

Contextual Subspace Variational Quantum Eigensolver

William M. Kirby¹, Andrew Tranter^{1,2}, and Peter J. Love^{1,3}

¹Department of Physics and Astronomy, Tufts University, Medford, MA 02155

²Cambridge Quantum Computing, 9a Bridge Street Cambridge, CB2 1UB United Kingdom

³Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973

We describe the *contextual subspace variational quantum eigensolver* (CS-VQE), a hybrid quantum-classical algorithm for approximating the ground state energy of a Hamiltonian. The approximation to the ground state energy is obtained as the sum of two contributions. The first contribution comes from a noncontextual approximation to the Hamiltonian, and is computed classically. The second contribution is obtained by using the variational quantum eigensolver (VQE) technique to compute a contextual correction on a quantum processor. In general the VQE computation of the contextual correction uses fewer qubits and measurements than the VQE computation of the original problem. Varying the number of qubits used for the contextual correction adjusts the quality of the approximation. We simulate CS-VQE on tapered Hamiltonians for small molecules, and find that the number of qubits required to reach chemical accuracy can be reduced by more than a factor of two. The number of terms required to compute the contextual correction can be reduced by more than a factor of ten, without the use of other measurement reduction schemes. This indicates that CS-VQE is a promising approach for eigenvalue computations on noisy intermediate-scale quantum devices.

1 Introduction

The *variational quantum eigensolver* (VQE) is the leading algorithm for quantum simulation on noisy intermediate-scale quantum (NISQ) com-

William M. Kirby: william.kirby@tufts.edu

puters, due to the limited resources it requires in both qubit count and coherence time [1–13]. VQE is a hybrid quantum-classical algorithm in which the expectation value of the Hamiltonian, or other observable, is computed for an ansatz state generated by a parameterized quantum circuit. Optimization of the ansatz parameters is performed iteratively using an optimization algorithm running on a classical computer. VQE algorithms require a large number of measurements to be performed, and give approximate results due to limitations of the ansatz that can be prepared and noise on the quantum device. As a result, the largest experiments to date either do not reach chemical accuracy [6], do not include all Hamiltonian terms [10], or simulate restricted models such as Hartree-Fock [13].

We address these limitations of VQE by providing an approximate simulation method for the full Hamiltonian that can be adjusted to use any amount of available quantum resources. We show that in many cases our method reduces the resources required to reach chemical accuracy. The method is based on the concept of a noncontextual Hamiltonian, in which definite values can be assigned to all Pauli terms simultaneously without contradiction [14]. In [15], we gave a classical algorithm for computing the ground state energies of noncontextual Hamiltonians based on a quasi-quantized model [16, 17]. This raises the question of whether a truly hybrid algorithm can be developed for simulating general Hamiltonians, in which the noncontextual parts of the Hamiltonians are computed classically and contextual quantum corrections are computed using VQE. The method described in this paper is such an algorithm.

The contextual quantum correction is obtained by performing VQE restricted to quantum states that are consistent with the noncontextual ground state. We refer to the space of

such quantum states as the *contextual subspace*. The contextual subspace represents the degrees of freedom that remain after the noncontextual degrees of freedom have been fixed. The resulting quantum computations only involve Hamiltonian terms in the complement of the noncontextual Hamiltonian. We also show how to adjust the noncontextual part of the Hamiltonian, in order to move more of the computation onto the quantum computer, while preserving the structure of the quasi-quantized model. The technique for accomplishing this is related to “subspace-search VQE” [18], in which excited energies are found by restricting the search space to be orthogonal to the (previously approximated) ground state. In our case, we are not looking for excited states, but we are implementing VQE in a restricted search space, and part of the technique for achieving this is similar to that in [18] (see Section 2 for details of the technique).

As an example, suppose we want to apply VQE to a Hamiltonian on n qubits, but the available quantum processor has only q qubits. CS-VQE permits us to adjust the noncontextual approximation method so that the associated quantum correction uses exactly q qubits. Increasing the number of qubits used on the quantum processor monotonically improves the quality of the overall approximation, interpolating between the noncontextual approximation with no quantum correction and full VQE. Thus, we can tune the quantum part to fit the available quantum resources, with the classical method making up the difference.

CS-VQE is the first hybrid quantum-classical algorithm of its kind, where a nonclassicality criterion (in our case, contextuality) is used to isolate the intrinsically quantum part of a quantum algorithm, and the classical remainder of the algorithm is simulated classically. Our prior works [14, 15] defined contextual and noncontextual Hamiltonians, and gave a classical model for noncontextual Hamiltonians, respectively, so the current algorithm is distinct because it is applicable to arbitrary Hamiltonians and because it has a quantum component as well as a classical component.

In the remainder of the introduction, we review the necessary information from [15] about classical simulation of noncontextual Hamiltonians. In Section 2, we describe the quantum cor-

rection procedure. In Section 3, we describe how to implement CS-VQE, and show the results of simulating CS-VQE on Hamiltonians for small molecules. Finally, in Section 4, we summarize our results and discuss their implications for NISQ computing.

1.1 Noncontextual Pauli Hamiltonians

A set of observables is *noncontextual* when it is possible to assign values to them simultaneously without contradiction [19–36]. How can we determine if a set $\mathcal{S}$ of Pauli operators is noncontextual? First remove the subset $\mathcal{Z}$ of terms that commute with the whole of $\mathcal{S}$. The set $\mathcal{S}$ is noncontextual if and only if commutation is an equivalence relation on $\mathcal{S} \setminus \mathcal{Z}$ [14]. The equivalence classes of commutation in $\mathcal{S} \setminus \mathcal{Z}$ are *cliques* C_i for $i = 1, 2, \dots, N$. Operators in the same clique commute, while operators in different cliques anticommute.

Let $\mathcal{S}$ be the set of Pauli terms in a general Hamiltonian H . We divide $\mathcal{S}$ into a noncontextual subset $\mathcal{S}_{\text{nc}}$ and its complement $\mathcal{S}_{\text{c}}$. This induces a decomposition of H into a noncontextual part H_{nc} whose Pauli terms are $\mathcal{S}_{\text{nc}}$, and H_{c} whose Pauli terms are $\mathcal{S}_{\text{c}}$ [15]:

$$H = H_{\text{nc}} + H_{\text{c}}. \quad (1)$$

We also require that $\mathcal{S}_{\text{nc}}$ be *closed under inference within $\mathcal{S}$* :

Definition 1. $\mathcal{S}_{\text{nc}}$ is *closed under inference within $\mathcal{S}$* if any operators in $\mathcal{S}$ whose values can be inferred from the values of operators in $\mathcal{S}_{\text{nc}}$ must be included in $\mathcal{S}_{\text{nc}}$ [15].

Closure under inference is reviewed in detail in Appendix C. The decomposition (1) is the basis of the CS-VQE method. We will obtain an efficient classical description of the eigenspaces of H_{nc} , and use this and H_{c} to quantum compute a correction to the noncontextual approximation to the ground-state energy.

Because all terms in H_{nc} can simultaneously be assigned definite values without contradiction we can introduce a phase space description of its eigenspaces [14, 36]. The phase space points are the possible joint value assignments to a set of observables derived from $\mathcal{S}_{\text{nc}}$, which we describe below. The eigenstates of H_{nc} are probability distributions over this phase space. This is a quasi-quantized model: a classical phase-space model

with an uncertainty relation imposed upon the allowed probability distributions (sometimes called *epistemic states*) on the phase space [16, 17]. We refer the reader to [15] and Appendix C for further general points about quasi-quantized models.

We now describe the states of the model, which we call *noncontextual states*. We first identify a set of observables that define the phase-space points in the model:

$$G \cup \{A_1, A_2, \dots, A_N\}. \quad (2)$$

Each A_i is one element of the corresponding clique C_i , so the A_i pairwise anticommute. G is an independent generating set for the Abelian group $\overline{\mathcal{Z}}$, which includes $\mathcal{Z}$ as well as all products of pairs of operators in the same clique. The phase space points are all assignments of values ± 1 to the observables in the set (2) [15].

The noncontextual states are probability distributions over the phase space points generated by (2). Probability distributions corresponding to valid quantum states must obey an uncertainty relation [16, 17]. A sufficient condition is that the commuting generators $G_j \in G$ take definite values, and that the expectation values of the A_i form a unit vector [15]. This means that each noncontextual state is defined by parameters $(\vec{q}, \vec{r})$ such that

$$\langle G_j \rangle = q_j = \pm 1, \quad \langle A_i \rangle = r_i, \quad |\vec{r}| = 1. \quad (3)$$

In [15], we showed how these expectation values for the set (2) induce expectation values for all terms $\mathcal{S}_{\text{nc}}$ in the noncontextual part H_{nc} of the Hamiltonian; consequently, a noncontextual state induces an expectation value for H_{nc} . We also proved that all expectation values for H_{nc} can be generated in this way. Minimizing this expectation value by varying the noncontextual state $(\vec{q}, \vec{r})$ thus provides a variational estimate of the ground state energy of H_{nc} [15]. We refer to the minimizing assignment $(\vec{q}, \vec{r})$ as the *noncontextual ground state*.

The A_i are anticommuting Pauli operators and $\vec{r}$ is a real unit vector, so the observable

$$\mathcal{A} \equiv \sum_{i=1}^N r_i A_i, \quad (4)$$

is a rotated Pauli operator, and thus has eigenvalues ± 1 . The unitary that maps $\mathcal{A}$ to a single

Pauli operator is a sequence of $N - 1$ rotations generated by Pauli operators, all of which preserve the G_j . If each of the A_i has expectation value r_i then $\mathcal{A}$ has expectation value $+1$, and vice versa [15]. Thus, the noncontextual state with parameters $(\vec{q}, \vec{r})$ is equivalent to a joint value assignment for the set of observables

$$G \cup \{\mathcal{A}\}, \quad (5)$$

where the value assignments are $G_j \mapsto q_j = \pm 1$ for each j and $\mathcal{A} \mapsto +1$. We refer to the observables in (5) as the *noncontextual generators*.

