Lawrence Berkeley National Laboratory
Recent Work

Title
A non-orthogonal variational quantum eigensolver

Permalink
https://escholarship.org/uc/item/71b0d1rn

Journal
New Journal of Physics, 22(7)

ISSN
1367-2630

Authors
Huggins, WJ
Lee, J
Baek, U
et al.

Publication Date
2020-07-01

DOI
10.1088/1367-2630/ab867b

Peer reviewed

PAPER • OPEN ACCESS

A non-orthogonal variational quantum eigensolver

View the article online for updates and enhancements.

PAPER

A non-orthogonal variational quantum eigensolver

William J Huggins[1,2,3,7] , Joonho Lee[1,2,3,6], Unpil Baek[2,3,4], Bryan O'Gorman[2,3,5] and K Birgitta Whaley[1,2,3]

1   Department of Chemistry, University of California, Berkeley, CA 94720, United States of America
2   Berkeley Quantum Information and Computation Center, University of California, Berkeley, CA 94720, United States of America
3   Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States of America
4   Department of Physics, University of California, Berkeley, CA 94720, United States of America
5   Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, United States of America
6   Present address: Department of Chemistry, Columbia University, New York, New York 10027, United States of America.
7   Author to whom any correspondence should be addressed.

E-mail: wjhuggins@gmail.com

Keywords: variational quantum eigensolver, noisy intermediate-scale quantum, quantum computing, quantum chemistry

Abstract

Variational algorithms for strongly correlated chemical and materials systems are one of the most promising applications of near-term quantum computers. We present an extension to the variational quantum eigensolver that approximates the ground state of a system by solving a generalized eigenvalue problem in a subspace spanned by a collection of parametrized quantum states. This allows for the systematic improvement of a logical wavefunction ansatz without a significant increase in circuit complexity. To minimize the circuit complexity of this approach, we propose a strategy for efficiently measuring the Hamiltonian and overlap matrix elements between states parametrized by circuits that commute with the total particle number operator. This strategy doubles the size of the state preparation circuits but not their depth, while adding a small number of additional two-qubit gates relative to standard variational quantum eigensolver. We also propose a classical Monte Carlo scheme to estimate the uncertainty in the ground state energy caused by a finite number of measurements of the matrix elements. We explain how this Monte Carlo procedure can be extended to adaptively schedule the required measurements, reducing the number of circuit executions necessary for a given accuracy. We apply these ideas to two model strongly correlated systems, a square configuration of H_4 and the π-system of hexatriene (C_6H_8).

1. Introduction

Large, error-corrected quantum computers are expected to provide powerful new tools for understanding quantum many-body physics. For example, such devices will be able to efficiently simulate long-time dynamics [1], and through phase estimation, measure the energy of a trial wavefunction while projecting it into the eigenbasis of the Hamiltonian [2]. Prior to the availability of such devices, it is natural to ask how today's noisy, intermediate-scale quantum (NISQ) platforms may be used for similar ends. One appealing strategy, the variational quantum eigensolver (VQE) [3, 4], uses a potentially noisy quantum computer as a black box to prepare parametrized wavefunctions and measure their energy. By optimizing over the wavefunction parameters in a classical outer loop, one obtains a variational upper bound on the true ground state energy.

While it is believed that even a noisy, modestly-sized quantum computer can prepare and measure states that are out of reach for a classical computer [5], it will still likely be difficult to take advantage of this fact to surpass the capabilities of classical variational methods [6–9]. One serious challenge is that noise is particularly damaging for quantum chemical calculations that demand a high degree of precision [8–10]. Recent works have presented a variety of approaches to overcoming this difficulty, including combining error detection schemes with post selection [10–12], extrapolating to the zero-noise limit [13–15], and using additional measurements and post-processing to construct better energy estimators [12, 16–18]. A

complementary body of research has focused on developing new variational ansätze that use fewer gates and thus offer less opportunity for errors to occur [19–21]. We shall present a new approach in this latter direction that allows for a systematic increase in wavefunction complexity without a growing circuit depth.

The standard VQE approach uses a quantum computer to measure the expectation value of the Hamiltonian for some parametrized wavefunction, $|\psi(\boldsymbol{\theta})\rangle$, in conjunction with a classical coprocessor that interprets the measurement outcomes and suggests new values for the $\boldsymbol{\theta}$ parameters in order to minimize the energy [4]. In our approach, we define instead a logical ansatz

$$|\psi(\mathbf{c}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)})\rangle = \sum_{i=1}^{M} c_i |\phi_i(\boldsymbol{\theta}^{(i)})\rangle, \tag{1}$$

where each $|\phi_i(\boldsymbol{\theta}^{(i)})\rangle$ is an independently parametrized wavefunction with a compact quantum circuit description. For brevity, we shall sometimes omit the parameters and refer to these wavefunctions more compactly as $|\psi\rangle$ and $|\phi_i\rangle$. Rather than preparing the state $|\psi\rangle$ directly on our device and measuring its energy, we use our quantum computer to prepare simpler pairwise superpositions of the states $\{|\phi_i\rangle\}$. We then measure the matrix elements of the Hamiltonian and overlap matrices,

$$
\begin{aligned}
H_{ij} &= \langle \phi_i | \hat{H} | \phi_j \rangle, \\
S_{ij} &= \langle \phi_i | \phi_j \rangle.
\end{aligned}
\tag{2}
$$

This allows us to classically solve a generalized eigenvalue problem,

$$H\mathbf{c} = ES\mathbf{c}, \tag{3}$$

thereby finding the optimal $\mathbf{c}$ parameters and minimizing the energy in the subspace spanned by the set of states $\{|\phi_i\rangle\}$. The $\boldsymbol{\theta}^{(i)}$ values that parametrize each basis function $|\phi_i(\boldsymbol{\theta}^{(i)})\rangle$ can then be optimized by a classical outer loop to lower the energy further, solving a new generalized eigenvalue problem at each step.

Our approach shares certain features with a variety of recent proposals for quantum algorithms that involve solving generalized eigenvalue problems [16, 22–26]. However, our approach also differs from these works in some key respects. Most importantly, we make no assumptions about the form of the component wavefunctions $|\phi_i\rangle$, other than that they have efficient quantum circuit implementations. In the context of quantum algorithms, prior work has assumed that these wavefunctions are generated by excitations from a fixed reference state [16], by real or imaginary time evolution [22, 24–26], or by the simultaneous rotation of a set of orthogonal reference wavefunctions [23]. Two of these works in particular, references [25] and [26], were released contemporaneously with our own and provide an interesting contrast to our approach. Specifically, they require the same off-diagonal matrix element measurements used in this work but construct the non-orthogonal basis function by real-time propagation of trial wavefunctions rather than the variational approach we take here.

In the context of classical simulations, multi reference methods which make use of a superposition of configurations have a long and storied history [27–36]. Most directly similar to this work are those which demand each of the $|\phi_i\rangle$ wavefunctions to be a Slater determinant (not necessarily in the same single particle basis) [30]. This basic direction has been elaborated upon under a variety of names, including the non-orthogonal configuration interation (NOCI) method [31, 32, 35], the non-orthogonal multicomponent adaptive greedy iterative compression (NOMAGIC) algorithm [34], and the non-orthogonal multi-Slater determinant (NOMSD) expansion approach [33, 36], among others. The restriction to Slater determinants allows for the efficient classical evaluation of the required matrix elements, while the relaxation of the requirement that the determinants be orthogonal to one another allows for more flexible and accurate wavefunctions when compared to orthogonal CI expansions with the same number of determinants.

The difference between these various approaches mainly lies in the way in which they obtain a set of non-orthogonal determinants. For example, NOCI separately optimizes individual determinants by finding a collection of different solutions to the Hartree−Fock equations before performing a single diagonalization of the Hamiltonian matrix [31, 32, 35]. Other approaches more closely parallel the one we take here, iteratively adding new states and variationally optimizing their parameters [30, 33, 36]. We do not exhaustively review the classical literature here, but note that the variational approach has been found to be prone to optimization challenges and that a number of the methods we cite arise out of attempts to ameliorate this difficulty [33, 34, 36].

By taking the basis functions $|\phi_i\rangle$ to be independently parametrized quantum circuits rather than single Slater determinants, we obtain an extremely flexible form for our logical ansatz, $|\psi\rangle = \sum c_i |\phi_i\rangle$. For a wide

variety of ansatz circuits, we shall show that the required matrix element measurements between any $|\phi_i\rangle$ and $|\phi_j\rangle$ pair can be implemented efficiently using a number of quantum gates that is equal to the sum of the gates required to prepare $|\phi_i\rangle$ and $|\phi_j\rangle$, plus a small factor that scales linearly with the system size. Notably, the circuit size required is independent of the number of wavefunctions in the logical ansatz, making it possible to systematically add flexibility to $|\psi\rangle$ without increasing the required gate fidelity or coherence times of the quantum hardware.

This flexibility, however, comes at the cost of demanding more matrix element measurements. To ameliorate this cost we propose using a Monte Carlo technique to estimate the uncertainty in the ground state energy and to adaptively allocate our measurements of the matrix elements. Essentially, this scheme involves sampling from the distributions representing the uncertainty in the estimates of the Hamiltonian and overlap matrices, and solving a small generalized eigenvalue problem for each sampled matrix pair. We then characterize the resulting distribution of ground state energy values by a sample variance. We suggest a heuristic that repeatedly determines which measurement to perform by calculating the sensitivity of this sample variance to additional measurements of each of the matrix elements.

We apply these ideas to two model chemical systems, a square configuration of H_4 and the π-system of hexatriene (C_6H_8), both of which exhibit mixed strong correlation and dynamical correlation effects. In terms of strong correlation, we shall focus on a pair of strongly entangled electrons. Specifically, entanglement between a pair of electrons can lead to two exactly degenerate determinants for certain geometries of these systems, with the rest of the electrons contributing to dynamical correlation. We present two types of numerical experiments. In the first, we explore how well the ground state of these systems can be represented by an NOVQE logical ansatz, varying both the complexity of the constituent basis functions and the size of the subspace. In the second, we take a fixed set of basis wavefunctions and compare our adaptive protocol for scheduling measurements with a simpler alternative.

