergence with the number of operators added is observed for finite shot simulations for ADAPT-VQE. While HG calculations with a finite amount of shots are converged with about 25 parameters to an error of 10^{-9} hartree, the precision for finite shot calculations using ADAPT-VQE is orders of magnitudes lower, at about 10^{-3} hartree at the same point.

With respect to the resulting *Ansatz* depth, we observe that *Ansätze* constructed with the order of operators resulting from state vector simulations using the ADAPT-VQE metric result in the most compact circuits, followed by HG and HSCI. *Ansätze* constructed with finite shot simulations for ADAPT-VQE are the least compact.

The total amounts of shots for ADAPT-VQE and FAST-VQE are very different. Both HSCI and HG converge with a total number of shots about two orders of magnitude lower than the number of shots required for finite shot simulations using ADAPT-VQE. HG requires fewer shots to obtain a given precision compared to HSCI.

B. LiH

Similar observations as for H_4 also hold true for the LiH calculations presented in Fig. 2, even though the overall convergence is slower. There are some other features to be observed in the convergence for this system. For example, the finite shot simulations for ADAPT-VQE with 100 and 500 shots per operator evaluation showed no sign of convergence and remained on the level of the HF reference state. The finite shot simulation with 1000 shots per operator evaluation shows early signs of convergence but is not able to go much below an energy difference of 10^{-2} hartree. The ADAPT-VQE state vector simulation, and all simulations for HG and HSCI converge to an energy difference of 10^{-3} hartree at roughly the same rate; here the HG convergence flattens out, while the remaining calculations continue to converge at a similar rate. Beyond the addition of roughly 30 parameters HG gets a dramatic increase in precision while the other calculations start flattening out, displaying an unintuitive and seemingly erratic behavior of convergence. It is notable that HG converges below the ADAPT-VQE state vector simulation.

With respect to *Ansatz* compactness and the number of shots required, similar conclusions hold true, displaying the same overall tendencies as observed for H_4 including the specific features described for the energy evaluation above.

C. Discussion

The dramatic difference in the number of shots between FAST-VQE and ADAPT-VQE is due to the excessive amount