The noncontextual states therefore correspond to subspaces of quantum states that are stabilized by the operators $q_j G_j$ and by $\mathcal{A}$. These are almost stabilizer subspaces in the usual sense (see e.g. [37, Sec. 10.3]), except that $\mathcal{A}$ is not a single Pauli operator, but is unitarily equivalent to one. Therefore, a noncontextual state can be thought of as a stabilizer subspace, one of whose stabilizers has been rotated by an efficiently-describable unitary.

2 Quantum correction

Let $(\vec{q}, \vec{r})$ be the noncontextual ground state. If we take the resulting energy of H_{nc} as a classical estimate of the ground state energy of the full Hamiltonian H , we can obtain a quantum correction by minimizing the energy of the remaining terms in the Hamiltonian over the quantum states that are consistent with the noncontextual ground state. As discussed above, this common eigenspace is a stabilizer subspace up to a rotation on one of the stabilizers. We refer to this subspace as the *contextual subspace*.

Before we discuss how to find quantum corrections, we establish when such corrections can appear:

Theorem 1. *Let $\mathcal{S}$ be a set of Pauli operators, and let $\mathcal{S}_{\text{nc}}$ be a noncontextual subset that is closed under inference within $\mathcal{S}$ (see Definition 1). Then for any noncontextual state $(\vec{q}, \vec{r})$ as in (3) describing $\mathcal{S}_{\text{nc}}$, there exists a quantum state consistent with $(\vec{q}, \vec{r})$ (i.e., that gives the same expectation values for $\mathcal{S}_{\text{nc}}$ as $(\vec{q}, \vec{r})$) for which the expectation value of every operator in $\mathcal{S}_{\text{c}} \equiv \mathcal{S} \setminus \mathcal{S}_{\text{nc}}$ is zero.*

The proof may be found in Appendix A, and follows from the fact that $\mathcal{S}_{\text{nc}}$ is closed under inference: no value of any operator in $\mathcal{S}_{\text{c}}$ can be

inferred from the values of operators in the non-contextual part of the Hamiltonian [15]. Theorem 1 has two useful corollaries.

Corollary 1.1. *If the noncontextual states of H_{nc} uniquely identify quantum states, then for any noncontextual state the expectation value of every term in H_c is zero, i.e., no quantum correction is possible.*

Proof. By Theorem 1, there exists a quantum state $|\psi\rangle$ consistent with the noncontextual state for which the expectation value of every term not in the noncontextual part is zero. Thus if the noncontextual state uniquely identifies a quantum state, it must be $|\psi\rangle$. $\square$

Corollary 1.2. *As an approximation method for a general Hamiltonian, a quasi-quantized model of the noncontextual part is variational, i.e., the resulting approximate ground state energies are lower-bounded by the true ground state energy. It remains variational when the quantum correction is included.*

Proof. From Theorem 1, it follows that there exists a ground state of the noncontextual part for which the expectation value of every other term in the Hamiltonian is zero. Hence, the ground state energy of the noncontextual part is a possible expectation value for the energy of the full Hamiltonian, so it is lower-bounded by the ground state energy of the full Hamiltonian. To compute the quantum correction we minimize the expectation value of the full Hamiltonian over quantum states consistent with a given noncontextual state. This produces a variational estimate of the energy with the contribution from H_{nc} given by the noncontextual state. $\square$

2.1 Mapping a contextual subspace to a stabilizer subspace

We now show how to map the contextual subspace corresponding to the noncontextual ground state to a subspace stabilized by single-qubit Z operators. To achieve this goal we rotate the G_j and subsequently the single operator $\mathcal{A}$ to single-qubit Z operators.

The number of G_j is some $M < n$, as discussed in [15]. Therefore, the G_j can be mapped to single-qubit Z operators by a sequence of at

most $2M \frac{\pi}{2}$ -rotations¹ (for completeness, we provide a constructive proof as Lemma A.2 in Appendix A). These rotations are Clifford operators, so they map the remaining Pauli operators in the Hamiltonian back to single Pauli operators while preserving their commutation relations. Let D denote the composition of these rotations, and let $H' \equiv DHD^\dagger$ be the rotated Hamiltonian. After applying D , the noncontextual generators G_j have been mapped to single-qubit Z operators G'_j . Without loss of generality, since we have already found the noncontextual ground state at this point we can choose the signs of the G'_j such that they all have eigenvalue $+1$ in the noncontextual ground state, and thus stabilize it. We refer to this basis as the “rotated basis.”

Once D has been applied, we map $\mathcal{A}' = DAD^\dagger$ to a single-qubit Z operator as well. $\mathcal{A}'$ is a normalized linear combination of the anticommuting Pauli operators $A'_i = DA_iD^\dagger$, as in (4), so we use the sequence of rotations employed in unitary partitioning [39, 40]. The result is a sequence of $N - 1$ rotations generated by products of pairs of the A'_i ; we denote by R the composition of these rotations, and the result of applying it to $\mathcal{A}'$ is

$$R^\dagger \mathcal{A}' R = A'_1, \quad (6)$$

where A'_1 is a single Pauli operator.

The rotations forming R are generated by products of the A'_i , and the A'_i commute with the operators G'_j , so R commutes with the operators G'_j . Thus A'_1 commutes with and is independent of the operators in G' , so since it is a single Pauli operator, we can use Lemma A.2 to map it to a single-qubit Z operator as well, without disturbing the operators in G' . Let $D_{\mathcal{A}'}$ denote the full rotation that maps $\mathcal{A}'$ to a single-qubit Z operator A'' .

2.2 Restricting the Hamiltonian to a contextual subspace

In the rotated basis, we will restrict the Hamiltonian to the subspace stabilized by the noncontextual generators G'_j . We will then obtain the quantum correction by minimizing the expectation value of this restricted Hamiltonian over $+1$ -eigenvectors of the remaining noncontextual generator $\mathcal{A}'$.

¹See for example the solution to problem 3 in [38].

Let $\mathcal{H}_1$ denote the Hilbert space of the n_1 qubits acted on by the single-qubit Pauli Z operators G'_j , and let $\mathcal{H}_2$ denote the Hilbert space of the remaining n_2 qubits. Thus the full Hilbert space is $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ and the total number of qubits is $n = n_1 + n_2$. The contextual part of the Hamiltonian in the rotated basis is:

$$H'_c = \sum_{P \in \mathcal{S}'_c} h_P P. \quad (7)$$

The set of Pauli terms in H'_c is $\mathcal{S}'_c$ and terms in $\mathcal{S}'_c$ in general act on both of the subspaces $\mathcal{H}_1$ and $\mathcal{H}_2$.

We can write the terms P in (7) as

$$P = P_1 \otimes P_2, \quad (8)$$

where P_1 is a Pauli operator acting on $\mathcal{H}_1$ and P_2 is a Pauli operator acting on $\mathcal{H}_2$. P commutes with an element of G' if and only if $P_1 \otimes \mathbb{1}_{\mathcal{H}_2}$ does (where $\mathbb{1}_{\mathcal{H}_2}$ denotes the identity operator acting on $\mathcal{H}_2$), since the operators in G' act only on $\mathcal{H}_1$. Since the noncontextual state corresponds to a subspace stabilized by G' , if P anticommutes with any element of G' , then its expectation value in the noncontextual state is zero. Hence, any P that admits a quantum correction must commute with all elements of G' , and thus $P_1 \otimes \mathbb{1}_{\mathcal{H}_2}$ must as well. The elements of G' comprise all single-qubit Z operators acting in $\mathcal{H}_1$, so P_1 must be a product of such operators. Hence the expectation value of P_1 is some $p_1 = \pm 1$, determined by the noncontextual state.

Let $|\psi_{(\vec{q}, \vec{r})}\rangle$ be any quantum state consistent with the noncontextual state $(\vec{q}, \vec{r})$. The action of any term P that admits a quantum correction on $|\psi_{(\vec{q}, \vec{r})}\rangle$ has the form

$$\begin{aligned} P|\psi_{(\vec{q}, \vec{r})}\rangle &= (P_1 \otimes P_2)|\psi_{(\vec{q}, \vec{r})}\rangle \\ &= p_1(\mathbb{1}_{\mathcal{H}_1} \otimes P_2)|\psi_{(\vec{q}, \vec{r})}\rangle. \end{aligned} \quad (9)$$

Therefore, if we denote by $H'_c|_{(\vec{q}, \vec{r})}$ the restriction of H'_c to its action on the noncontextual ground state $(\vec{q}, \vec{r})$,

$$\begin{aligned} H'_c|_{(\vec{q}, \vec{r})} &= \sum_{\substack{P \in \mathcal{S}'_c \\ \text{s.t. } [P, G'_i] = 0 \\ \forall G'_i \in G'}} p_1 h_P \mathbb{1}_{\mathcal{H}_1} \otimes P_2 \\ &= \mathbb{1}_{\mathcal{H}_1} \otimes H'_c|_{\mathcal{H}_2} \end{aligned} \quad (10)$$

for

$$H'_c|_{\mathcal{H}_2} \equiv \sum_{\substack{P \in \mathcal{S}'_c \\ \text{s.t. } [P, G'_i] = 0 \\ \forall G'_i \in G'}} p_1 h_P P_2. \quad (11)$$

This is a Hamiltonian on n_2 qubits, for n_2 given by

$$n_2 = n - |G|, \quad (12)$$

where $|G|$ is the number of noncontextual generators G_j . Furthermore, if terms $P \in \mathcal{S}'_c$ are distinct only on their tensor factors P_1 , the remaining operators P_2 in $H'_c|_{\mathcal{H}_2}$ will be identical. Also, any terms that anticommute with any of the noncontextual generators G_j are dropped entirely (since their expectation values are zero). Thus, the restricted Hamiltonian $H'_c|_{\mathcal{H}_2}$ may contain fewer than $|\mathcal{S}'_c|$ terms. This is illustrated in Fig. 3, in Section 3.2.