2. Theory

2.1. Matrix element measurement

The off-diagonal matrix elements of the Hamiltonian, $H_{ij} = \langle\phi_i|\hat{H}|\phi_j\rangle$, do not correspond to physical observables and therefore cannot be measured directly in the usual manner. Nevertheless, it is possible to construct circuits that allow us to estimate them, for example, by using the Hadamard test [37]. In this section we present a simple strategy for measuring these matrix elements. We combine ideas from recent proposals for measuring off-diagonal matrix elements that appear in other contexts [24, 38] with a trick inspired by the literature on the impossibility of black box protocols for adding controls to arbitrary unitaries [40]. Our strategy offers several benefits over a naive application of the Hadamard test. Namely, it does not require implementing controlled versions of the ansatz preparation circuits, and it enables the simultaneous measurement of matrix elements of multiple commuting observables while also yielding information about the overlap matrix elements, $S_{ij} = \langle\phi_i|\phi_j\rangle$.

For simplicity, we will describe below the case where $\hat{H}$ is a sum of commuting operators, which can easily be simultaneously measured. In the more general case, the usual Hamiltonian averaging approach of grouping the terms into multiple sets that are each simultaneously measurable and measuring the sets separately can be applied without modification [4, 8, 39, 41].

We begin by preparing the state

$$|+_{ij}\rangle := \frac{1}{\sqrt{2}}(|\phi_i\rangle|0\rangle + |\phi_j\rangle|1\rangle), \tag{4}$$

where the second register is an ancilla qubit. This task can be accomplished by using controlled versions of the unitaries $\hat{U}_i$ and $\hat{U}_j$ that prepare $|\phi_i\rangle$ and $|\phi_j\rangle$ from a fixed reference state. Given some quantum circuit that implements the unitaries $\hat{U}_i$ and $\hat{U}_j$, it is possible to construct circuits that implement the controlled version of $\hat{U}_i$ and $\hat{U}_j$, by replacing each gate in the original circuits with its controlled form. Even setting aside the difficulty of compiling such a circuit on a physical device with limited connectivity, the cost of implementing such a circuit on a near-term device (quantified by counting the number of two-qubit gates) will be substantially increased. For example, it is known that the decomposition of the Toffoli gate (the controlled–controlled-NOT gate) into a collection of single qubit and CNOT gates requires the use of six CNOT gates [42]. Given the limited coherence times and two-qubit gate fidelities of near-term hardware, we must ask if there are alternatives for implementing controlled versions of $\hat{U}_i$ and $\hat{U}_j$.

An ideal protocol might allow us to implement a controlled version of an arbitrary $\hat{U}$ using a single execution of the original, unmodified circuit that implements $\hat{U}$. Unfortunately, a single use of oracle (blackbox) access to a general $\hat{U}$ is insufficient for implementing a controlled version of $\hat{U}$ in the quantum circuit model [40]. However, if $\hat{U}_i$ and $\hat{U}_j$ preserve fermionic (or bosonic) excitation number and act

trivially on the vacuum state, then we can circumvent this no-go result. We now show how this can be accomplished in the construction of a controlled unitary operator,

$$\hat{U}_i, \hat{U}_j \rightarrow \hat{U}_i \otimes |0\rangle|0\rangle + \hat{U}_j \otimes |1\rangle|1\rangle. \tag{5}$$

We begin with a generic input state $|\psi_0\rangle|0\rangle + |\psi_1\rangle|1\rangle$, subject to the restriction that $|\psi_0\rangle$ and $|\psi_1\rangle$ are both states that are orthogonal to the state with zero particles, $|\text{vac}\rangle$.

(a) First, we adjoin an ancilla system register in the vacuum state to obtain

$$|\psi_0\rangle \otimes |\text{vac}\rangle \otimes |0\rangle + |\psi_1\rangle \otimes |\text{vac}\rangle \otimes |1\rangle.$$

(b) Treating the final qubit as the control, we apply a controlled-SWAP operation between the two system registers, resulting in

$$|\psi_0\rangle \otimes |\text{vac}\rangle \otimes |0\rangle + |\text{vac}\rangle \otimes |\psi_1\rangle \otimes |1\rangle.$$

(c) Next, we execute the unmodified circuit for $\hat{U}_i$ on the first system register, while doing the same with $\hat{U}_j$ on the second system register, yielding

$$\hat{U}_i|\psi_0\rangle \otimes |\text{vac}\rangle \otimes |0\rangle + |\text{vac}\rangle \otimes \hat{U}_j|\psi_1\rangle \otimes |1\rangle.$$

(d) We follow this with a second controlled-SWAP operation to produce the state,

$$\hat{U}_i|\psi_0\rangle \otimes |\text{vac}\rangle \otimes |0\rangle + \hat{U}_j|\psi_1\rangle \otimes |\text{vac}\rangle \otimes |1\rangle.$$

(e) Finally, we discard the now unentangled second system register to show completion of the action of the controlled unitary gate and obtain the desired result,

$$\hat{U}_i|\psi_0\rangle \otimes |0\rangle + \hat{U}_j|\psi_1\rangle \otimes |1\rangle.$$

For our purposes, we can take $|\psi_0\rangle$ and $|\psi_1\rangle$ to be the same fixed reference state, usually a Hartree–Fock state $|\psi_{\text{HF}}\rangle$. Then $|\phi_i\rangle = \hat{U}_i|\psi_{\text{HF}}\rangle$ and $|\phi_j\rangle = \hat{U}_j|\psi_{\text{HF}}\rangle$ and we see that with the last step we have successfully prepared the desired state, $|+_{ij}\rangle := \frac{1}{\sqrt{2}}(|\phi_i\rangle|0\rangle + |\phi_j\rangle|1\rangle)$. We then apply a Hadamard gate on the ancilla qubit and perform a $\hat{Z}$ measurement. It is easy to see that the expectation value of $\hat{Z}$ for the ancilla qubit will be $\langle\hat{Z}_{\text{anc}}\rangle = \text{Re}\langle\phi_i|\phi_j\rangle$. Furthermore, the post-measurement state of the system register is either

$$\frac{|\phi_i\rangle + |\phi_j\rangle}{\sqrt{2 + 2\,\text{Re}\langle\phi_i|\phi_j\rangle}}, \tag{6}$$

if the ancilla qubit was found to be in the $+1$ eigenstate, or

$$\frac{|\phi_i\rangle - |\phi_j\rangle}{\sqrt{2 - 2\,\text{Re}\langle\phi_i|\phi_j\rangle}}, \tag{7}$$

if the measurement outcome was -1. These outcomes occur with probabilities $\frac{1+\text{Re}\langle\phi_i|\phi_j\rangle}{2}$ and $\frac{1-\text{Re}\langle\phi_i|\phi_j\rangle}{2}$ respectively.

In both cases, we proceed to measure the Hamiltonian $\hat{H}$ on the system register. Depending on the result of the ancilla qubit measurement, the resulting expectation values will be either

$$\langle\hat{H}\rangle^{(+1)} = \frac{\langle\hat{H}\rangle_i + \langle\hat{H}\rangle_j + 2\,\text{Re}\langle\phi_i|\hat{H}|\phi_j\rangle}{2 + 2\,\text{Re}\langle\phi_i|\phi_j\rangle}, \tag{8}$$

or

$$\langle\hat{H}\rangle^{(-1)} = \frac{\langle\hat{H}\rangle_i + \langle\hat{H}\rangle_j - 2\,\text{Re}\langle\phi_i|\hat{H}|\phi_j\rangle}{2 - 2\,\text{Re}\langle\phi_i|\phi_j\rangle}. \tag{9}$$

Now we consider the expectation value of the operator $\hat{H}\hat{Z}_{\text{anc}}$. By multiplying each of the conditional expectation values of $\hat{H}$ by the corresponding eigenvalue of $\hat{Z}_{\text{anc}}$ and taking the appropriate weighted average, we find that

$$\langle\hat{H}\hat{Z}_{\text{anc}}\rangle = \text{Re}\langle\phi_i|\hat{H}|\phi_j\rangle. \tag{10}$$

Furthermore, if $\hat{H}$ is a sum of Pauli operators, then the usual Hamiltonian averaging approach and upper bounds on the variance of a VQE observable apply to equation (10) [41]. Therefore, by repeated measurement we can estimate $\text{Re}\langle\phi_i|\hat{H}|\phi_j\rangle$ to a fixed precision ϵ using approximately the same number of measurements that we would need to measure a diagonal matrix element to the same accuracy. A similar

approach allows us to estimate $\text{Im}\langle\phi_i|\phi_j\rangle$ and $\text{Im}\langle\phi_i|\hat{H}|\phi_j\rangle$ by starting with the state $\frac{1}{\sqrt{2}}(|0\rangle|\phi_i\rangle + i|1\rangle|\phi_j\rangle)$.

Consider an ansatz $|\theta\rangle = U(\theta)|\psi_0\rangle$ for N spin-orbitals, represented by N qubits after a Jordan-Wigner transformation, such that the size and depth of the circuit for U is independent of θ; this is typical of VQE ansätze, but the following can be easily generalized when it is not the case. Suppose also that we have a protocol for measuring the Hamiltonian H on the N-qubit register. What are the additional resources required to implement NOVQE? First, we require $2N$ qubits and at least one ancilla. The variational unitaries U_i and U_j can be applied in parallel, doubling the size of the circuit but not the depth. The measurement protocol for H can be applied without modification to the first register. For the two controlled swaps, there is a space-time tradeoff. First, consider the case without geometric constraints. Each controlled swap of the registers can be implemented using the single ancilla and N 3-qubit CSWAP gates in series on pairs of the corresponding qubits from the two registers, adding $2N\tau_{\text{CSWAP}}$ to the depth, where τ_{CSWAP} is the effective depth of the CSWAP gate. Alternatively, we can use N ancillas and in $\lceil\log_2 N\rceil$ depth produce a logical ancilla in the form of a GHZ state. Then the N CSWAPS can be done in parallel, adding only $2\tau_{\text{CSWAP}}$ to the depth.

