Supplemental Material
Unitary Partitioning and the Contextual Subspace Variational Quantum Eigensolver

Alexis Ralli,[1,*] Tim Weaving,[1,†] Andrew Tranter,[2,3,‡] William M. Kirby,[2,§] Peter J. Love,[2,4,¶] and Peter V. Coveney[1,5,**]

[1] *Centre for Computational Science, Department of Chemistry, University College London, WC1H 0AJ, United Kingdom*
[2] *Department of Physics and Astronomy, Tufts University, Medford, MA 02155, USA*
[3] *Cambridge Quantum Computing, 9a Bridge Street Cambridge, CB2 1UB, United Kingdom*
[4] *Computational Science Initiative, Brookhaven National Laboratory, Upton, NY 11973, USA*
[5] *Informatics Institute, University of Amsterdam, Amsterdam, 1098 XH, Netherlands*

(Dated: January 20, 2023)

CONTENTS

I. CONTEXTUAL-SUBSPACE VQE OVERVIEW

In the following subsections we summarise all the details required to implement the CS-VQE algorithm, where a problem Hamiltonian is split into a contextual and noncontextual part [1]:

$$H_{\text{full}} = H_{\text{con}} + H_{\text{noncon}}. \tag{S.1}$$

A. Noncontextual Part

1. *Testing for contexuality*

Let $\mathcal{S}^{H_{\text{full}}}$ be the set of Pauli operators, in the full system Hamiltonian, requiring measurement in a VQE experiment. It was shown in [2, 3], that a set of four Pauli operators $\{A, B, C, D\}$ is strongly contextual if A commutes with both B and C, but B anticommutes with C. If this condition is not present, then the set is noncontextual.

Given an arbitrary set of Pauli operators $\mathcal{P}$, this gives an algorithm to check for contextuality [3]. A pseudo algorithm is given in algorithm 1. First an $\mathcal{O}(|\mathcal{P}|^2)$ routine is used to remove completely commuting operators $\mathcal{Z}$, leaving the remaining set $\mathcal{T}$. Then a procedure taking $\mathcal{O}(|\mathcal{T}|^3)$ steps is used to determine whether $\mathcal{P}$ is contextual. This check for contextuality is implemented in OpenFermion [4].

* alexis.ralli.18@ucl.ac.uk
† timothy.weaving.20@ucl.ac.uk
‡ tufts@atranter.net
§ william.kirby@tufts.edu
¶ peter.love@tufts.edu
** p.v.coveney@ucl.ac.uk

Algorithm 1 Test for strong contextuality in a given set of Pauli operators [3]

Input: $\mathcal{P} = \{P_0, P_1, P_2, ...\}$ ▷ Input $\mathcal{P}$ is a set of Pauli operators.
Output: *contextual* $(True/False)$ ▷ Whether the set $\mathcal{P}$ is strongly contextual.

$\mathcal{Z} \leftarrow \{\}$
$\mathcal{T} \leftarrow \{\}$
$contextual \leftarrow False$

for $i = 0$ to $|\mathcal{P}| - 1$ **do**
  if $[P_i, P_j] = 0\ \forall j \neq i$ where $j = 0$ to $|\mathcal{P}| - 1$ **then**
    $\mathcal{Z} \leftarrow \mathcal{Z} \cup \{P_i\}$
  else
    $\mathcal{T} \leftarrow \mathcal{T} \cup \{P_i\}$
  end if
end for

for $i = 0$ to $|\mathcal{T}| - 3$ **do**
  for $j = i + 1$ to $|\mathcal{T}| - 2$ **do**
    for $k = j + 1$ to $|\mathcal{T}| - 1$ **do**

      if $[P_i, P_j] = 0$, $[P_i, P_k] = 0$ and $\{P_j, P_k\} = 0$ **then**
        $contextual \leftarrow True$
        return *contextual*
      else
        continue
      end if

    end for
  end for
end for
return *contextual*

2. Obtaining Noncontexual Hamiltonian

To begin CS-VQE, we first need to define the contextual and noncontextual parts. The task of finding the largest noncontextual subset of Pauli operators in $\mathcal{S}^{H_{\text{full}}}$ is a generalization of the disjoint cliques problem [5, 6], which is NP-complete. However, different heuristics can be used to approximately solve this problem.