2.3 Optimizing within a noncontextual subspace

To obtain the quantum correction, we perform n_2 -qubit VQE on the restricted Hamiltonian $H'_c|_{\mathcal{H}_2}$ within the contextual subspace. The contextual subspace is the subspace of $\mathcal{H}_2$ that forms the +1-eigenspace of $\mathcal{A}'$, the remaining noncontextual generator in the rotated basis. To search within the +1-eigenspace of $\mathcal{A}'$, we can prepare ansätze in the +1-eigenspace of $\mathcal{A}'$ (a single-qubit Z operator), and then apply the inverse of $D_{\mathcal{A}'}$. This guarantees that every ansatz state is consistent with the noncontextual ground state $(\vec{q}, \vec{r})$.

Note that we only explicitly restrict the rotated Hamiltonian to the subspace stabilized by G'_j , whereas we restrict to the +1-eigenbasis of $\mathcal{A}'$ at the level of the ansatz. This is because although the operation $D_{\mathcal{A}'}$ that diagonalizes $\mathcal{A}'$ can be efficiently implemented on a quantum computer, it is not a Clifford operation. Conjugating the Hamiltonian with $D_{\mathcal{A}'}$ can increase the number of terms by a factor of $\Theta(2^N)$, where N is the number of cliques. Thus if N is small, we could classically compute the Hamiltonian restricted to the +1 eigenbasis of $\mathcal{A}'$ and then perform VQE on this Hamiltonian. This would save one qubit and permit an unconstrained search for the quantum correction, but since N can in principle scale as $\Theta(n)$ this approach will not be efficient in general.

2.4 Example

As an example, we construct a Hamiltonian for which most of the terms are included in the non-contextual part. Let $\mathcal{S} = \mathcal{S}_{\text{nc}} \cup \mathcal{S}_{\text{c}}$, where

$$\begin{aligned}\mathcal{S}_{\text{nc}} &= \{ZII, IXI, IYI, IZX, IZY, IZZ, \\ &\quad ZXI, ZYI, ZZX, ZZY, ZZZ\}, \\ \mathcal{S}_{\text{c}} &= \{IIX, IYY, IIZ\}.\end{aligned}\quad (13)$$

The set of terms $\mathcal{S}_{\text{nc}}$ is noncontextual, partitioning into $\mathcal{Z} = \{ZII\}$ (recall that $\mathcal{Z}$ is the set of terms that commute with all others), and five cliques, $\{IXI, ZXI\}$, $\{IYI, ZYI\}$, $\{IZX, ZZX\}$, $\{IZY, ZZY\}$, and $\{IZZ, ZZZ\}$. Thus we may choose

$$\begin{aligned}A_1 &= IXI, \quad A_2 = IYI, \quad A_3 = IZX, \\ A_4 &= IZY, \quad A_5 = IZZ.\end{aligned}\quad (14)$$

The extra terms $\mathcal{S}_{\text{c}}$ all commute with $\mathcal{Z}$. In this case, $G = \mathcal{Z}$ since $\mathcal{Z}$ contains only one operator, and this operator is already a single-qubit Z operator, so D is the identity.

Thus $\mathcal{H}_2$ is the Hilbert space of the second two qubits, so for

$$H'_c = H_c = h_{IIX}IIX + h_{IYY}IYY + h_{IIZ}IIZ \quad (15)$$

for some coefficients $h_{IIX}, h_{IYY}, h_{IIZ}$, the restriction to $\mathcal{H}_2$ is

$$H'_c|_{\mathcal{H}_2} = H_c = h_{IIX}IX + h_{IYY}IY + h_{IIZ}IZ. \quad (16)$$

We also have

$$\mathcal{A}' = \mathcal{A} = r_1 A_1 + r_2 A_2 + r_3 A_3 + r_4 A_4 + r_5 A_5 \quad (17)$$

for some unit vector $\vec{r}$; the restriction of $\mathcal{A}'$ to $\mathcal{H}_2$ is thus

$$\mathcal{A}'|_{\mathcal{H}_2} = r_1 XI + r_2 YI + r_3 ZX + r_4 ZY + r_5 ZZ, \quad (18)$$

so $D_{\mathcal{A}'}$ is the rotation that maps this to a single-qubit Z operator, as described in Section 2.1. We can choose

$$D_{\mathcal{A}'} \mathcal{A}'|_{\mathcal{H}_2} D_{\mathcal{A}'}^\dagger = ZI; \quad (19)$$

in this case, for an ansatz we may prepare any state whose value is $|0\rangle$ for the first qubit in $\mathcal{H}_2$, and then apply $D_{\mathcal{A}'}$ to this state.

Thus, we reduce an initial Hamiltonian on three qubits to a noncontextual approximation

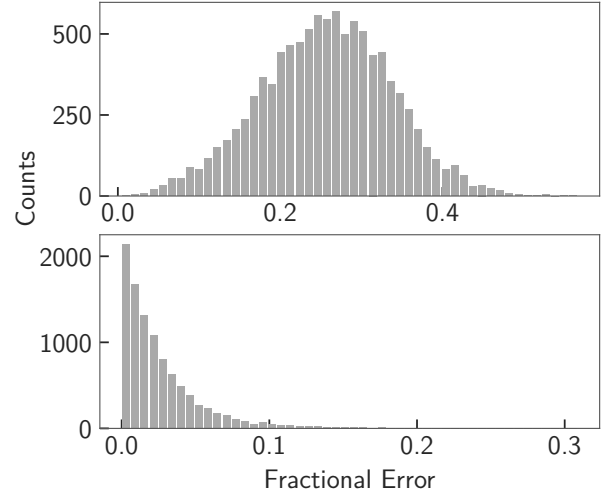


Figure 1: Comparison of fractional errors in the noncontextual approximation of the ground state energy (upper plot), and in the noncontextual approximation with quantum correction (lower plot). The histogram points were generated by 10000 Hamiltonians with terms (13) and uniformly random coefficients in $[-1, 1]$. The mean fractional error without quantum correction is 0.257, and the mean fractional error with quantum correction is 0.0268.

and a quantum correction that may be implemented on a two-qubit quantum processor.

To evaluate the performance of the resulting approximations, we generated 10000 Hamiltonians with the terms (13) by choosing coefficients for them uniformly at random from $[-1, 1]$. The resulting fractional errors in the ground state energies are plotted in Fig. 1; the average fractional error is 0.257 for the noncontextual approximation alone, and 0.0268 when the quantum correction is included. The quantum corrections were simulated classically by directly evaluating the lowest eigenvalues of the Hamiltonians restricted to the noncontextual ground states.

3 Contextual subspace VQE

The quantum correction to noncontextual approximations discussed in Section 2 allows us to use limited quantum resources to improve a classical simulation result. In this section we explain how we can systematically step back from the original noncontextual approximation in order to enlarge the contextual subspace, thus improving the overall accuracy of the approximation by using more quantum resources. This provides a pa-

parameter that can be specified based on the quantum resources available, taking us from the optimal noncontextual approximation at one extreme to full VQE at the other. We call this method *contextual subspace VQE*.

3.1 Method

We begin with a Hamiltonian H whose noncontextual approximation is H_{nc} . As discussed above, the noncontextual ground state corresponds to a joint eigenspace of the noncontextual generators $G \cup \{\mathcal{A}\}$. We can trade accuracy of the noncontextual approximation for an improved quantum correction by decreasing the size of $G \cup \{\mathcal{A}\}$, which increases the dimension of the contextual subspaces. We accomplish this by decreasing the size of G , the set of generators for the commuting part of H_{nc} . Since the number of qubits used in the quantum correction procedure is the total number of qubits minus the number of generators in G (see (12)), reducing the size of G increases the dimension of the search space for the quantum processor.

We work in the rotated basis, as in Section 2. In this basis, we select some subset of the noncontextual generators G'_j , and remove all terms generated by them from the noncontextual part. Since the G'_j are single-qubit Z operators, this means that for each generator to be dropped we remove from the noncontextual part all terms containing the corresponding single-qubit Z operator as a tensor factor. All the terms thus removed should be added to the quantum correction Hamiltonian H'_c (as in Section 2.2). We now implement the quantum correction on this expanded H'_c , keeping the same noncontextual ground state that we began with, but only applying its value assignments to the generators that remain in the noncontextual part.

The new noncontextual approximation on its own will in general be worse than the original noncontextual approximation. However, after including the new quantum correction the overall approximation cannot be worse, and will in general be better. This is because the values assigned in the original noncontextual approximation and quantum correction are still consistent with the noncontextual ground state, so quantum states that obtain those values are included in the new quantum search space. Thus in the worst case the new approximation will only recover the original

approximation. If the new quantum correction is nonzero for any additional terms, the new approximation will be strictly better than the original approximation. In the limit where we remove all terms from the noncontextual part and simulate them on the quantum computer, there will be no noncontextual approximation left, and we will have recovered full VQE.

The additional terms that can have nonzero quantum corrections after the removal procedure are those that anticommute with any of the generators G'_j that were removed, but commute with the remaining generators. These terms were previously restricted to null expectation values only because the noncontextual state was required to be a joint eigenstate of the removed generators, so when that is no longer enforced their expectation values can vary. Therefore, we can choose which subset of the generators to remove based on which will permit the optimal quantum correction.

Note that classically simulating the noncontextual part of the Hamiltonian is NP-complete, so in worst cases the classical simulation part of CS-VQE will not perform well [15]. However, worst case Hamiltonians for standard VQE are QMA-complete, meaning that a similar argument applies to VQE in general. Hence, in both cases we are interested in heuristic performance for specific Hamiltonians of interest, rather than worst cases. Framed in this way, what CS-VQE does is take standard VQE, which is a heuristic for an optimization problem over a set of parameters for a quantum circuit, and transform it into two heuristics (one classical and one quantum) for two smaller optimization problems. In practice, we have found that a combination of Monte-Carlo and gradient descent methods works well for the classical part of the algorithm, but continuing to optimize this is a topic for future work.