Suppose now that we are restricted, e.g., to some subgraph of a 2D square grid, and that $U(\theta)$ can be implemented only using gates on linearly adjacent qubits. Then we can place the computational registers on adjacent rows and the ancilla at the end of one. Now, in addition to the CSWAP gates, we must use N 2-qubit SWAP gates to move the ancilla through the line, so that the contribution to the depth is now $2N(\tau_{\text{CSWAP}} + \tau_{\text{SWAP}})$. Alternatively, we can use a whole row of ancillas between the two computational rows, and in $\lceil N/2\rceil\,\tau_{\text{CNOT}}$ prepare the ancilla GHZ state as we did without geometric constraints, and again the CSWAP gates can be done in parallel[8].

2.2. Diagonalization with uncertainty

Given a collection of states $\{|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_n\rangle\}$, we are interested in determining the minimum energy state in the subspace that they span. To do this, we use our protocol described above to measure the matrix elements of the Hamiltonian and overlap matrices (equation (2)), and solve the generalized eigenvalue problem (equation (3)). However, because we perform only a finite number of measurements of each of matrix element, we have some level of statistical uncertainty. In this section, we shall lay out a simple Monte Carlo strategy to estimate the resulting uncertainty in the minimum eigenvalue of equation (3). We shall aim to provide a self-contained presentation for convenience, but we note that this approach is related to a long tradition of applying Monte Carlo methods to statistical problems, including the diagonalization of noisy matrices [43–45].

We model the experimentally determined values of each matrix element using a normal distribution. In practice, the experimental measurements of the matrix elements are individually described by draws from Bernoulli random variables, but variational quantum algorithms typically work in the regime where the average of such measurements are well-approximated by a normal distribution [4]. In the context of an actual experiment, one could approximately determine the parameters of these distributions from the experimental measurement record of the Hamiltonian and overlap matrix elements.

For the purposes of the numerical experiments in this work, we determine the variance of the Hamiltonian matrix element measurements using the upper bounds described in references [4] and [41]. Similarly, we observe that our scheme for measuring the overlap matrix elements will have a variance that is at most $\frac{1}{m}$, where m is the number of measurements performed, and we use this upper bound as an approximation to the true variance. We use these approximations both in our simulation of the experimental measurement record and in our subsequent protocol to determine the uncertainty in the ground state energy. Throughout this section, we use a notation which separates the intrinsic component of the variance, which we denote by σ^2, from the scaling with the number of measurements, m.

Experimentally, we only have access to estimates of $\langle\phi_i|\hat{H}|\phi_j\rangle$ and $\langle\phi_i|\phi_j\rangle$ from our measurement record, which we denote by $\tilde{h}_{ij}$ and $\tilde{s}_{ij}$. Taken together with our estimates of the variances, $\tilde{\sigma}^2_{H_{ij}}$ and $\tilde{\sigma}^2_{S_{ij}}$, we can define the random variables

$$\tilde{H}'_{ij} = \tilde{h}_{ij} + \frac{\tilde{\sigma}_{H_{ij}}}{\sqrt{m_{H_{ij}}}}\mathcal{N}(0,1), \tag{11}$$

$$\tilde{S}'_{ij} = \tilde{s}_{ij} + \frac{\tilde{\sigma}_{S_{ij}}}{\sqrt{m_{S_{ij}}}}\mathcal{N}(0,1). \tag{12}$$

---

[8] Note that technically we should distinguish between different values for τ_{CSWAP} depending on geometric constraints on the 2-qubit gates into which the CSWAP is decomposed, e.g., between when the control qubit is in the middle of the three on a line and when it is at one of the ends.

These distributions represent our uncertainty about the true value of the matrix elements given the limited information provided by our experimental data.

To quantify the corresponding uncertainty in the ground state energy in the NOVQE subspace, we use a Monte Carlo sampling procedure. We accomplish this by repeatedly drawing from the distributions $\tilde{H}'_{ij}$ and $\tilde{S}'_{ij}$, and solving the resulting generalized eigenvalue problems. However, it is possible that the noise in our matrix element measurements and subsequent sampling destroys the positive semi-definite character of the overlap matrix. To deal with this, we follow the canonical orthogonalization procedure described in reference [46], discarding the eigenvalues of the sampled overlap matrices that are less than some numerical cutoff (and their associated eigenvectors). Each sampled pair of matrices yields a sample from the unknown distribution over possible NOVQE ground state energies. We then quantify our uncertainty in our estimate of this lowest eigenvalue by calculating the sample variance of this distribution of possible energies, σ_{MC}^2.

It is important to note that this distribution is not Gaussian and that its mean is not an unbiased estimate of the ground state energy in the NOVQE subspace [43]. This is true for a number of reasons, but it can be seen, for example, by considering the fact that the usual second order correction to the energy is quadratic in the off-diagonal matrix elements. Therefore, even unbiased and normally distributed noise in the matrix elements leads to a bias in the estimated eigenvalues. Furthermore, the rate at which our estimate of the mean and variance of the distribution over possible NOVQE ground state energies converges (with respect to the number of Monte Carlo samples) will vary according to the underlying distribution. The most meaningful consequences of this for our purposes are that convergence with respect to the number of Monte Carlo samples should be checked before being relied upon and that one should be cautious in using the standard error to generate error bars. As the number of measurements made increases and the amount of uncertainty diminishes, these effects are naturally suppressed.

2.2.1. Experiment design heuristic

In the previous section, we proposed a Monte Carlo scheme for estimating the uncertainty in the NOVQE ground state energy caused by a finite number of measurements of the individual matrix elements. By repeatedly sampling from $\tilde{H}'_{ij}$ and $\tilde{S}'_{ij}$ and solving the resulting generalized eigenvalue problems, we obtained a distribution over NOVQE ground state energies with some mean μ_{MC} and standard deviation σ_{MC}. Here we build on this proposal to determine the relative impact of performing additional measurements. Ultimately, our goal is to create a reasonable heuristic for adaptively scheduling measurements to most efficiently use a limited amount of device time.

We determine the impact of additional measurements of the matrix elements on the uncertainty in the ground state energy by calculating the derivatives of the sample standard deviation, σ_{MC}, with respect to the number of measurements performed, $m_{H_{ij}}$ and $m_{S_{ij}}$. The resultant quantities, $\frac{\mathrm{d}\sigma_{\mathrm{MC}}}{\mathrm{d}m_{H_{ij}}}$ and $\frac{\mathrm{d}\sigma_{\mathrm{MC}}}{\mathrm{d}m_{S_{ij}}}$, estimate how much we expect the sample deviation to shrink if we perform additional measurements of H_{ij} or S_{ij}. Note that we take these derivatives with respect to $m_{H_{ij}}$ and $m_{S_{ij}}$ only in the Monte Carlo sampling procedure of equations (11) and (12), not in the original measurements on the device. Therefore, no additional quantum resources are required. We use the tensor flow software package to perform the Monte Carlo sampling of $\tilde{H}'_{ij}$ and $\tilde{S}'_{ij}$, to calculate of the ground state energies, and to estimate σ_{MC} [47]. This enables us to evaluate the analytical expressions for each of $\frac{\mathrm{d}\sigma_{\mathrm{MC}}}{\mathrm{d}m_{H_{ij}}}$ and $\frac{\mathrm{d}\sigma_{\mathrm{MC}}}{\mathrm{d}m_{S_{ij}}}$ (for a fixed set of samples drawn from $\tilde{H}'_{ij}$ and $\tilde{S}'_{ij}$) without explicitly deriving the equations.

To optimally allocate our experimental measurements, we begin by performing a small number of measurements of each matrix element. We then estimate the derivatives $\frac{\mathrm{d}\sigma_{\mathrm{MC}}}{\mathrm{d}m_{H_{ij}}}$ and $\frac{\mathrm{d}\sigma_{\mathrm{MC}}}{\mathrm{d}m_{S_{ij}}}$. Using these estimates, we simply choose to perform additional measurements on the matrix element whose corresponding derivative is the most negative. In practice, we perform these measurements in small batches so that the time taken by the classical processing of the measurement results is small compared to the time performing the measurements. By repeating this process for many steps, until we either achieve the desired accuracy or exhaust a pre-defined measurement budget, we aim to approximately optimize allocation of measurements between the different terms.

2.3. Implementation

The tools presented above are applicable for use with a variety of different ansätze, and are subject only to the constraint that the circuits act on a common reference state and conserve fermionic excitation number, in order to benefit from the efficient implementation of the matrix element measurements. For our numerical experiments, we shall focus on a particular class of wavefunctions known as **k**-fold products of **u**nitary **p**aired **c**oupled **c**luster with **g**eneralized **s**ingle and **d**ouble excitations [20] (*k*-UpCCGSD). These wavefunctions have the appealing properties that (1) the required circuit depth scales only linearly in the size of the system, and (2) they can be systematically improved by increasing the refinement parameter *k*.

We briefly review this ansatz below and then describe in more detail the implementation details of our numerical experiments.

2.3.1. The k-UpCCGSD ansatz

The essential idea behind the k-UpCCGSD ansatz is to act on a reference state, Hartree–Fock in the case of this paper, with a product of k elementary blocks. Each block is an independently parametrized approximation to a unitary coupled cluster circuit generated by a sparse cluster operator containing only single and paired double excitations [48, 49]. To this end, the wavefunction (before the Trotter approximation involved in compiling the circuits) is defined as follows.

$$|\psi\rangle = \prod_{x=1}^{k} \left(e^{\hat{T}^{(x)} - \hat{T}^{(x)\dagger}} \right) |\phi_0\rangle, \tag{13}$$

where each cluster operator

$$\hat{T} = \sum_{ia} t_{ii}^{aa} \hat{a}_{a\alpha}^{\dagger} \hat{a}_{a\beta}^{\dagger} \hat{a}_{i\beta} \hat{a}_{i\alpha} + t_i^a (\hat{a}_{a\alpha}^{\dagger} \hat{a}_{i\alpha} + \hat{a}_{a\beta}^{\dagger} \hat{a}_{i\beta}) \tag{14}$$

possesses an independent collection of variational parameters. (We omit the (x) superscript for simplicity and use latin and greek letters for spatial and spin indices respectively.)