To date, VQE experiments have mainly focused on chemistry Hamiltonians, where Hartree-Fock accounts for most of the energy. Such Hamiltonians contain Pauli operators that l_1 norms are dominated by diagonal terms - Pauli operators made up of tensor products of single qubit I and Pauli Z matrices. To find a noncontextual set in such a scenario, a greedy heuristic selecting high weight terms from the full Hamiltonian first can be used, while checking the set remains noncontextual using algorithm 1. This gives a noncontextual set containing mainly diagonal terms, with some additional operators [6]. Alternative procedures to find the largest noncontextual subsets remain an open question for the CS-VQE algorithm.

3. Noncontextual hidden variable model

Once the noncontextual Hamiltonian H_{noncon} is determined, we can define the set $\mathcal{S}^{H_{\text{noncon}}}$ to be the Pauli operators in H_{noncon}. We split $\mathcal{S}^{H_{\text{noncon}}}$ into two subsets $\mathcal{Z}$ and $\mathcal{T}$ - representing the set of universally commuting Pauli operators $\mathcal{Z}$ and their complement respectively [1, 6]:

$$\mathcal{S}^{H_{\text{noncon}}} = \mathcal{Z} \cup \mathcal{T} = \left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{Z}}}^{|\mathcal{Z}|-1} P_i \right\} \cup \left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{T}}}^{|\mathcal{T}|-1} P_i \right\}. \tag{S.2}$$

Slight modifications to Algorithm 1 achieve this - where $\mathcal{P}$ would be set to be $\mathcal{S}^{H_{\text{noncon}}}$ and both $\mathcal{Z}, \mathcal{T}$ should be returned.

The operators in $\mathcal{Z}$ are noncontextual, as by definition they are universally commuting and represent symmetries of H_{noncon}. For the overall super-set $\mathcal{S}^{H_{\text{noncon}}}$ to be noncontextual, the remaining operators in $\mathcal{T}$ must be made up of N disjoint cliques C_j [6], where operators within a clique must all commute with each other and operators between cliques pairwise anticommute. This is because commutation forms an equivalence relation on $\mathcal{T}$ if and only if $\mathcal{S}^{H_{\text{noncon}}}$ is noncontextual [1, 2, 6]. We can write $\mathcal{T}$ as:

$$\mathcal{T} = \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{T}}}^{|\mathcal{T}|-1} P_i = \bigcup_{j=0}^{N-1} C_j = \bigcup_{j=0}^{N-1} \Big(\bigcup_{\substack{k=0 \\ \text{where} \\ [P_k, P_l]=0 \\ \forall P_k, P_l \in C_j}}^{|C_j|-1} P_k^{(j)} \Big). \tag{S.3}$$

We re-define each clique C_j using the identity operation defined by the first operator of the jth clique, $P_0^{(j)} P_0^{(j)} = \mathcal{I}$, which represents of the first operator in each of the N cliques. We write the jth clique as [6]:

$$C_j = \bigcup_{\forall P_k \in C_j} P_k^{(j)} = \bigcup_{\forall P_k \in C_j} P_k^{(j)} P_0^{(j)} P_0^{(j)} = \bigcup_{\forall P_k \in C_j} \Big(P_k^{(j)} P_0^{(j)} \Big) P_0^{(j)} = \bigcup_{k=0}^{|C_j|-1} A_k^{(j)} P_0^{(j)} \tag{S.4}$$

The new operators $A_k^{(j)} = P_k^{(j)} P_0^{(j)}$ are just Pauli operators up to a complex phase. The new operators $A_k^{(j)}$ must still commute with the universally commuting operators in $\mathcal{Z}$, but now must also commute with all the other terms in the $N-1$ cliques C_j [6]. Using this, the noncontextual set (Equation S.2) can be rewritten as:

$$\mathcal{S}^{H_{\text{noncon}}} = \left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{Z}}}^{|\mathcal{Z}|-1} P_i \right\} \cup \left\{ \bigcup_{j=0}^{N-1} C_j \right\} \tag{S.5a}$$

$$= \underbrace{\left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{Z}}}^{|\mathcal{Z}|-1} P_i \right\}}_{\mathcal{Z}} \cup \underbrace{\left\{ \bigcup_{j=0}^{N-1} \Big(\bigcup_{\substack{k=0 \\ \text{where} \\ [P_k, P_l]=0 \\ \forall P_k, P_l \in C_j}}^{|C_j|-1} A_k^{(j)} P_0^{(j)} \Big) \right\}}_{\mathcal{T}}. \tag{S.5b}$$