3.2 Applications

We tested CS-VQE on a set of electronic structure Hamiltonians in the Jordan-Wigner mapping [41]. In order to distinguish CS-VQE from qubit tapering, we first tapered the Hamiltonians using symmetries as in [42, 43], then implemented CS-VQE in order to remove even more qubits. The initial noncontextual approximation Hamiltonians were chosen via a greedy classical algorithm, as described in [15]. This algorithm runs

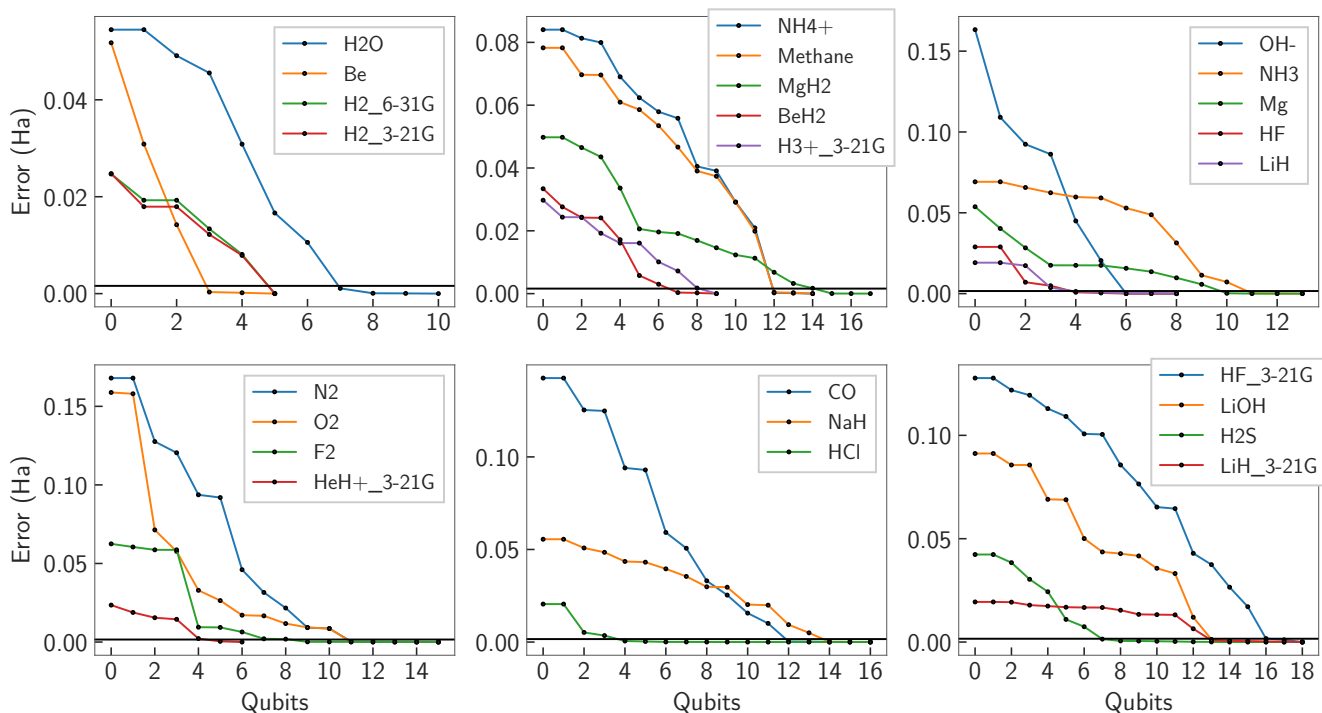


Figure 2: CS-VQE approximation errors versus number of qubits used on the quantum computer, for tapered molecular Hamiltonians. All Hamiltonians whose curves overlap in the region below chemical accuracy have the same total numbers of qubits. The solid black lines indicate chemical accuracy. Within each subplot, the ordering of the legend matches the vertical ordering of the leftmost points in the curves.

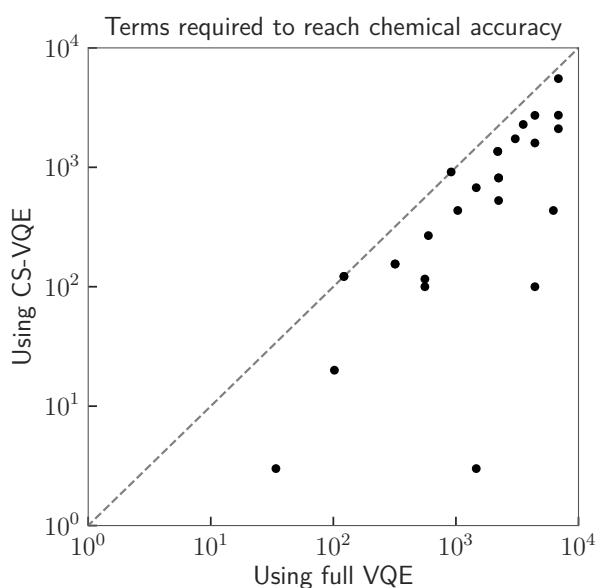


Figure 3: Number of terms simulated on the quantum processor required to reach chemical accuracy using CS-VQE versus using full VQE. The dashed line marks equality. All points represent either one, two, or three Hamiltonians.

in $O(N^5)$ time for an N -term Hamiltonian, since testing a particular Hamiltonian for noncontextuality takes $O(N^3)$ time [14], and a greedy algorithm that adds optimal terms one at a time requires $O(N^2)$ steps. This method is not optimal, but is efficient. The quantum parts of the procedures were simulated classically by directly evaluating the lowest eigenvalues of the quantum correction Hamiltonians. The results are given in Fig. 2, which shows the overall CS-VQE approximation errors versus the number of qubits used on the quantum computer, and in Fig. 3, which shows the number of terms that must be simulated on the quantum computer in order to reach chemical accuracy using CS-VQE. Our code is available on GitHub², and may be used to reproduce our results or to apply CS-VQE to new Hamiltonians of the reader's choosing.

As noted at the end of Section 3.1, CS-VQE is sensitive to the order in which the qubits are moved from the noncontextual approximation to the quantum processor. In the calculations to obtain Figs. 2 and 3, we used a heuristic that begins

²Source code: <https://github.com/wmkirby1/ContextualSubspaceVQE>

with the noncontextual approximation, then adds qubits to the quantum correction two at a time, greedily choosing each pair to maximize the decrease in the ground state energy estimate. This method is informed by the structure of the non-contextual and contextual parts of the molecular Hamiltonians, and performed best out of the heuristics we tried that can be implemented efficiently without performing full VQE. Details of the implementation of the heuristic are given in Appendix B.

This heuristic involves running CS-VQE repeatedly, since for sufficiently large applications one would have to use the quantum processor to compute the quantum corrections on the way to choosing the set of qubits for the final quantum correction. However, these preliminary computations would only be necessary once the number of qubits chosen becomes unfeasible for classical simulation, and from Fig. 3 we see that the number of terms required to reach chemical accuracy can be many times smaller than the number of terms required to implement full VQE. Therefore, even the repeated runs of CS-VQE required for this heuristic can require fewer measurements overall than full VQE, and of course they also require fewer qubits.

Alternatively, one could use a heuristic to determine the order without evaluating energies at all. However, all variants of this that we tried had substantially worse performance than the heuristic discussed above, so we suggest using that heuristic for real applications. We also tested an inefficient "optimal" heuristic that begins from full VQE and moves qubits to the noncontextual approximation one at a time, greedily minimizing the error penalty for each. Actually implementing this heuristic is even more costly than full VQE, but it did identify the optimal qubit orderings in cases small enough for us to find these by brute-force search. The first heuristic discussed above performed nearly as well as the "optimal" heuristic, requiring the same number of qubits to reach chemical accuracy in most cases and only one extra in the remaining few cases. Details of all of the heuristics and their relative performance are discussed in Appendix B.

4 Conclusion

In this paper, we showed how to use a quantum computer to obtain a correction to a non-contextual approximation of a ground state energy. We then showed how to adjust the number of qubits used on the quantum computer in order to increase the accuracy of the hybrid approximation. This method, contextual subspace VQE or CS-VQE, is a true hybrid quantum-classical algorithm, in which the quantum resources used may be set to match whatever resources are available, and the classical approximation algorithm accounts for the remainder. The method is approximate, but variational, as is VQE itself. Exact methods will only achieve approximate results on NISQ devices due to their noisy character. CS-VQE allows the quantum resources used to be increased systematically until the desired precision is achieved, if possible.

Standard VQE is a heuristic algorithm: there are no analytic characterizations of its performance for general Hamiltonians, or even for special classes like electronic structure Hamiltonians, upon scaling the system size. This is also true for CS-VQE: its performance is sensitive to the specific problem to which it is applied. We do not analytically characterize the errors as a function of the number of qubits used on the quantum processor. However, the examples in Section 3.2 illustrate that CS-VQE performs well in many cases of interest going well beyond the scale of VQE implementations to date, so we hope that as the available quantum processors continue to grow, CS-VQE can be used to allow larger systems to be simulated using those processors.

The technique for restricting the quantum correction to the subspace consistent with the non-contextual ground state may appear reminiscent of using qubit tapering to exploit symmetries as described in [42, 43]. However, in CS-VQE the symmetries are intrinsic to the noncontextual ground state, rather than to the Hamiltonian (as in [42, 43]), and are thus under the experimenter's control. We illustrated this point in Section 3.2 by applying CS-VQE to Hamiltonians that were already tapered using the methods of [42, 43]; using CS-VQE we can eliminate additional qubits at will.

In this paper we did not explore how to implement ansätze for the restricted VQE instance used by CS-VQE, instead finding the exact

ground states of the contextual parts. However, standard ansatz classes, like unitary coupled-cluster (UCC) for electronic structure Hamiltonians [44–46], can be transformed into ansätze for CS-VQE by projecting the gates onto the contextual subspace, just as the contextual part of the Hamiltonian is restricted to the contextual subspace. Detailed study of this is a topic for future research.