In contrast with the standard unitary coupled cluster single and doubles (UCCSD), k-UpCCGSD only includes doubles excitations which collectively move a pair of electrons from one spatial orbital to another. The resulting loss of flexibility is ameliorated by the use of generalized excitations that do not distinguish between occupied and unoccupied orbitals [50, 51], and by the k-fold repetition of the elementary circuit block. As a result, the number of free parameters in the ansatz scales as $O(kN^2)$. We make use of the generalized swap networks of reference [52] to implement a single Trotter step approximation to the k-UpCCGSD ansatz with the open source Cirq and OpenFermion-Cirq libraries [53, 55].

The circuits consist of the following elementary gates:

- $FSIM_2(w_0, w_1) = \exp(iH)$ for $H = \left(w_0 |10\rangle\langle 01| + \text{h.c.} \right) + w_1 |11\rangle\langle 11|$,
- $FSWAP = SWAP \cdot CZ$, and
- $FSIM_4(w) = \exp(iH)$ for

  $H = w|0011\rangle\langle 1100| + \text{h.c.}$

Because each FSWAP immediately follows an $FSIM_2$, we can compile them together to get an effective duration τ_2. Let τ_4 be the effective duration of $FSIM_4$. The overall depth then is $kN(\tau_2 + \tau_4/2)$. There are $\binom{N}{2}$ pairs of $FSIM_2$ and FSWAP gates, and $\binom{N/2}{2}$ $FSIM_4$ gates. This is simply an upper bound; the depth may be compressed further by combining the compilation of each $FSIM_4$ with the immediately following 2-qubit gates.

2.3.2. Computational details

The quantum chemical calculations of the full configuration interaction (FCI) ground states and Hartree–Fock (HF) reference wavefunctions were performed using the open source packages Psi4 and OpenFermion [54, 55]. We optimized the ground state energy in the NOVQE subspace by varying the parameters of the most recently added ansatz wavefunction, diagonalizing the Hamiltonian and overlap matrices at each step. Inspired by recent proposals for adaptive ansatz construction [21, 56, 57], each k-UpCCGSD wavefunction was grown iteratively by adding a single UpCCGSD block at a time, as described in more detail below. We performed this optimization using the Scipy implementation of the quasi-Newton limited-memory BFGS (L-BFGS-B) method [58, 59], treating the ground state energy in the NOVQE subspace as the objective function. We calculated the gradient at each step using a finite difference method with a step size of $\delta = 10^{-6}$. Each circuit was optimized using up to 2000 gradient evaluations, terminating early if the magnitude of all components of the gradient fell below 10^{-5}.

In order to escape local minima, we repeatedly applied random kicks to the variational parameters. After each 500 gradient evaluations we compared the current value of the objective function to the best observed value and reset the parameters if appropriate. Subsequently, we added random values drawn from the a normal distribution with zero mean and variance $\sigma^2 = 1$ (after 500 steps), $\sigma^2 = 10^{-1}$ (after 1000 steps), or $\sigma^2 = 10^{-2}$ (after 1500 steps). The best observed value of the energy across this whole procedure is the one we ultimately report. We randomly initialized the parameters of the $k = 1$ circuits by drawing from a normal distribution with mean 0 and variance $\sigma^2 = 10^{-6}$. Parameters for circuits with higher values of k were initialized by taking the parameters from an optimized circuit with $k - 1$ UpCCGSD blocks and

appending a new block with random variational parameters drawn from the same distribution, $\mathcal{N}(0, 10^{-6})$.

3. Results

H$_4$ is often used as a small test bed for single-reference coupled-cluster methods [60–64]. We shall focus on the square (D$_{4h}$) geometry here. The system exhibits two exactly degenerate determinants at the D$_{4h}$ geometry, leading to a mix of strong and weak correlation effects. Another important class of chemical systems to investigate is hydrocarbons. In this work, we shall study a simple hydrocarbon, hexatriene (C$_6$H$_8$). The interesting aspect of this molecule is that the torsional PES of a double bond leads to a strong correlation problem. At $\theta = 90°$, it exhibits two exactly degenerate determinants and therefore it is strongly correlated. To form the active space, we include the entire set of π electrons in the system along with both Π and Π^* orbitals. The resulting active space is then (6e, 6o), and this also possesses a good mixture of weak and strong correlation.

In the following subsections, we present the results of two types of experiments related to our proposed NOVQE approach on these chemical systems and discuss the potential utility of our approach for more general chemical problems. With the first class of experiments, we focus on understanding how effectively the ground state can be represented by a linear combination of parametrized wavefunctions, optimized using the gradient-based approach we described above. We vary both the complexity of the individual ansatz wavefunctions by adjusting the number of circuit blocks (k) in the k-UpCCGSD ansatz and the number of states (M) in the NOVQE subspace. For these calculations, we neglect the challenges posed by a finite number of measurements and the impact of circuit noise. In our second set of numerical experiments, we explore the extent to which our proposal for an adaptive measurement scheme is successful in reducing the number of circuit repetitions required to resolve the NOVQE ground state energy to a fixed precision. In both cases we compare the NOVQE energies with the numerically exact FCI energies.

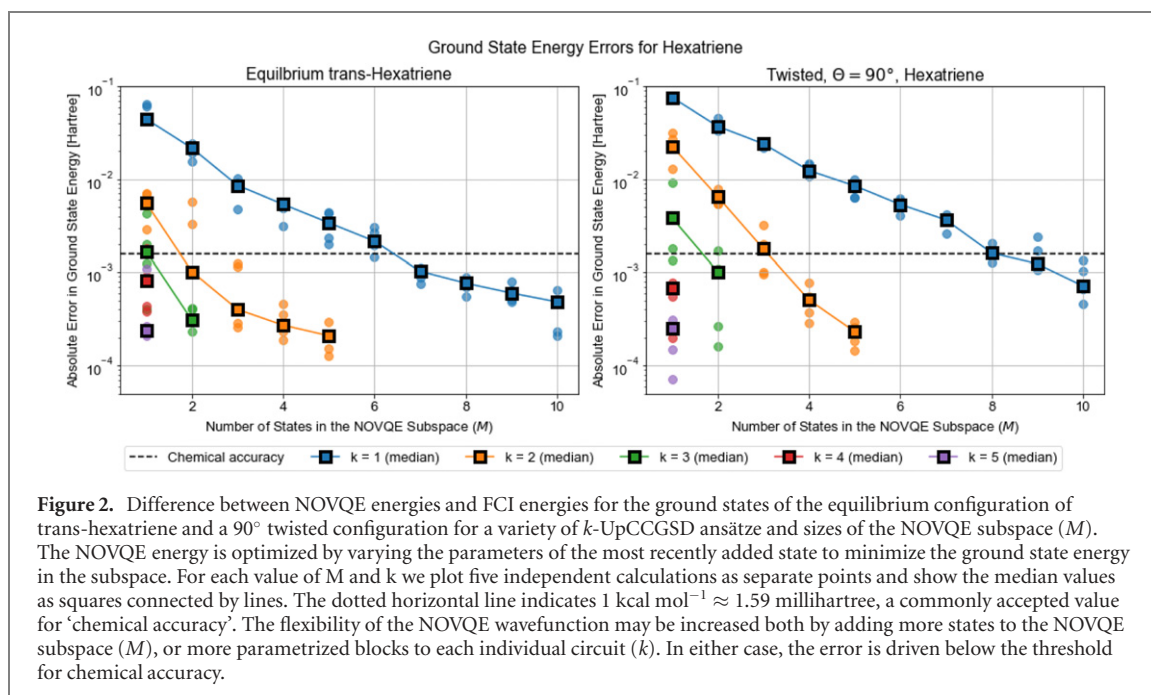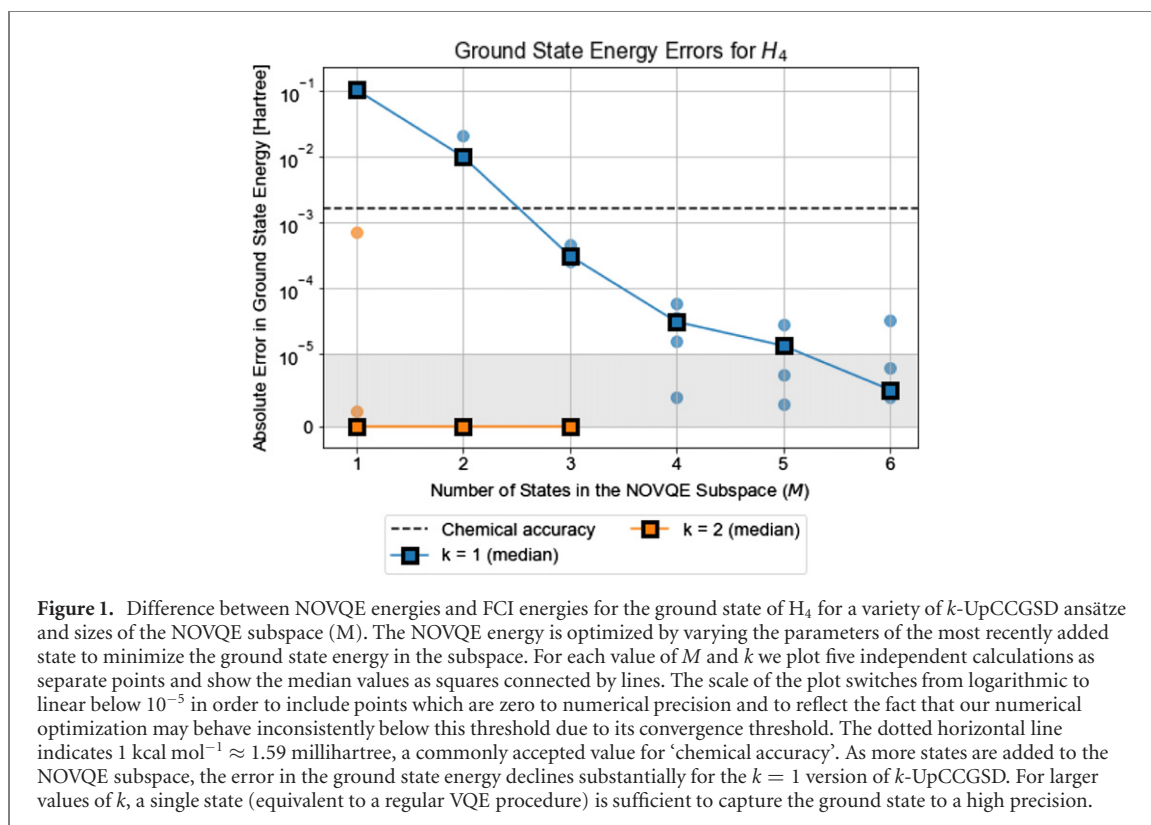### 3.1. NOVQE ground state energies