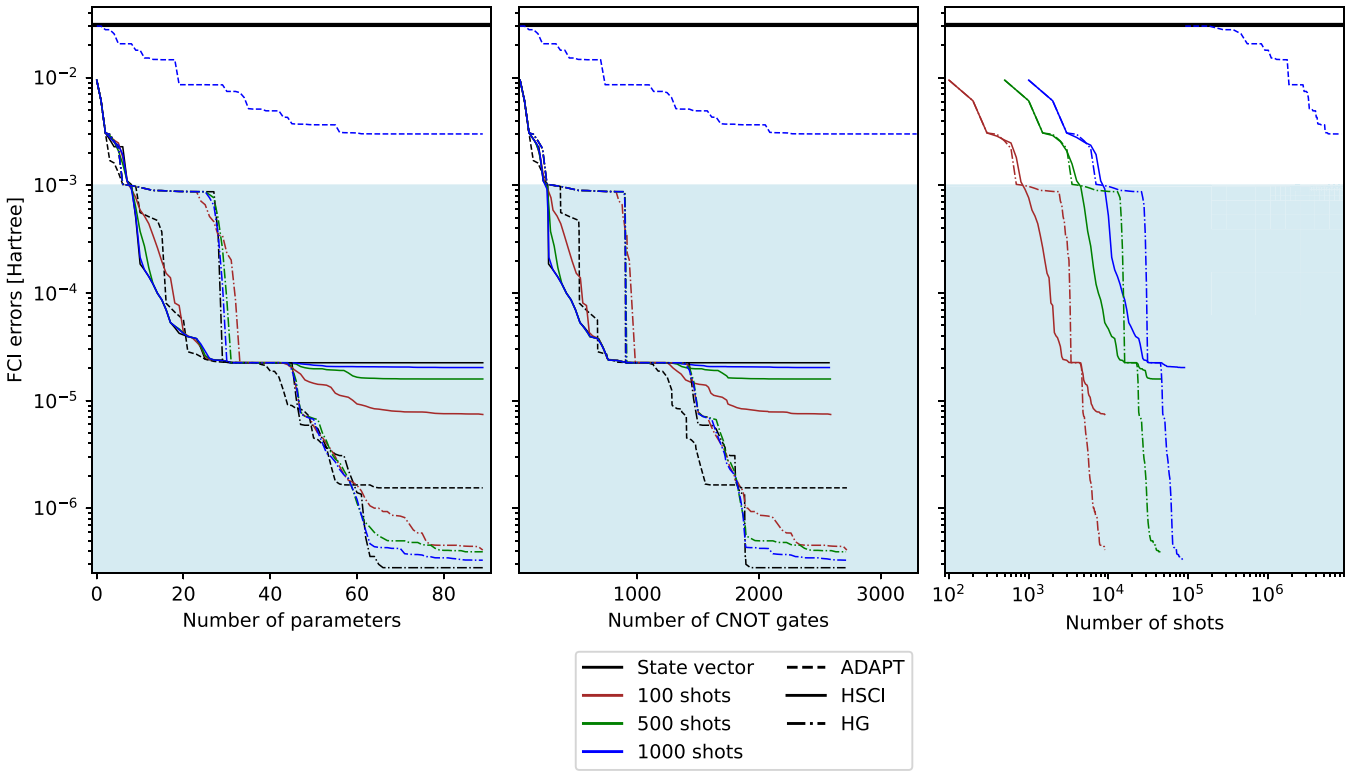

FIG. 2. Convergence of ADAPT-VQE (dashed), HSCI (full), and HG (dotted) with respect to the number of parameters (left), CNOT gates (middle), and the total number of shots (right) at a given precision of the obtained wave function for a linear LiH molecule with 1.5 Å separation between the atoms. The numbers for the finite shot simulations for ADAPT-VQE with 100 and 500 shots are not shown as the energy remained on the HF level. The upper horizontal line corresponds to the HF energy. The blue shaded region corresponds to chemical precision (1 kcal/mol $\approx$ 1.59 mhartree).

of shots necessary to measure the gradients of the operator pool, $\mathcal{A}$, of ADAPT-VQE. We can write the total amount of shots as iterations times shots for FAST-VQE whereas for ADAPT-VQE it reads iterations times shots times $|\mathcal{A}|$. Such a fact also provides another reason for the slow convergence of ADAPT-VQE when using a finite amount of shots, as the evaluation of the gradient in Eq. (3) is prone to sampling error. In contrast, Eqs. (12) and (11) for HSCI and HG are evaluated on a classical computer from states that are generated by the measurement of the energy. However, note that sampling error also effects HSCI and HG, as these methods are dependent on a representation of the weights of the determinants in the current wave function $|\Psi^{(k)}\rangle$. For ADAPT-VQE, more precise measurements of the gradients are required in order to improve convergence, while more precise sampling of the Slater determinants (diagonal elements of the Hamiltonian) becomes necessary for FAST-VQE. This is especially important when the electronic structure becomes more correlated; i.e., when many determinants are required to describe the chemical system accurately, the necessary sampling depth may become a challenge.

It must also be noted that none of the proposed metrics for selecting the next operator is optimal and that there is room for improvement. For example, despite being the overall most competitive metric, HG seems to select some suboptimal operators for LiH below 10^{-3} hartree, yet it converges at an order of magnitude below the error which the ADAPT-VQE state vector simulation achieves beyond 60 parameters; HSCI does not exhibit the same behavior. Additionally, for LiH with the HSCI metric the finite shot simulations with fewer shots achieve higher precisions, indicating that this metric does not capture some important correlations in this particular system.

The results shown here suggest that for practical purposes the introduced heuristic metrics are good enough, since they converge at a similar rate as the ADAPT-VQE state vector simulations using a finite amount of shots. However, the systems shown here are rather small and the basis sets are limited. With the two different systems investigated, we have observed quite different detailed behaviors of convergence with no clear indication for why the ordering behaves so differently with different metrics. A better theoretical understanding of the limits of this method and a more rigorous derivation of metrics could make the convergence more robust across many systems and ensure that a similar convergence rate is retained for more complicated molecules and larger basis sets.

VI. CONCLUSION

In this work, we have presented FAST-VQE, a method for selecting operators based on the populations of Slater determinants in the wave function. We have introduced two different importance metrics HG and HSCI and compared them to ADAPT-VQE in terms of the convergence to the FCI ground state energy. As was demonstrated, FAST-VQE mitigates the significant measurement overhead for ADAPT-VQE by utilizing information about the population of Slater

determinants in the wave function, whereas ADAPT-VQE must evaluate the expectation value of gradient operators. For infinite shots, ADAPT-VQE provides the most compact wave function in terms of CNOT gates but with equal amounts of parameters compared to FAST-VQE. For finite shot simulations, FAST-VQE yields more compact wave functions with dramatically reduced execution times. Of the two introduced importance metrics HG converged most rapidly and resulted in more compact circuits compared to HSCI. However, we expect that a more systematic construction of importance metrics may improve the performance and eliminate some erratic features seen, e.g., for LiH. It remains to be seen how this method performs on real quantum hardware and in combination with other operator pools and other improvements available for ADAPT-VQE. This will be the topic of future investigations.

ACKNOWLEDGMENTS

We thank Niels Kristian Kjærgård Madsen, Mads Bøttger Hansen, Mogens Dalgaard, and Stig Elkjær Rasmussen from Kvantify ApS and Mads Greisen Højlund, Rasmus Berg Jensen, and Ove Christiansen from Aarhus University for fruitful discussions.

N.T.Z. is a cofounder of Kvantify Aps. The authors have filed a provisional patent application covering the method described here.

[1] V. E. Elfving, B. W. Broer, M. Webber, J. Gavartin, M. D. Halls, K. P. Lorton, and A. Bochevarov, How will quantum computers provide an industrially relevant computational advantage in quantum chemistry? arXiv:2009.12472.