One concern for standard VQE as well as for CS-VQE is that the ansatz may suffer from the *barren plateau problem* [47–49], where the gradient of the cost function (in this case expected energy) vanishes exponentially with the system size. It is hoped that for standard VQE, using physically-motivated ansätze like UCC may avoid the barren plateau problem, so since we can use projections of the same ansätze for CS-VQE, this same hope transfers to our case. However, even physically-motivated ansätze may be subject to *noise-induced barren plateaus* [50]: to the best of our knowledge, all variational quantum algorithms have the potential to fail in this way, including CS-VQE. Nonlinear optimization and its attendant problems, including barren plateaus, may be avoided by the use of quantum imaginary time evolution (QITE) or similar methods [51, 52]. In our case, QITE could be applied directly to the contextual part of the Hamiltonian.

It is possible that some of the qubit and term reductions we obtained using CS-VQE have explanations in terms of chemistry. However, in such cases CS-VQE identifies and exploits such features using principles that are derived from the foundations of quantum mechanics, and are consequently agnostic any specific, high-level chemistry arguments. Identifying such chemical arguments would illustrate the role contextuality plays in chemistry, which would be of independent interest.

By using CS-VQE it is possible to reach chemical accuracy for ground state energies of numerous small molecules using many fewer qubits than would be required to implement full VQE on the tapered Hamiltonians. The number of terms and thus number of measurements required is also substantially reduced by using CS-VQE, since groups of terms become equivalent under the symmetry imposed by the noncontextual ground state. The number of measurements needed to obtain the quantum correction could

be further reduced by the techniques described in [39, 40, 53–55]. We leave this and other optimizations of the method to future work. Current VQE implementations are limited in both qubit count and number of measurements by the available hardware, so we expect CS-VQE to be of immediate practical value in accessing new molecular simulation applications on NISQ computers.

Acknowledgements

W. M. K. acknowledges support from the National Science Foundation, Grant No. DGE-1842474. P. J. L. acknowledges support from the National Science Foundation, Grant No. PHY-1720395, and from Google Inc. This work was supported by the NSF STAQ project (PHY-1818914).

References

- [1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, *Nature Communications* **5**, 4213 EP (2014).
- [2] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, *Phys. Rev. X* **6**, 031007 (2016).
- [3] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, D. P. Tew, X. Zhou, J. L. O’Brien, and M. G. Thompson, *Science Advances* **4** (2018).
- [4] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, *Physical Review A* **95**, 020501 (2017).
- [5] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O’Brien, and M. G. Thompson, *Phys. Rev. Lett.* **118**, 100503 (2017).

- [6] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Nature* **549**, 242 (2017).
- [7] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, *Phys. Rev. X* **8**, 031022 (2018).
- [8] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski, *Phys. Rev. Lett.* **120**, 210501 (2018).
- [9] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, *Phys. Rev. X* **8**, 011021 (2018).
- [10] Y. Nam, J.-S. Chen, N. C. Pienti, K. Wright, C. Delaney, D. Maslov, K. R. Brown, S. Allen, J. M. Amini, J. Apisdorf, K. M. Beck, A. Blinov, V. Chaplin, M. Chmielewski, C. Collins, S. Debnath, K. M. Hudek, A. M. Ducore, M. Keesan, S. M. Kreikemeier, J. Mizrahi, P. Solomon, M. Williams, J. D. Wong-Campos, D. Moehring, C. Monroe, and J. Kim, *npj Quantum Information* **6**, 33 (2020).
- [11] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, and P. Zoller, *Nature* **569**, 355 (2019).
- [12] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, *Nature* **567**, 491 (2019).
- [13] Google AI Quantum and Collaborators, *Science* **369**, 1084 (2020).
- [14] W. M. Kirby and P. J. Love, *Phys. Rev. Lett.* **123**, 200501 (2019).
- [15] W. M. Kirby and P. J. Love, *Phys. Rev. A* **102**, 032418 (2020).
- [16] R. W. Spekkens, *Phys. Rev. A* **75**, 032110 (2007).
- [17] R. W. Spekkens, “Quasi-quantization: Classical statistical theories with an epistemic restriction,” in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer Netherlands, Dordrecht, 2016) pp. 83–135.
- [18] K. M. Nakanishi, K. Mitarai, and K. Fujii, *Phys. Rev. Research* **1**, 033062 (2019).
- [19] J. S. Bell, *Physics* **1**, 195 (1964).
- [20] J. S. Bell, *Rev. Mod. Phys.* **38**, 447 (1966).
- [21] S. Kochen and E. Specker, *J. Math. Mech.* **17**, 59 (1967).
- [22] R. W. Spekkens, *Phys. Rev. A* **71**, 052108 (2005).
- [23] S. Abramsky and A. Brandenburger, *New Journal of Physics* **13**, 113036 (2011).
- [24] R. Raussendorf, *Phys. Rev. A* **88**, 022322 (2013).
- [25] M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Nature* **510**, 351 EP (2014).
- [26] A. Cabello, S. Severini, and A. Winter, *Phys. Rev. Lett.* **112**, 040401 (2014).
- [27] A. Cabello, M. Kleinmann, and C. Budroni, *Phys. Rev. Lett.* **114**, 250402 (2015).
- [28] R. Ramanathan and P. Horodecki, *Phys. Rev. Lett.* **112**, 040404 (2014).
- [29] N. de Silva, *Phys. Rev. A* **95**, 032108 (2017).
- [30] B. Amaral and M. T. Cunha, arXiv preprint (2017), [arXiv:1709.04812 \[quant-ph\]](https://arxiv.org/abs/1709.04812).
- [31] Z.-P. Xu and A. Cabello, *Phys. Rev. A* **99**, 020103 (2019).
- [32] R. Raussendorf, *Quantum Information and Computation* **19**, 1141 (2019).
- [33] C. Duarte and B. Amaral, *Journal of Mathematical Physics* **59**, 062202 (2018).
- [34] S. Mansfield and E. Kashefi, *Phys. Rev. Lett.* **121**, 230401 (2018).
- [35] C. Okay, E. Tyhurst, and R. Raussendorf, *Quantum Information and Computation* **18**, 1272 (2018).
- [36] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurel, *Phys. Rev. A* **101**, 012350 (2020).
- [37] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2001).
- [38] J. Yan and D. Bacon, arXiv preprint (2012), [arXiv:1203.3906 \[quant-ph\]](https://arxiv.org/abs/1203.3906).
- [39] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, *Journal of Chemical Theory and Computation* **16**, 190 (2020).
- [40] A. Zhao, A. Tranter, W. M. Kirby, S. F. Ung, A. Miyake, and P. J. Love, *Phys. Rev. A* **101**, 062322 (2020).
- [41] P. Jordan and E. Wigner, *Z. Phys.* **47**, 631 (1928).

- [42] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, arXiv preprint (2017), [arXiv:1701.08213 \[quant-ph\]](#).
- [43] K. Setia, R. Chen, J. E. Rice, A. Mezzacapo, M. Pistoia, and J. D. Whitfield, [Journal of Chemical Theory and Computation](#) **16**, 6091 (2020).
- [44] M. H. Yung, J. Casanova, A. Mezzacapo, J. McClean, L. Lamata, A. Aspuru-Guzik, and E. Solano, [Scientific Reports](#) **4**, 3589 EP (2014).
- [45] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, [New Journal of Physics](#) **18**, 023023 (2016).
- [46] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, [Quantum Science and Technology](#) **4**, 014008 (2018).
- [47] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, [Nature Communications](#) **9**, 4812 (2018).
- [48] A. Uvarov and J. Biamonte, [Journal of Physics A: Mathematical and Theoretical](#) (2021).
- [49] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, [Nature Communications](#) **12**, 1791 (2021).
- [50] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, arXiv preprint (2020), [arXiv:2007.14384 \[quant-ph\]](#).
- [51] M. Motta, C. Sun, A. T. K. Tan, M. J. O’Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, [Nature Physics](#) **16**, 205 (2020).
- [52] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, [npj Quantum Information](#) **5**, 75 (2019).
- [53] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, [The Journal of Chemical Physics](#) **152**, 124114 (2020).
- [54] T.-C. Yen, V. Verteletskyi, and A. F. Izmaylov, [Journal of Chemical Theory and Computation](#) **16**, 2400 (2020).
- [55] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, arXiv preprint (2019), [arXiv:1907.13623 \[quant-ph\]](#).

A Proofs

We will use Lemma 1 from [15]:

Lemma A.1 (Lemma 1 in [15]). *Let $P_1, P_2, \dots, P_N$ be an anticommuting set of Pauli operators. For any unit vector $\vec{a} \in \mathbb{R}^N$, the operator $\sum_{i=1}^N a_i P_i$ has eigenvalues ± 1 . From this it follows that for any state, $\sum_{i=1}^N \langle P_i \rangle^2 \leq 1$.*

Theorem 1 *Let $\mathcal{S}$ be a set of Pauli operators, and let $\mathcal{S}_{nc}$ be a noncontextual subset that is closed under inference within $\mathcal{S}$ (see Definition 1). Then for any noncontextual state $(\vec{q}, \vec{r})$ as in (3) describing $\mathcal{S}_{nc}$, there exists a quantum state consistent with $(\vec{q}, \vec{r})$ (i.e., that gives the same expectation values for $\mathcal{S}_{nc}$ as $(\vec{q}, \vec{r})$) for which the expectation value of every operator in $\mathcal{S}_c \equiv \mathcal{S} \setminus \mathcal{S}_{nc}$ is zero.*

Proof. Let $G \cup \{\mathcal{A}\}$ be the independent, commuting set of observables associated to the noncontextual state $(\vec{q}, \vec{r})$ describing $\mathcal{S}_{nc}$ (see (3) and (5)): the values assigned to $G \cup \{\mathcal{A}\}$ in the noncontextual state $(\vec{q}, \vec{r})$ are

$$G_j \mapsto q_j = \pm 1 \quad (20)$$

for each $G_j \in G$, and

$$\mathcal{A} \equiv \sum_{i=1}^N r_i A_i \mapsto +1 \quad (21)$$

(see (4) and the associated discussion). Let P be a Pauli operator in $\mathcal{S} \setminus \mathcal{S}_{nc}$.