3.1.1. A hydrogen complex, H$_4$

Figure 1 presents data on the application of NOVQE to the square geometry of H$_4$ with fixed bond distance $R_{H-H} = 1.23$ Å in a minimal STO-3G basis set, an $N = 8$ qubit problem. We consider the performance of the k-UpCCGSD ansatz for $k = 1$ and $k = 2$ with $M = 1$ up to $M = 6$ states in the NOVQE subspace, noting that $M = 1$ is equivalent to the regular VQE procedure. For each value of k and M we perform five independent calculations and consider the error in the median ground state energy found by the optimization procedure as a proxy for the ansatz's ability to describe the ground state. Note that figure 1 switches from a logarithmic scale to a linear one below 10^{-5}, in order to include points that are zero to numerical precision and to reflect the fact that our numerical optimization may behave inconsistently below this threshold due to its convergence threshold.

Focusing first on understanding the behavior of the wavefunctions in the context of the standard VQE approach ($M = 1$), we note that for $k \geqslant 2$ the k-UpCCGSD ansatz is essentially exact for this problem. Looking more closely at the data for $k = 2$, $M = 1$ in figure 1, one can see that two of the five calculations failed to find the global optimum (the pale orange points). In general, we found that the optimization of this ansatz was challenging. We expect these challenges to become more severe with increasing system size, and when the stochastic nature of the quantum measurements are taken into account.

In the case of $k = 1$ we observe that we can systematically improve the accuracy of the estimated ground state energy by increasing the number of states included in the NOVQE subspace (M). Given $M = 3$ independent copies, even this relatively simple ansatz is able to represent that ground state almost exactly. This supports our thesis that a collection of ansatz states which are individually not capable of targeting a desired state may be fruitfully combined to yield a sufficiently powerful logical ansatz. However, the measurements of the off-diagonal matrix elements for NOVQE require slightly more than twice the gate count necessary for the measurements of individual ansatz states in the regular VQE formalism. For this particular system, it may therefore be more effective to use a single $k = 2$ ansatz than multiple $k = 1$ circuits.

3.1.2. Hexatriene

Here we present our results for the ground state energy of two molecular configurations of hexatriene (C$_6$H$_8$) in an STO-3G basis with an active space of 6 electrons in 6π orbitals ($N = 12$ qubits). Here, due to the system's increased complexity, we consider circuits with up to $k = 5$ UpCCGSD blocks and subspace sizes as large as $M = 10$. In figure 2 we show the calculations for an equilibrium geometry (the trans isomer, obtained by performing geometry optimization using density functional theory) and a configuration with a

Figure 1. Difference between NOVQE energies and FCI energies for the ground state of H$_4$ for a variety of k-UpCCGSD ansätze and sizes of the NOVQE subspace (M). The NOVQE energy is optimized by varying the parameters of the most recently added state to minimize the ground state energy in the subspace. For each value of M and k we plot five independent calculations as separate points and show the median values as squares connected by lines. The scale of the plot switches from logarithmic to linear below 10^{-5} in order to include points which are zero to numerical precision and to reflect the fact that our numerical optimization may behave inconsistently below this threshold due to its convergence threshold. The dotted horizontal line indicates 1 kcal mol^{-1} ≈ 1.59 millihartree, a commonly accepted value for 'chemical accuracy'. As more states are added to the NOVQE subspace, the error in the ground state energy declines substantially for the $k = 1$ version of k-UpCCGSD. For larger values of k, a single state (equivalent to a regular VQE procedure) is sufficient to capture the ground state to a high precision.

Figure 2. Difference between NOVQE energies and FCI energies for the ground states of the equilibrium configuration of trans-hexatriene and a 90° twisted configuration for a variety of k-UpCCGSD ansätze and sizes of the NOVQE subspace (M). The NOVQE energy is optimized by varying the parameters of the most recently added state to minimize the ground state energy in the subspace. For each value of M and k we plot five independent calculations as separate points and show the median values as squares connected by lines. The dotted horizontal line indicates 1 kcal mol^{-1} ≈ 1.59 millihartree, a commonly accepted value for 'chemical accuracy'. The flexibility of the NOVQE wavefunction may be increased both by adding more states to the NOVQE subspace (M), or more parametrized blocks to each individual circuit (k). In either case, the error is driven below the threshold for chemical accuracy.

90° twist on the central carbon–carbon double bond respectively. We provide the geometries for these two configurations in appendix A, tables 1 and 2.

Once again we notice that increasing the circuit complexity by taking larger values of k provides a substantial benefit, driving the estimated ground state energy well below the threshold for chemical accuracy without resorting to the multiple states of the NOVQE formalism. Likewise, as the number of NOVQE states increases, the NOVQE ground state energy reaches chemical accuracy even with the most limited ansatz. For hexatriene we see that multiple $k = 1$ states are able to achieve a performance on par with a single $k = 4$ state. The NOVQE procedure for the $k = 1$ states requires almost a factor of four less circuit depth and half as many quantum gates as performing a single VQE with the $k = 4$ state.

Interestingly, for the $k = 1$ case in both configurations, and the $k = 2$ case in the twisted configuration, figure 2 shows regimes where the error in the ground state energy decreases exponentially as a function of M. We compare this with the observation in classical non-orthogonal electronic structure calculations, where a small number of determinants are often sufficient to capture most of the wavefunction, but a long tail of dynamic correlation can result in a slow convergence to the true ground state as determinants are added to the variational space [30–36, 65]. The classical intractability of calculating matrix elements between different coupled cluster wavefunctions means that relatively little work has been done on the representational power of wavefunctions like those used in NOVQE. This is in contrast with another class of quantum non-orthogonal methods which, by virtue of building their basis states by time-evolving a set of reference wavefunctions, demonstrate exponential convergence by construction [25, 26]. In the future, it would be interesting to determine whether the rapid convergence with respect to M we observe here gives way to a regime of slow convergence, like that observed in classical non-orthogonal methods, when more challenging systems are treated.

3.2. NOVQE matrix element measurements

In the previous subsection we presented data on the performance of NOVQE in the absence of noise during the circuit execution and measurement process. Now we consider the effects of statistical noise during measurement. Specifically, we determine how many circuit repetitions are necessary to evaluate the ground state energy within a target precision, for a subspace defined by a fixed set of NOVQE states. For simplicity, we do not combine this analysis with an investigation of the optimization procedure. Instead, we take the optimized circuit parameters for a collection of M NOVQE states and compare the effectiveness of the adaptive protocol we described in section 2.2.1 to a simpler alternative for determining the ground state energy in the subspace spanned by the optimized states, which we shall explain below.

The simpler protocol, which we shall refer to as non-adaptive, consists of measuring each matrix element of the Hamiltonian and overlap matrices the same number of times. For the adaptive protocol, we repeatedly use the procedure described in section 2.2.1 to select a particular matrix element and perform measurements in batches of $\approx 10^5$ circuit repetitions. For the purpose of this comparison, we treat a 'measurement' of a particular Hamiltonian or overlap matrix element as a draw from a Gaussian random variable whose mean is the true value of the matrix element and whose variance is set by the upper bound described in reference [4], scaled by the number of measurements performed. Note that in a real experiment, or a finer-grained simulation, the Hamiltonian has to be decomposed into groups of terms that can be simultaneously measured [1]: one could apply an adaptive scheme like the one we propose to schedule measurements between these groups as well. For both kinds of numerical experiments we calculate a 2σ error bar using a bootstrapping sample size of 200 with the techniques of section 2.2.