[2] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, Practical quantum advantage in quantum simulation, Nature (London) **607**, 667 (2022).

[3] J. F. Gonthier, M. D. Radin, C. Buda, E. J. Doskocil, C. M. Abuan, and J. Romero, Measurements as a roadblock to near-term practical quantum advantage in chemistry: Resource analysis, Phys. Rev. Res. **4**, 033154 (2022).

[4] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, Science **309**, 1704 (2005).

[5] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum (NISQ) algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[6] D. S. Abrams and S. Lloyd, Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors, Phys. Rev. Lett. **83**, 5162 (1999).

[7] 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).

[8] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

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

[10] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, Qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient *Ansätze* on a quantum processor, PRX Quantum **2**, 020310 (2021).

[11] Y. S. Yordanov, V. Armaos, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, Qubit-excitation-based adaptive variational quantum eigensolver, Commun. Phys. **4**, 228 (2021).

[12] Z. Lan and W. Liang, Amplitude reordering accelerates the adaptive variational quantum eigensolver algorithms, J. Chem. Theory Comput. **18**, 5267 (2022).

[13] P. G. Anastasiou, Y. Chen, N. J. Mayhall, E. Barnes, and S. E. Economou, TETRIS-ADAPT-VQE: An adaptive algorithm that yields shallower, denser circuit ansätze, arXiv:2209.10562.

[14] L. W. Bertels, H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, Symmetry breaking slows convergence of the ADAPT variational quantum eigensolver, J. Chem. Theory Comput. **18**, 6656 (2022).

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

[16] I. G. Ryabinkin, R. A. Lang, S. N. Genin, and A. F. Izmaylov, Iterative qubit coupled cluster approach with efficient screening of generators, J. Chem. Theory Comput. **16**, 1055 (2020).

[17] 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).

[18] A. Anand, P. Schleich, S. Alperin-Lea, P. W. K. Jensen, S. Sim, M. Díaz-Tinoco, J. S. Kottmann, M. Degroote, A. F. Izmaylov, and A. Aspuru-Guzik, A quantum computing view on unitary coupled cluster theory, Chem. Soc. Rev. **51**, 1659 (2022).

[19] 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 (2019).

[20] B. Huron, J. P. Malrieu, and P. Rancurel, Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions, J. Chem. Phys. **58**, 5745 (1973).

[21] L. Bytautas and K. Ruedenberg, *A priori* identification of configurational deadwood, Chem. Phys. **356**, 64 (2009).

[22] J. S. M. Anderson, F. Heidar-Zadeh, and P. W. Ayers, Breaking the curse of dimension for the electronic Schrödinger equation with functional analysis, Comput. Theor. Chem. **1142**, 66 (2018).

[23] C. F. Bender and E. R. Davidson, Studies in configuration unteraction: The first-row diatomic hydrides, Phys. Rev. **183**, 23 (1969).

[24] J. L. Whitten and M. Hackmeyer, Configuration interaction studies of ground and excited states of polyatomic molecules. I. The CI formulation and studies of formaldehyde, J. Chem. Phys. **51**, 5584 (1969).

[25] S. Evangelisti, J.-P. Daudey, and J.-P. Malrieu, Convergence of an improved CIPSI algorithm, Chem. Phys. **75**, 91 (1983).
[26] P. S. Epstein, The Stark effect from the point of view of Schroedinger's quantum theory, Phys. Rev. **28**, 695 (1926).
[27] R. K. Nesbet, Configuration interaction in orbital theories, Proc. R. Soc. London, Ser. A: Math. Phys. Eng. Sci. **230**, 312 (1955).
[28] K. Kanno, M. Kohda, R. Imai, S. Koh, K. Mitarai, W. Mizukami, and Y. O. Nakagawa, Quantum-selected configuration interaction: Classical diagonalization of Hamiltonians in subspaces selected by quantum computers, arXiv:2302.11320 [quant-ph].
[29] M. Kohda, R. Imai, K. Kanno, K. Mitarai, W. Mizukami, and Y. O. Nakagawa, Quantum expectation-value estimation by computational basis sampling, Phys. Rev. Res., **4**, 033173 (2022).
[30] F. A. Evangelista, G. K.-L. Chan, and G. E. Scuseria, Exact parameterization of fermionic wave functions via unitary coupled cluster theory, J. Chem. Phys. **151**, 244112 (2019).
[31] P. Jordan and E. Wigner, Über das Paulische Äquivalenzverbot, Z. Phys. **47**, 631 (1928).
[32] T. Tsuchimochi, M. Taii, T. Nishimaki, and S. L. Ten-no, Adaptive construction of shallower quantum circuits with quantum spin projection for fermionic systems, Phys. Rev. Res. **4**, 033100 (2022).
[33] Qiskit contributors, Qiskit: An Open-source Framework for Quantum Computing (2023), doi:10.5281/zenodo.2573505.