Case 1. If P anticommutes with any operator in G , then $\langle P \rangle = 0$, since any quantum state consistent with $(\vec{q}, \vec{r})$ is a simultaneous eigenstate of G .

Case 2. If P commutes with the operators in G and also with the A_i , then:

1. if P is a product of operators in G , then P can be inferred from G , so P must in fact be included in $\mathcal{S}_{nc}$, since by assumption $\mathcal{S}_{nc}$ is closed under inference within $\mathcal{S}$. This follows immediately from Definition 1, the definition of closure under inference with $\mathcal{S}$.
2. if P is not a product of operators in G , then P is unconstrained by the noncontextual state, and may take any expectation value, including zero.

Case 3. Finally, suppose P commutes with the operators in G , but anticommutes with at least one of the A_i . In this case we want to prove that there exists a $+1$ -eigenstate of $\mathcal{A}$ for which $\langle P \rangle = 0$, as follows:

Let I_P be the set of indices such that

$$\begin{aligned} i \in I_P &\Rightarrow [P, A_i] = 0, \\ i \notin I_P &\Rightarrow \{P, A_i\} = 0. \end{aligned} \quad (22)$$

If I_P is empty, then P anticommutes with all of the A_i : thus since

$$\sum_{i=1}^N \langle A_i \rangle^2 = \sum_{i=1}^N r_i^2 = 1 \quad (23)$$

(see (4) and the associated discussion), and

$$\langle P \rangle^2 + \sum_{i=1}^N \langle A_i \rangle^2 \leq 1 \quad (24)$$

(by Lemma A.1), it follows that $\langle P \rangle = 0$.

The remaining case is when I_P is nonempty; there also exist $i \notin I_P$ by assumption. Let

$$K \equiv \sum_{i \in I_P} r_i A_i, \quad L \equiv \sum_{i \notin I_P} r_i A_i; \quad (25)$$

thus

$$\mathcal{A} = K + L \quad (26)$$

and

$$[K, P] = 0, \quad \{L, P\} = 0, \quad \{K, L\} = 0. \quad (27)$$

Since K and L are linear combinations of anti-commuting Pauli operators, their eigenvalues are $\pm k$ and $\pm l$, respectively, where

$$k \equiv \sqrt{\sum_{i \in I_P} r_i^2}, \quad l \equiv \sqrt{\sum_{i \notin I_P} r_i^2}. \quad (28)$$

Therefore,

$$K^2 = k^2 \mathbb{1}, \quad L^2 = l^2 \mathbb{1}, \quad k^2 + l^2 = 1. \quad (29)$$

Since P commutes with K and is a Pauli operator, PK is also an observable with eigenvalues $\pm k$, which commutes with L (since both P and K anticommute with L). Thus, PK commutes with $\mathcal{A}$, so within the $+1$ -eigenspace of $\mathcal{A}$ there exist eigenstates $|\pm\rangle$ of PK with eigenvalues $\pm k$, i.e.,

$$\mathcal{A}|\pm\rangle = |\pm\rangle, \quad PK|\pm\rangle = \pm k|\pm\rangle. \quad (30)$$

Note that since $P \notin \mathcal{S}_{\text{nc}}$, P cannot be written as a product of operators in G with any of the A_i , so both of the states $|\pm\rangle$ are consistent with the noncontextual state $(\vec{q}, \vec{r})$.

The noncontextual state gives us the expectation values of the A_i :

$$\langle \pm | A_i | \pm \rangle = r_i \quad (31)$$

(see (4) and the corresponding discussion). This means that in addition to (30), we have

$$\langle \pm | K | \pm \rangle = \sum_{i \in I_P} r_i \langle \pm | A_i | \pm \rangle = \sum_{i \in I_P} r_i^2 = k^2. \quad (32)$$

Define $|\psi\rangle \equiv \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$; then

$$\begin{aligned} \langle \psi | P | \psi \rangle &= \frac{1}{k^2} \langle \psi | P K^2 | \psi \rangle \quad \text{by (29)} \\ &= \frac{1}{2k^2} (\langle + | + \langle - |) K (P K | + \rangle + P K | - \rangle) \\ &= \frac{1}{2k} (\langle + | + \langle - |) K (| + \rangle - | - \rangle) \quad \text{by (30)} \\ &= \frac{1}{2k} (\langle + | K | + \rangle - \langle + | K | - \rangle + \langle - | K | + \rangle - \langle - | K | - \rangle) \\ &= -\frac{1}{2k} (\langle + | K | - \rangle - \langle - | K | + \rangle) \quad \text{by (32)}. \end{aligned} \quad (33)$$

But since K and P commute, we similarly have

$$\begin{aligned} \langle \psi | P | \psi \rangle &= \frac{1}{k^2} \langle \psi | P K^2 | \psi \rangle \quad \text{by (29)} \\ &= \frac{1}{2k^2} (\langle + | K P + \langle - | K P) K (| + \rangle + | - \rangle) \\ &= \frac{1}{2k} (\langle + | - \langle - |) K (| + \rangle + | - \rangle) \quad \text{by (30)} \\ &= \frac{1}{2k} (\langle + | K | + \rangle + \langle + | K | - \rangle - \langle - | K | + \rangle - \langle - | K | - \rangle) \\ &= \frac{1}{2k} (\langle + | K | - \rangle - \langle - | K | + \rangle) \quad \text{by (32)}. \end{aligned} \quad (34)$$

Together, (33) and (34) imply that

$$\langle \psi | P | \psi \rangle = -\langle \psi | P | \psi \rangle = 0, \quad (35)$$

so since $|\psi\rangle$ is a state in the $+1$ -eigenspace of $\mathcal{A}$, we are done. $\square$

Lemma A.2. *For any set of M independent, commuting Pauli operators, there exists an efficiently calculable unitary rotation U , given by a sequence of at most $2M \frac{\pi}{2}$ -rotations generated by Pauli operators, that maps the set to a set of distinct single-qubit Z operators.*

Proof. Let $\{B_i \mid i = 1, 2, \dots, M\}$ be a commuting set of Pauli operators. We may write B_i as

$$B_i = \bigotimes_{k=1}^N \sigma_k^{(B_i)}, \quad (36)$$

where each $\sigma_k^{(B_i)} \in \{I, X, Y, Z\}$. We consider two cases:

Case 1. Suppose B_i is not diagonal, meaning that there is some k such that $\sigma_k^{(B_i)} \in \{X, Y\}$. Consider the Pauli operator

$$J_i = \bigotimes_{k=1}^N \sigma_k^{(J_i)}, \quad (37)$$

defined as follows: for each k ...

$$\sigma_k^{(B_i)} = I \Rightarrow \sigma_k^{(J_i)} = I, \quad (38)$$

$$\sigma_k^{(B_i)} = Z \Rightarrow \sigma_k^{(J_i)} = Z, \quad (39)$$

$$\sigma_k^{(B_i)} = X \text{ or } Y \Rightarrow \sigma_k^{(J_i)} = X \text{ or } Y, \quad (40)$$

where the values $\sigma_k^{(J_i)} = X$ or Y in (40) are chosen so that $\sigma_k^{(B_i)}$ and $\sigma_k^{(J_i)}$ differ for exactly one k (as we noted above, at least one of the $\sigma_k^{(B_i)}$ is X or Y if B_i is not diagonal.) This guarantees that J_i anticommutes with B_i .

Consider a rotation by $\pi/2$ generated by J_i , i.e.,

$$\exp\left(i\frac{\pi}{4}J_i\right) = \frac{1}{\sqrt{2}}(1 + iJ_i). \quad (41)$$

Upon conjugating B_i by this operator, we obtain

$$\begin{aligned} & \frac{1}{2}(1 + iJ_i)B_i(1 - iJ_i) \\ &= \frac{1}{2}(B_i - iB_iJ_i + iJ_iB_i + J_iB_iJ_i) \\ &= \frac{1}{2}(B_i + 2iJ_iB_i - J_iJ_iB_i) \\ &= iJ_iB_i \\ &= i \bigotimes_{k=1}^N \sigma_k^{(J_i)} \sigma_k^{(B_i)}, \end{aligned} \quad (42)$$

where the third line follows because J_i anticommutes with B_i , and the fourth line follows because J_i is self-inverse. By the conditions on the $\sigma_k^{(J_i)}$, we see that $\sigma_k^{(J_i)} \sigma_k^{(B_i)} = I$ or $\pm iZ$ for each k , and $\pm iZ$ appears exactly once, so (42) becomes

$$\frac{1}{2}(1 + iJ_i)B_i(1 - iJ_i) = \pm \bigotimes_{k=1}^N \sigma_k^{(D_i)}, \quad (43)$$

where all $\sigma_k^{(D_i)} = I$ except one, which is Z . In other words, the rotation about J_i has mapped B_i to a single-qubit Z operator, as desired.

In each step, we apply the rotation $\exp(i\frac{\pi}{4}J_i)$ to all operators in the set. Thus we might worry that, having already mapped some subset of the $B_{i'}$ to single-qubit Z operators $D_{i'}$, applying some later rotation $\exp(i\frac{\pi}{4}J_i)$ to map B_i to a single-qubit Z operator could change the previously obtained $D_{i'}$. This turns out not to be the case, as we now show:

Consider some particular one of the $D_{i'}$, whose expansion as a tensor product is

$$D_{i'} = \bigotimes_{k=1}^N \sigma_k^{(D_{i'})} \quad (44)$$

where one of the $\sigma_k^{(D_{i'})}$ is Z and the others are I . $D_{i'}$ commutes with B_i , since the previously applied rotations preserve commutation relations, so for all values of m such that $\sigma_m^{(D_{i'})} = Z$, $\sigma_m^{(B_i)}$ (as defined in (36)) must be I or Z . But this implies that J_i also commutes with $D_{i'}$, since we know that $\sigma_m^{(J_i)}$ is I or Z exactly when $\sigma_m^{(B_i)}$ is I or Z , and thus $\sigma_m^{(J_i)}$ is I or Z .