3.2.1. A hydrogen complex, H_4

In figure 3 we plot the actual trajectories of the estimates for the ground state energies, together with their error bars for both the adaptive and non-adaptive approaches to measurement. We show two realizations of this numerical experiment applied to an NOVQE simulation of H_4 with $M = 4$ of the 1-UpCCGSD states. In both cases, we see that the adaptive protocol converges more quickly towards the NOVQE ground state energy than the non-adaptive one. We find that the data qualitatively supports the assumption that the variance in the ground state energy estimate settles into an asymptotic regime where its behavior is well described by the relationship
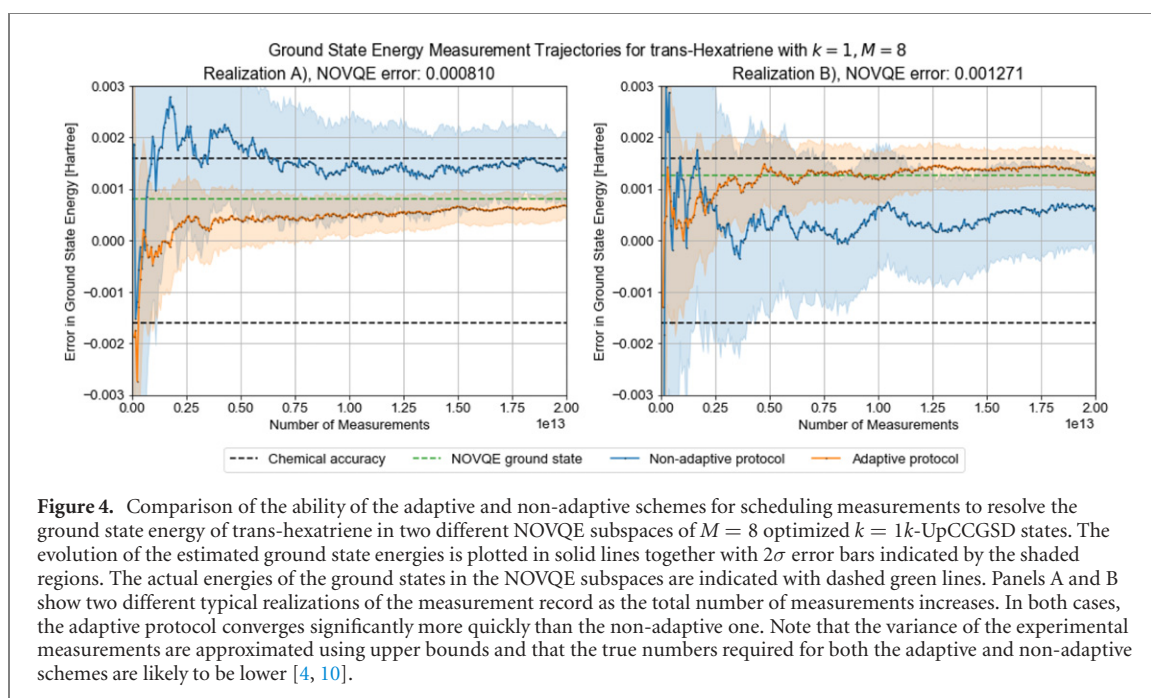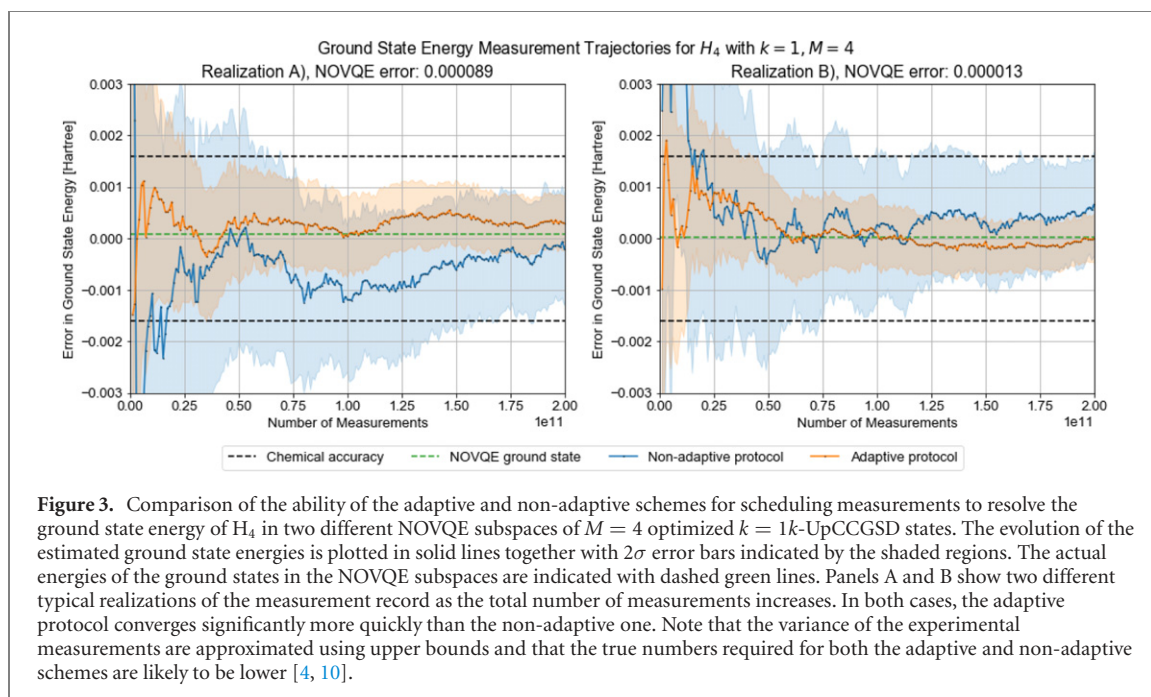
$$\sigma^2(N_m) \approx \frac{\kappa}{(N_m)}, \tag{15}$$

where (N_m) indicates the total number of measurements performed and κ is a constant, which we shall refer to as the 'intrinsic variance'. For these particular realizations, we find κ to be approximately $5.3 \cdot 10^4 \, E_h^2$ and $5.5 \cdot 10^4 \, E_h^2$ for the non-adaptive scheme in panels (A) and (B), and approximately $1.4 \cdot 10^4 \, E_h^2$ and $9.7 \cdot 10^3 \, E_h^2$ for the adaptive scheme. Using the same upper bounds to calculate the variance for a regular VQE calculation performed on the same system would yield $\kappa \approx 28 \, E_h^2$.

For some fixed precision $\sigma^2(N_m)$, the relative increase (or decrease) in the number of measurements between the two approaches can be determined by setting the ratio κ/N_m to be the same, i.e.,

$$N_m^{(1)}/N_m^{(2)} \approx \kappa^{(1)}/\kappa^{(2)},$$

where 1 refers to non-adaptive and 2 to adaptive. Therefore, for these applications to H_4, our scheme for iterative measurement achieves a modest reduction in variance. When targeting a fixed accuracy this would translate into a few-fold (≈ 3.7 for realization A and ≈ 5.7 for realization B) savings in the total number of measurements. Carrying out the same comparison between our adaptive measurement scheme for NOVQE and ordinary VQE, we see that, unfortunately, the overall measurement cost is still two orders of magnitude

Figure 3. Comparison of the ability of the adaptive and non-adaptive schemes for scheduling measurements to resolve the ground state energy of H_4 in two different NOVQE subspaces of $M = 4$ optimized $k = 1k$-UpCCGSD states. The evolution of the estimated ground state energies is plotted in solid lines together with 2σ error bars indicated by the shaded regions. The actual energies of the ground states in the NOVQE subspaces are indicated with dashed green lines. Panels A and B show two different typical realizations of the measurement record as the total number of measurements increases. In both cases, the adaptive protocol converges significantly more quickly than the non-adaptive one. Note that the variance of the experimental measurements are approximated using upper bounds and that the true numbers required for both the adaptive and non-adaptive schemes are likely to be lower [4, 10].

Figure 4. Comparison of the ability of the adaptive and non-adaptive schemes for scheduling measurements to resolve the ground state energy of trans-hexatriene in two different NOVQE subspaces of $M = 8$ optimized $k = 1k$-UpCCGSD states. The evolution of the estimated ground state energies is plotted in solid lines together with 2σ error bars indicated by the shaded regions. The actual energies of the ground states in the NOVQE subspaces are indicated with dashed green lines. Panels A and B show two different typical realizations of the measurement record as the total number of measurements increases. In both cases, the adaptive protocol converges significantly more quickly than the non-adaptive one. Note that the variance of the experimental measurements are approximated using upper bounds and that the true numbers required for both the adaptive and non-adaptive schemes are likely to be lower [4, 10].

($\approx 300-500$) larger than that required for energy measurement in an ordinary VQE approach. In order for NOVQE, or other forms of quantum non-orthogonal methods to be made practically useful, this increased measurement time will have to be accounted for and minimized.

3.2.2. Hexatriene

As in our analysis of H_4, we compare the proposed adaptive approach to distributing measurements between the elements of the Hamiltonian and overlap matrices with a non-adaptive one. We do so by choosing collections of optimized NOVQE states and applying both methods to determine the ground state energy in the resulting subspaces. In this case we choose to use $M = 8$ states, each of which is generated by a 1-UpCCGSD circuit, and focus on the equilibrium configuration of trans-hexatriene. Examining the two realizations of this experiment plotted in figure 4, we see immediately that the increased difficulty of this problem compared to H_4 is reflected in the much larger gaps between the FCI ground states and the ground states in the NOVQE subspaces, as well in the larger numbers of measurements required for convergence.

Figure 4 shows the same substantial difference between the performances of the adaptive and non-adaptive approaches that was observed for H_4. In panel B, we see that the true ground state of the subspace lies outside of the error bars for the non-adaptive scheme during small portions of the measurement procedure. This is a manifestation of the phenomenon mentioned in section 2.2, where using an insufficient number of Monte Carlo samples may result in misestimating the magnitude of the uncertainty in the ground state energy. We note that the adaptive scheme moves quickly to a regime where the uncertainty estimates are reliable even with a small number of samples. We once again observe that the variance qualitatively converges with the expected long-time $\frac{1}{N}$ behavior of equation (15) for most of the numerical experiment. Therefore, we can determine the intrinsic variance κ, defined in equation (15), for each method and compare their statistical efficiencies.

For the non-adaptive scheme we observe $\kappa \approx 2.4 \cdot 10^6$ and $\kappa \approx 2.9 \cdot 10^6$ for panels A and B, while for the adaptive scheme we see $\kappa \approx 3.7 \cdot 10^5$ and $\kappa \approx 6.5 \cdot 10^5$. The reference value for a regular VQE calculation is $\kappa \approx 1.6 \cdot 10^2$, determined using the same bounds assumed throughout this comparison [4, 41]. Comparing with the simpler H_4 example, we see that the adaptive scheme for measuring the NOVQE ground state energy of hexatriene results in a slightly larger gain when compared to the non-adaptive scheme, but still falls short of the goal of reducing the number of measurements to an experimentally plausible number. One promising avenue to further reducing this cost is the adaptation of recently proposed strategies for measurement in the context of regular VQE to NOVQE [10]. These strategies have been shown to reduce the number of circuit executions by orders of magnitude when compared with the bounds used to derive the number of measurements in this work. Further study is required in order to determine if one can alter the optimization process of the NOVQE states themselves or their coefficients, in order to achieve an additional reduction in the measurement cost.