Therefore, the rotation that maps B_i to a single-qubit Z preserves the previously obtained $D_{i'}$.

Case 2. Suppose B_i is diagonal, so $\sigma_k^{(B_i)} \in \{I, Z\}$ for all k . Since any previously-obtained $D_{i'}$ are single-qubit Z operators and we assumed that the entire set is independent, B_i cannot be the product of any subset of the previously-obtained $D_{i'}$. Therefore, there must exist some $m \in \{1, 2, \dots, n\}$ such that

$$\sigma_m^{(B_i)} = Z \quad (45)$$

and

$$\sigma_m^{(D_{i'})} = I \quad (46)$$

for all of the previously-obtained $D_{i'}$.

Apply the rotation $\exp(i\frac{\pi}{4}K_i)$, for K_i defined by

$$K_i = \bigotimes_{k=1}^N \sigma_k^{(K_i)}, \quad (47)$$

where

$$\sigma_m^{(K_i)} = Y \quad (48)$$

and

$$\sigma_k^{(K_i)} = I \quad (49)$$

for all $k \neq m$. Thus $\exp(i\frac{\pi}{4}K_i)$ commutes with and therefore does not change any previously-obtained D_i .

As in Case 1, applying $\exp(i\frac{\pi}{4}K_i)$ to B_i obtains

$$\exp\left(i\frac{\pi}{4}K_i\right) B_i \exp\left(-i\frac{\pi}{4}K_i\right) = i \bigotimes_{k=1}^N \sigma_k^{(K_i)} \sigma_k^{(B_i)}, \quad (50)$$

where by construction,

$$\sigma_m^{(K_i)} \sigma_m^{(B_i)} = iX \quad (51)$$

and

$$\sigma_k^{(K_i)} \sigma_k^{(B_i)} = \sigma_k^{(B_i)} \quad (52)$$

for all $k \neq m$. In other words, the rotation has changed the Z at the m th spot in B_i into an X , and left B_i otherwise unchanged. We also apply this rotation to all other operators in the set, which does not change those that have already been mapped to single-qubit Z operators, as we noted above.

Now B_i is no longer diagonal, so we proceed as described in Step 1 above, applying a second Pauli $\frac{\pi}{2}$ -rotation to map B_i to a single-qubit Pauli operator. $\square$

B CS-VQE implementation details

B.1 Moving qubits from the noncontextual approximation to the quantum correction

We describe in detail how the noncontextual approximation is truncated to make room for improved quantum corrections. As in Section 2, we work in the rotated basis, denote by $\mathcal{H}_1$ the subspace of n_1 qubits acted upon by the noncontextual generators G' , and denote by $\mathcal{H}_2$ the subspace of the n_2 remaining qubits, which are used to implement the quantum correction. Let $\mathcal{I}_2$ be the set of indices of these latter n_2 qubits (those not acted upon by G'), whose Hilbert space is the quantum search space $\mathcal{H}_2$. To increase the size of $\mathcal{H}_2$, we first need to select some subset of G' that we want to remove from H_{nc} . Since the elements of G' are single-qubit Z operators, this subset defines a set $\mathcal{I}_{\text{add}}$ of indices for qubits whose states are initially fixed by the noncontextual state, but that we will switch to simulating on the quantum processor. To begin with, from $\mathcal{S}'_{\text{nc}}$ (the terms in the noncontextual part of the Hamiltonian, in the rotated basis) we remove all terms that act on

any of the qubits in $\mathcal{I}_{\text{add}}$, including the elements of G' that act on these qubits. The remaining elements of G' form a new generating set $G'' \subset G'$ satisfying

$$|G''| = |G'| - |\mathcal{I}_{\text{add}}|. \quad (53)$$

Let the new noncontextual set of terms be denoted $\mathcal{S}''_{\text{nc}}$, and let

$$H''_{\text{nc}} \equiv \sum_{P \in \mathcal{S}''_{\text{nc}}} h_P P. \quad (54)$$

All terms that were removed from $\mathcal{S}'_{\text{nc}}$ to obtain $\mathcal{S}''_{\text{nc}}$ should be added to $\mathcal{S}'_c$ to obtain an expanded contextual set of terms $\mathcal{S}''_c$, whose corresponding Hamiltonian is

$$H''_c \equiv \sum_{P \in \mathcal{S}''_c} h_P P. \quad (55)$$

We can see that this removal operation preserves closure under inference of $\mathcal{S}''_{\text{nc}}$ by again thinking of $G' \cup \mathcal{A}'$ as stabilizers (up to some signs) for the contextual subspace. The elements of G' are single Pauli operators that are generators for the stabilizer group of the subspace. In order to increase the dimension of the stabilized subspace, therefore, for each element G'_i of G' that we remove we must also remove all elements of the stabilizer group that include G'_i as a factor. In other words, in this instance preserving closure under inference is equivalent to preserving closure of a stabilizer group.

We can now implement the quantum correction on H''_{nc} and H''_c , keeping the same noncontextual ground state that we began with, but only applying its value assignments to terms in H''_{nc} . Let n''_2 denote the new number of qubits used in the quantum correction procedure: then by (53),

$$n''_2 = n - |G''| = n - |G'| + |\mathcal{I}_{\text{add}}| = n_2 + |\mathcal{I}_{\text{add}}|, \quad (56)$$

where n_2 was the initial number of qubits used in the quantum correction procedure.

B.2 Details of heuristics

The heuristic described in the main text starts from the pure noncontextual approximation and moves qubits two at a time to the quantum correction search space, greedily maximizing the improvement to the error at each step. We guessed that this heuristic might perform well for the following reasons. For the molecular Hamiltonians

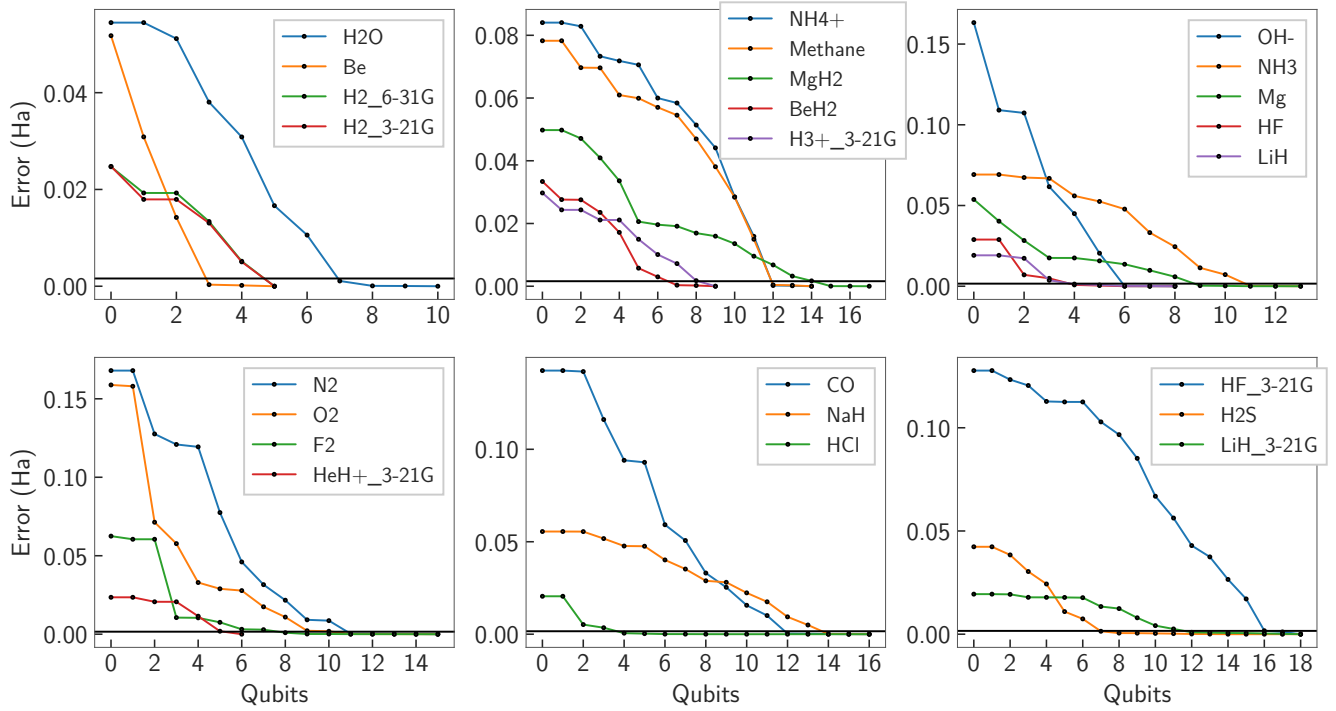


Figure 4: CS-VQE approximation errors versus number of qubits used on the quantum computer, for tapered molecular Hamiltonians, using the qubit ordering chosen by the optimal heuristic. All Hamiltonians whose curves overlap in the region below chemical accuracy have the same total numbers of qubits. The solid black lines indicate chemical accuracy. Within each subplot, the ordering of the legend matches the vertical ordering of the leftmost points in the curves.

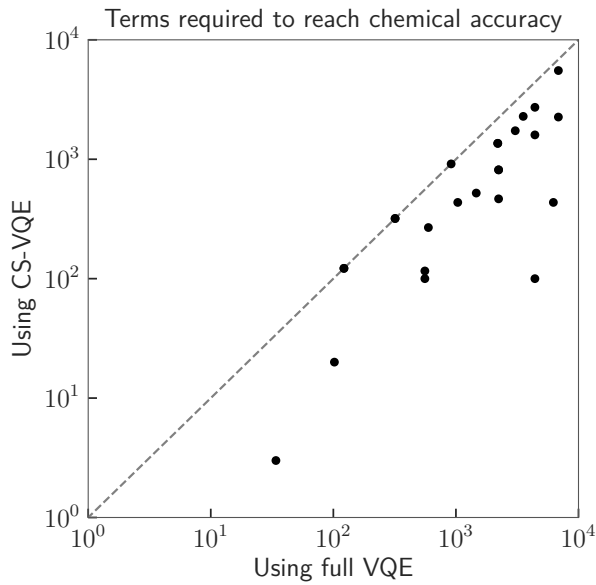


Figure 5: Number of terms simulated on the quantum processor required to reach chemical accuracy using CS-VQE versus using full VQE, for qubit ordering chosen by the optimal heuristic. The dashed line marks equality. All points represent either one, two, or three Hamiltonians.

we tested, the noncontextual Hamiltonians chosen via the greedy heuristic discussed in [15] contain the diagonal terms in the full Hamiltonian (i.e., tensor products of combinations of Pauli Z and single-qubit identity), together with a single clique containing some off-diagonal terms. In particular, this means that the generating set G comprises single-qubit Z operators in the original basis, so the rotated basis is identical to the original basis. The highest weight terms that are not included in the noncontextual part are those containing only one or two off-diagonal Pauli tensor factors.

As discussed in the main text, terms in the quantum correction Hamiltonian are “freed” to be optimized when the generators in the noncontextual part that they anticommute with are dropped. In the present case, these generators are the single-qubit Z operators that act on the qubits for which the terms in the quantum correction Hamiltonian are off-diagonal. Thus, greedily dropping pairs of generators allows a new subset of the highest weight terms in the quantum correction Hamiltonian to be freed for optimization

at each step in the heuristic.

For other heuristics, we refer to the one that led to the best errors of any we tested as the optimal heuristic. The optimal heuristic begins with full VQE, then adds qubits to the noncontextual approximation one at a time while greedily minimizing the error penalty at each step. This heuristic is consequently as hard as performing full VQE, so using it in practical implementations would negate the value of CS-VQE. However, it is informative because it provides a good approximation to the optimal orderings and consequently to the ideal performance of CS-VQE.

For comparison to Fig. 2 and Fig. 3, the errors versus qubits and terms to reach chemical accuracy figures for the heuristic in the main text, we include here the corresponding figures for the optimal heuristic, as Fig. 4 and Fig. 5. Notably, the heuristic included in the main text matches the number of qubits required to reach chemical accuracy using the optimal heuristic in all cases except F_2 , LiH in the 3-21G basis, and Mg: in these cases the heuristic in the main text requires one more qubit than the optimal heuristic. In fact, for HeH+ in the 3-21G basis, the heuristic in the main text requires one fewer qubit to reach chemical accuracy than the optimal heuristic.

As an alternative to heuristics that calculate actual energies, we tested two heuristics based on the total weight of the Pauli terms associated to each qubit. Both starting from the full VQE end (greedily minimizing the penalty for each qubit removed) and starting from the noncontextual end (greedily maximizing the improvement for each qubit added) had identical performance for the examples we tested, but unfortunately that performance was substantially worse than the performance of the heuristic discussed in the main text (as well as the optimal heuristic).

C Noncontextual Hamiltonians

As noted in the main text, a set of observables is *noncontextual* when it admits consistent joint valuations. The kind of contradiction that might prevent such a joint valuation is closely related to the notion of *inference*, which we can introduce as follows. If a pair of observables A, B commute, then they can be measured simultaneously together with their product AB . Thus, if we attempt to construct a classical, ontological model

for some set of observables including A and B , in any assignment of values, the value assigned to AB must be the product of the values assigned to A and B . This is based on the fact that if an observer measures A , B , and AB , the values they obtain will always be consistent with the product relation. Hence, we say that given an assignment of values to A and B , we may *infer* the value assignment to AB .

Definition 2 (see [14]). Given an arbitrary set $\mathcal{S}$ of Pauli operators, the *closure under inference* $\overline{\mathcal{S}}$ of $\mathcal{S}$ is the minimal set of Pauli operators, containing $\mathcal{S}$ as a subset, such that for every commuting pair A, B in $\overline{\mathcal{S}}$, AB is also in $\overline{\mathcal{S}}$.

Definition 3 (see [14]). A set $\mathcal{S}$ of Pauli operators is *noncontextual* if it is possible to assign values (± 1) to $\overline{\mathcal{S}}$ that respect all inference relations in $\overline{\mathcal{S}}$, i.e., such that for every commuting pair $A, B \in \overline{\mathcal{S}}$, the value assigned to AB is the product of the values assigned to A, B .

A set of Pauli operators $\mathcal{S}_{nc}$ is noncontextual if and only if it has the form

$$\mathcal{S}_{nc} = \mathcal{Z} \cup C_1 \cup C_2 \cup \dots \cup C_N, \quad (57)$$

where $\mathcal{Z}$ is the subset of $\mathcal{S}_{nc}$ containing all operators in $\mathcal{S}_{nc}$ that commute with all other operators in $\mathcal{S}_{nc}$, operators in the same C_i commute, and operators in different C_i anticommute [14]. In other words, $\mathcal{S}_{nc}$ is noncontextual if and only if commutation is transitive on $\mathcal{S}_{nc} \setminus \mathcal{Z}$. If commutation is transitive on a set, then because it is always reflexive and symmetric, it is an equivalence relation: as noted in the main text, we label its equivalence classes C_i , and also refer to these as cliques.

As in the main text, we partition a general Hamiltonian H into a noncontextual part H_{nc} and the remaining terms H_c , where the associated sets of Pauli operators are $\mathcal{S}$, $\mathcal{S}_{nc}$, and $\mathcal{S}_c$, respectively. $\mathcal{S}_{nc}$ must be closed under inference within $\mathcal{S}$, which we may now define rigorously in terms of Definition 2: $\mathcal{S}_{nc}$ is closed under inference within $\mathcal{S}$ if and only if

$$\mathcal{S}_{nc} = \overline{\mathcal{S}_{nc}} \cap \mathcal{S}. \quad (58)$$

This is simply a formalization of Definition 1 in the main text. Subject to these constraints, we can choose H_{nc} in any way we like.

The key step in building a quasi-quantized model for the noncontextual part H_{nc} is construction of $\mathcal{R}$, a new set of Pauli operators such that $\overline{\mathcal{R}} = \overline{\mathcal{S}_{\text{nc}}}$ (so that value assignments to $\mathcal{R}$ induce value assignments to $\mathcal{S}_{\text{nc}}$ by inference), and $\mathcal{R}$ is independent:

Definition 4 (see [15]). A set $\mathcal{R}$ of Pauli operators is *independent* if no value of any operator A in $\mathcal{R}$ can be inferred from any value assignment to a subset of $\mathcal{R}$ not containing A . Equivalently, $\mathcal{R}$ is independent if and only if for every commuting subset of $\mathcal{R}$, its product is not in $\mathcal{R}$.

Note that in a commuting set of Pauli operators, this notion of independence reduces to the usual definition of independence of subsets of an Abelian group [15]. Requiring $\mathcal{R}$ to be independent means that not only are *some* value assignments to $\mathcal{R}$ allowed (since $\mathcal{R}$ is noncontextual), but in fact *all* value assignments to $\mathcal{R}$ are allowed (which is not true for a general noncontextual set) [15].

The independent set $\mathcal{R}$ is given by

$$\mathcal{R} = G \cup \{A_i \mid i = 1, 2, \dots, N\} \quad (59)$$

where each $A_i \in C_i$, and G is an independent generating set for

$$\mathcal{Z} \cup \bigcup_{i=1}^N \{AA' \mid A, A' \in C_i\}. \quad (60)$$

Note that since the set (60) is composed of elements of $\mathcal{Z}$ and products of pairs of elements in the same clique, its elements commute with all elements of $\mathcal{S}_{\text{nc}}$. All elements of G therefore commute with all of the A_i , although the A_i pairwise anticommute (since each is an element of the corresponding C_i). As noted in (5) in the main text, states of the quasi-quantized model turn out to be equivalent to joint knowledge of the set of commuting observables

$$G \cup \{\mathcal{A}\}, \quad (61)$$

where the Pauli operators $G_j \in G$ have values $q_j = \pm 1$, and for some unit vector $\vec{r}$ the operator

$$\mathcal{A} \equiv \sum_{i=1}^N r_i A_i \quad (62)$$

has value $+1$ (and consequently each operator A_i has expectation value r_i). Since the Pauli operators A_i anticommute and $|\vec{r}| = 1$, $\mathcal{A}$ has eigenvalues ± 1 , as shown in [15] (see also Lemma A.1).

The states of the quasi-quantized model are thus parametrized by the values $(\vec{q}, \vec{r})$, so we call such a parameter set a *noncontextual state* (which is the same as an *epistemic state*, as in [15, 17]).

The resulting expression for the expectation value of H_{nc} is

$$\langle H_{\text{nc}} \rangle_{(\vec{q}, \vec{r})} = \sum_{B \in \overline{G}} \left(h_B + \sum_{i=1}^N h_{B,i} r_i \right) \prod_{j \in \mathcal{J}_B} q_j, \quad (63)$$

where the classical state parameters $(\vec{q}, \vec{r})$ can take any values such that $q_j = \pm 1$ for each j and $|\vec{r}| = 1$ [15]. The constants in (63), h_B and $h_{B,i}$, are the coefficients in the original Hamiltonian H_{nc} (under an efficiently classically calculable relabeling), and for each $B \in \overline{G}$, $\mathcal{J}_B$ is the set of indices such that

$$B = \prod_{j \in \mathcal{J}_B} G_j, \quad (64)$$

which is also efficiently classically calculable. Thus (63) expresses the expectation value of the noncontextual part of the Hamiltonian as a classical objective function of the parameters $(\vec{q}, \vec{r})$, which may be both obtained and evaluated classically efficiently [15].

Given the objective function (63), estimating the ground state energy of the noncontextual part of the Hamiltonian requires minimizing (63) over the parameters $(\vec{q}, \vec{r})$. For an Hamiltonian of n qubits, the total dimension of $(\vec{q}, \vec{r})$ is at most $2n + 1$. As noted in the main text, we refer to the setting $(\vec{q}, \vec{r})$ that minimizes (63) as the *noncontextual ground state*.