4. Discussion and outlook

We have introduced an extension to the variational quantum eigensolver that calls for the ground state energy to be approximated by solving a generalized eigenvalue problem in a subspace that is spanned by a linear combination of M parametrized quantum wavefunctions. The resulting logical wavefunction ansatz is a linear combination of all M states in the subspace, but its properties can be determined by only pairwise measurements of the Hamiltonian and overlap matrices. In particular, construction of all off-diagonal matrix elements required for the generalized eigenvalue equation can be made using the unmodified state preparation circuits for each parameterized basis wavefunction, together with $O(N)$ additional two-qubit gates, where N is the number of spin-orbitals. Therefore, it is possible to increase the flexibility of the ansatz without requiring additional coherent quantum resources. By analogy with the non-orthogonal configuration interaction method of classical quantum chemistry [31, 32, 35], we call our approach the non-orthogonal variational quantum eigensolver, NOVQE.

Our proposal necessitates off-diagonal measurements of the Hamiltonian and overlap matrices. We perform these using a modified Hadamard test. Naively, this would require us to implement controlled versions of the quantum circuits for state preparation. To avoid this cost, we demanded that the state preparation circuits all act on a common reference state and preserve fermionic excitation number. This allowed us to avoid the need to add controls to the ansatz circuits, by instead performing controlled swap operations between two copies of the system register, a cost that scales linearly and modestly with the system size.

To determine the ground state energy in the subspace, our approach requires that we measure all M^2 elements of the Hamiltonian and overlap matrices in the NOVQE subspace. We presented a statistical strategy for estimating the uncertainty in the resultant ground state energy estimate for a given uncertainty in the matrix elements. We also pointed out how the machinery that generates these estimates can be leveraged in a Monte Carlo sampling process to determine which matrix element should be chosen for additional measurements to optimally reduce the uncertainty. We proposed an iterative approach, in which small batches of measurements are repeatedly performed according to this Monte Carlo prescription, to minimize the overall number of circuit repetitions required by our NOVQE method.

We demonstrated an implementation of our approach using a collection of k-UpCCGSD wavefunctions to approximate the ground state of two model strongly-correlated systems, a square geometry of H_4 and the π-space of hexatriene in two configurations. Growing the NOVQE subspace by adding and optimizing one state at a time, we showed how a collection of ansätze which individually struggle to represent the ground state can be fruitfully combined combined to form a more powerful logical ansatz. In our numerical experiments we observed that the marginal utility of adding additional states to the NOVQE subspace remained large, even as the size of the space increased. It is interesting to compare this with the commonly noted behavior of classical non-orthogonal methods, which generate a collection of non-orthogonal Slater

determinants and diagonalize in the resulting subspace [30–36]. These approaches eventually enter a regime where convergence slows down significantly as states are added to the subspace, sometimes before the desired accuracy is reached. This suggests that the there is a benefit in NOVQE's ability to make use of wavefunctions more sophisticated than the Slater determinants available to classical non-orthogonal methods, allowing for a balance between the number of distinct wavefunctions and their flexibility.

To characterize our proposal for adaptively scheduling measurements to minimize the number of circuit repetitions required by our approach, we focused on quantifying the number of measurements required to approximate the ground state energy in a fixed NOVQE subspace. For the purposes of this investigation we approximated the variance of the individual matrix element measurements using the bounds described in references [4] and [41]. For both our square H_4 and our equilibrium configuration of trans-hexatriene, we optimized collections of NOVQE states and froze their parameters. We then applied our adaptive approach for scheduling measurements and compared it to a simpler non-adaptive scheme, in which each matrix element was measured the same number of times. We found that our adaptive approach used somewhat fewer measurements than a simpler non-adaptive strategy, but still dramatically more than it would take to measure the energy in the standard VQE formalism. It would be worthwhile to understand whether similar challenges appear for other proposed quantum non-orthogonal methods [22–26].

We can imagine several routes towards ameliorating this difficulty and developing NOVQE further. First, having states that are nearly linearly dependent in the NOVQE subspace can dramatically increase the cost of measurement. Developing an optimization strategy for the individual states, or their coefficients, that regularizes this behavior away would be useful. Related to this is the possibility of extending the tools for measuring analytical gradients of parametrized quantum circuits to work with the NOVQE formalism. Another avenue for future work would be the development of good initialization strategies for NOVQE, potentially using reference states derived from a classical NOCI calculation. Finally, recent work has shown that a measurement strategy based on factorizations of the two-electron integral tensor can dramatically reduce the cost of the standard VQE approach, lowering the number of separately measured terms from $O(N^4)$ to $O(N)$ [10]. The resulting cost reduction is especially large when compared to the type of bounds used throughout this paper [4, 41]. Adapting this approach for use with NOVQE is likely to offer a significant improvement.

Beyond these modifications to the NOVQE approach outlined in this paper, it is also conceivable that the tools we have presented might be usefully employed in other ways. For example, we have focused here on the variational optimization of a logical ansatz that is a superposition of individual parametrized wavefunctions. An alternative is to take inspiration from reference [22] and from the classical NOCI method [31, 32, 66], and optimize the individual wavefunctions separately, solving the generalized eigenvalue problem only once with the final collection of states. In this vein, there are several recent proposals which form a non-orthogonal basis using a collection of time-evolved reference states [25, 26]. Another possible direction to pursue is the inclusion of one or more states in the NOVQE subspace that can be classically optimized, only turning to the use of more general parametrized quantum circuits to prepare small corrections to the classically tractable states. All of these ideas have the potential to benefit from the tools we have developed for efficiently performing the required matrix element measurements.

In summary, this work has presented a promising new extension to the VQE formalism and highlighted both its advantages and its drawbacks, some of which may be of general concern for developers of other quantum non-orthogonal methods. We have also presented a strategy for compiling off-diagonal matrix element measurements and promoted a general approach to Monte Carlo estimation of uncertainty. The circuit simulations of the k-UpCCGSD ansatz presented here add to the analyses of references [20, 67, 68]. We believe that the ability of our NOVQE to trade off coherent quantum resources for additional measurements may prove to be a useful tool in making use of NISQ-era quantum hardware for studying challenging strongly correlated systems.

In the final stages of preparing this manuscript two works were posted which independently developed approaches using the matrix elements between collections of quantum states for other applications. One appears in the context of variational quantum algorithms for solving linear systems of equations [69], while the other proposes a strategy for approximating the low energy subspace of a Hamiltonian in terms of time-evolved trial wavefunctions [25].

Acknowledgments

Table 1. The geometry of the equilibrium configuration of trans-hexatriene.

Atom	X	Y	Z
C	(0.598 7833,	0.296 9975,	0.000 0000)
H	(0.652 0887,	1.382 2812,	0.000 0000)
C	(−0.598 7843,	−0.297 0141,	0.000 0000)
H	(−0.652 0904,	−1.382 2967,	0.000 0000)
C	(−1.860 7210,	0.419 5548,	0.000 0000)
H	(−1.801 0551,	1.503 6080,	0.000 0000)
C	(−3.053 1867,	−0.169 3136,	0.000 0000)
H	(−3.968 5470,	0.405 3361,	0.000 0000)
H	(−3.147 9810,	−1.248 5605,	0.000 0000)
C	(1.860 7264,	−0.419 5599,	0.000 0000)
H	(1.801 0777,	−1.503 6141,	0.000 0000)
C	(3.053 1816,	0.169 3296,	0.000 0000)
H	(3.968 5551,	−0.405 2992,	0.000 0000)
H	(3.147 9561,	1.248 5793,	0.000 0000)

Table 2. The geometry of the 90° twisted configuration of hexatriene.

Atom	X	Y	Z
C	(0.598 7833,	0.296 9975,	0.000 0000)
H	(1.371 6346,	−0.068 3717,	0.670 7370)
C	(−0.598 7843,	−0.297 0141,	0.000 0000)
H	(−1.371 6354,	0.068 3544,	0.670 7361)
C	(−0.948 4080,	−1.419 7297,	−0.850 4282)
H	(−0.172 1763,	−1.780 3215,	−1.518 3873)
C	(−2.139 0983,	−2.012 1775,	−0.852 0831)
H	(−2.355 4088,	−2.846 8591,	−1.503 7144)
H	(−2.935 3514,	−1.677 2360,	−0.198 2062)
C	(0.948 4189,	1.419 7134,	−0.850 4230)
H	(0.172 1980,	1.780 3171,	−1.518 3881)
C	(2.139 1167,	2.012 1462,	−0.852 0613)
H	(2.355 4502,	2.846 8291,	−1.503 6834)
H	(2.935 3585,	1.677 1903,	−0.198 1764)

Appendix A. Hexatriene Geometries

The equilibrium geometry was obtained from the geometry optimization with ωB97X-D [70] and cc-pVTZ [71] using a development version of Q-chem [72]. The 90° twisted configuration was obtained by rotating the middle C–C double bond out-of-plane. All distances are given in angstroms.

ORCID iDs

William J Huggins ⓘ https://orcid.org/0000-0003-2735-1380
K Birgitta Whaley ⓘ https://orcid.org/0000-0002-7164-4757

References

[1]   Lloyd S 1996 Universal quantum simulators *Science* **273** 1073−8
[2]   Aspuru-Guzik A, Dutoi A D, Love P J and Head-Gordon M 2005 Simulated quantum computation of molecular energies *Science* **309** 1704−7
[3]   Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 A variational eigenvalue solver on a photonic quantum processor *Nat. Commun.* **5** 4213
[4]   McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 The theory of variational hybrid quantum-classical algorithms *New J. Phys.* **18** 023023
[5]   Boixo S, Isakov S V, Smelyanskiy V N, Babbush R, Ding N, Jiang Z, Bremner M J, Martinis J M and Neven H 2018 Characterizing quantum supremacy in near-term devices *Nat. Phys.* **14** 595−600
[6]   Wecker D, Hastings M B and Troyer M 2015 Progress towards practical quantum variational algorithms *Phys. Rev.* A **92** 042303
[7]   McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 Barren plateaus in quantum neural network training landscapes *Nat. Commun.* **9** 4812
[8]   Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets *Nature* **549** 242−6
[9]   Colless J I, Ramasesh V V, Dahlen D, Blok M S, Kimchi-Schwartz M E, McClean J R, Carter J, de Jong W A and Siddiqi I 2018 Computation of molecular spectra on a quantum processor with an error-resilient algorithm *Phys. Rev.* X **8** 011021

[10] Huggins W J, McClean J, Rubin N and Jiang Z *et al* 2019 Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers arXiv:1907.13117 [quant-ph]

[11] McArdle S, Yuan X and Benjamin S 2019 Error-Mitigated digital quantum simulation *Phys. Rev. Lett.* **122** 180501

[12] Bonet-Monroig X, Sagastizabal R, Singh M and O′Brien T E 2018 Low-cost error mitigation by symmetry verification *Phys. Rev. A* **98** 062339

[13] Temme K, Bravyi S and Gambetta J M 2017 Error mitigation for short-depth quantum circuits *Phys. Rev. Lett.* **119** 180509

[14] Otten M and Gray S K 2019 Accounting for errors in quantum algorithms via individual error reduction *npj Quantum Information* **5** 11

[15] Kandala A, Temme K, Córcoles A D, Mezzacapo A, Chow J M and Gambetta J M 2019 Error mitigation extends the computational reach of a noisy quantum processor *Nature* **567** 491–5

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

[17] McClean J R, Zhang J, Rubin N C, Babbush R and Neven H 2020 Decoding quantum errors with subspace expansions *Nat. Commun.* **11** 636

[18] Sagastizabal R *et al* 2019 Error mitigation by symmetry verification on a variational quantum eigensolver arXiv:1902.11258 [quant-ph]

[19] Kivlichan I D, McClean J, Wiebe N, Gidney C, Aspuru-Guzik A, Chan G K-L and Ryan B 2018 Quantum simulation of electronic structure with linear depth and connectivity *Phys. Rev. Lett.* **120** 110501

[20] Lee J, Huggins W J, Head-Gordon M and Whaley K B 2019 Generalized unitary coupled cluster wave functions for quantum computation *J. Chem. Theory Comput.* **15** 311–24

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

[22] Motta M, Sun C, Tan A T K, O'Rourke M J, Ye E, Minnich A J, Brandão F G S L and Chan G K-L 2020 Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution *Nat. Phys.* **16** 205–10

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

[24] Kyriienko O 2020 Quantum inverse iteration algorithm for programmable quantum simulators *npj Quantum Information* **7**

[25] Parrish R M and McMahon P L 2019 Quantum filter diagonalization: quantum eigen decomposition without full quantum phase estimation (arXiv:1909.08925 [quant-ph])

[26] Stair N H, Huang R and Evangelista F A 2019 A multireference quantum krylov algorithm for strongly correlated electrons (arXiv:1911.05163 [physics.chem-ph])

[27] Condon E U, Condon E U and Shortley G H 1951 *The Theory of Atomic Spectra* (Cambridge: Cambridge University Press)

[28] Jeziorski B and J Monkhorst H 1981 Coupled-cluster method for multideterminantal reference states *Phys. Rev. A* **24** 1668–81

[29] Werner H-j and Reinsch E-a 1982 The self-consistent electron pairs method for multiconfiguration reference state functions *J. Chem. Phys.* **76** 3144–56

[30] Koch H and Dalgaard E 1993 Linear superposition of optimized non-orthogonal slater determinants for singlet states *Chem. Phys. Lett.* **212** 193–200

[31] Malmqvist P 1986 Calculation of transition density matrices by nonunitary orbital transformations *Int. J. Quantum Chem. Lecture Notes* **30** 479–94

[32] Thom A J W and Head-Gordon M 2009 Hartree–Fock solutions as a quasidiabatic basis for nonorthogonal configuration interaction *J. Chem. Phys.* **131** 124113

[33] Jiménez-Hoyos C A, Rodríguez-Guzmán R and Scuseria G E 2013 Multi-component symmetry-projected approach for molecular ground state correlations *J. Chem. Phys.* **139** 204102

[34] McClean J R and Aspuru-Guzik A 2015 Compact wavefunctions from compressed imaginary time evolution *RSC Adv.* **5** 102277 102283

[35] Sundstrom E J and Head-Gordon M 2014 Non-orthogonal configuration interaction for the calculation of multielectron excited states *J. Chem. Phys.* **140** 114103

[36] Borda E J, Gomez J and Morales M A 2019 Non-orthogonal multi-slater determinant expansions in auxiliary field quantum monte carlo *J. Chem. Phys.* **150** 074105

[37] Kitaev A Y 1995 Quantum measurements and the abelian stabilizer problem arXiv:quant-ph/9511026

[38] Schuld M, Bergholm V, Gogolin C, Izaac J and Killoran N 2019 Evaluating analytic gradients on quantum hardware *Phys. Rev. A* **99** 032331

[39] Izmaylov A F, Yen T-C, Lang R A and Verteletskyi V 2020 Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method *J. Chem. Theory Comput.* **16** 190–5

[40] Araújo M, Feix A, Costa F and Brukner Č 2014 Quantum circuits cannot control unknown operations *New J. Phys.* **16** 093026

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

[42] Shende V V and Markov I L 2008 On the CNOT-cost of TOFFOLI gates arXiv:0803.2316 [quant-ph]

[43] Shinozuka M and Astill C J 1972 Random eigenvalue problems in structural analysis *AIAA Journal* **10** 456–62

[44] Benaroya H 1992 Random eigenvalues, algebraic methods and structural dynamic models *Appl. Math. Comput.* **52** 37–66

[45] Soize C 2005 Random matrix theory for modeling uncertainties in computational mechanics *Comput. Methods Appl. Mech. Eng.* **194** 1333–66

[46] Szabo A and Ostlund N S 2012 *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Mineola, NY: Dover)

[47] Abadi M *et al* 2016 Tensor flow: a system for large-scale machine learning *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* pp 265–83

[48] Stein T, Henderson T M and Scuseria G E 2014 Seniority zero pair coupled cluster doubles theory *J. Chem. Phys.* **140** 214113

[49] Kowalski K 2018 Properties of coupled-cluster equations originating in excitation sub-algebras *J. Chem. Phys.* **148** 094104

[50] Nakatsuji H 2000 Structure of the exact wave function *J. Chem. Phys.* **113** 2949–56

[51] Nooijen M 2000 Can the eigenstates of a many-body hamiltonian be represented exactly using a general two-body cluster expansion? *Phys. Rev. Lett.* **84** 2108–11

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

[53]   The Cirq Developers 2019 (https://github.com/quantumlib/OpenFermion-Cirq)

[54]   Parrish R M *et al* 2017 Psi4 1.1: an open-source electronic structure program emphasizing automation, advanced libraries, and interoperability *J. Chem. Theory Comput.* **13** 3185–97

[55]   McClean J R *et al* 2017 OpenFermion: the electronic structure package for quantum computers arXiv:1710.07629 [quant-ph]

[56]   Tang H L, Barnes E, Grimsley H R, Mayhall N J and Economou S E 2019 Qubit-ADAPT-VQE: an adaptive algorithm for constructing hardware-efficient ansatze on a quantum processor arXiv:1911.10205 [quant-ph]

[57]   Ryabinkin I G, Lang R A, Genin S N and Izmaylov A F 2020 Iterative qubit coupled cluster approach with efficient screening of generators *J. Chem. Theory Comput.* **16** 1055–63

[58]   Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat Methods* **17** 261–72

[59]   Zhu C, Byrd R H, Lu P and Nocedal J 1997 Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization *ACM Trans. Math. Softw.* **23** 550–60

[60]   Paldus J, Piecuch P, Pylypow L and Jeziorski B 1993 Application of Hilbert-space coupled-cluster theory to simple $(H_2)_2$ model systems: planar models *Phys. Rev.* A **47** 2738–82

[61]   Mahapatra U S, Datta B and Mukherjee D 1999 A size-consistent state-specific multireference coupled cluster theory: Formal developments and molecular applications *J. Chem. Phys.* **110** 6171

[62]   Kowalski K and Piecuch P 2000 Complete set of solutions of multireference coupled-cluster equations: the state-universal formalism *Phys. Rev.* A **61** 052506

[63]   Jankowski K and Paldus J 1980 Applicability of coupled-pair theories to quasidegenerate electronic states: a model study *Int. J. Quantum Chem.* **18** 1243–69

[64]   Evangelista F A, Allen W D and Schaefer H F 2006 High-order excitations in state-universal and state-specific multireference coupled cluster theories: model systems *J. Chem. Phys.* **125** 154113

[65]   Lee J, Malone F D and Morales M A 2019 An auxiliary-field quantum monte carlo perspective on the ground state of the dense uniform electron gas: an investigation with hartree-fock trial wavefunctions *J. Chem. Phys.* **151** 064122

[66]   Eric J S and Head-Gordon M 2014 Non-orthogonal configuration interaction for the calculation of multielectron excited states *J. Chem. Phys.* **140** 114103

[67]   Grimsley H R, Claudino D, Economou S E, Barnes E and Mayhall N J 2020 Is the trotterized UCCSD ansatz chemically well-defined? *J. Chem. Theory Comput.* **16** 1–6

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

[69]   Huang H-Y, Bharti K and Rebentrost P 2019 Near-term quantum algorithms for linear systems of equations arXiv:1909.07344 [quant-ph]

[70]   Chai J-D and Head-Gordon M 2008 Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections *Phys. Chem. Chem. Phys.* **10** 6615–20

[71]   Dunning T H 1989 Gaussian basis sets for use in correlated molecular calculations: I. The atoms boron through neon and hydrogen *J. Chem. Phys.* **90** 1007–23

[72]   Shao Y *et al* 2015 Advances in molecular quantum chemistry contained in the Q-chem 4 program package *Mol. Phys.* **113** 184–215