So far we have considered the noncontextual set of Pauli operators $\mathcal{S}^{H_{\text{noncon}}}$, which in general will be a dependent set. By this we mean that some operators in the set can be written as a product of other commuting operators in the set. We need to reduce this set $\mathcal{S}^{H_{\text{noncon}}}$ to an independent set of Pauli operators, where all operators in the noncontextual Hamiltonian can be inferred from the values of other operators in the set under the Jordan product.

To obtain an independent set from $\mathcal{S}^{H_{\text{noncon}}}$, we first take the completely commuting Pauli operators:

$$\begin{aligned} \mathcal{G}' &\equiv \mathcal{Z} \cup \left\{ \bigcup_{j=0}^{N-1} \{ A_k^{(j)} | k = 1, 2, \ldots, |C_j| - 1 \} \right\} \\ &\equiv \left\{ \bigcup_{\substack{i=0 \\ \forall P_i \in \mathcal{Z}}}^{|\mathcal{Z}|-1} P_i \right\} \cup \left\{ \bigcup_{j=0}^{N-1} \{ A_k^{(j)} | k = 1, 2, \ldots, |C_j| - 1 \} \right\}, \end{aligned} \tag{S.6}$$

and using the procedure in [7] find an independent subset $\mathcal{G}$:

$$\mathcal{G} \equiv \{ P_i | i = 0, 1, \ldots, |\mathcal{G}| - 1 \}. \tag{S.7}$$

Appendix C in [6] gives all steps required.

Finally, we need to consider the N pairwise anticommuting $P_0^{(j)}$ operators defined by the N anticommuting cliques. As the operators in $\mathcal{G}$ universally commute with all operators in the noncontextual Hamiltonian, each operator in the

set $\{P_0^{(j)}|j=0,1,\ldots,(N-1)\}$ must be independent of $\mathcal{G}$ under the Jordan product [6]. Combining these results, we get the set $\mathcal{R}$:

$$\mathcal{R} \equiv \{P_0^{(j)}|j=0,1,\ldots,N-1\} \cup \mathcal{G} \tag{S.8}$$

Inspecting the properties of $\mathcal{R}$, one can bound its size. The set $\mathcal{G}$ has size at most $n-1$, as n independent commuting Pauli operators form a complete commuting set of observables for n qubits. In other words, as $\mathcal{G}$ is a universally commuting set, if its size was n (or more) then taking $\mathcal{G}$ and one operator $P_0^{(j)}$ (the set $\mathcal{G} \cup \{P_0^{(j)}\}$) is also a fully commuting set - and would be a commuting set of size $n+1$ (or more) [6]. The maximum number of independent anticomuting operators on n qubits was shown in [2] to be $2n+1$. This actually bounds the size of $\mathcal{R}$, which occurs when the set $\mathcal{G}$ (and thus $\mathcal{Z}$) is empty [6].

Looking at the noncontextual set of Pauli operators Equation S.5, making up the H_{noncon}, we see that the subset $\mathcal{G}$ in $\mathcal{R}$ (Equation S.8) includes all the generators for the terms in $\mathcal{Z}$ and each Pauli $A_k^{(j)}$ operator. Any operator in $\mathcal{Z}$ and each Pauli $A_k^{(j)}$ operator can therefore be found by a finite combination of operators in $\mathcal{G}$. Each operator in $\mathcal{T}$ can also be generated by a combination of one $P_0^{(j)}$ operator and some combination of operators in $\mathcal{G}$. Again $\mathcal{R}$ (Equation S.8) contains all the operators required. To summarise, the set $\mathcal{R}$ contains all the required terms to reproduce the expectation value of any operator in $\mathcal{S}^{H_{\text{noncon}}}$ under the Jordan product.

We have shown how the Pauli operators making H_{noncon} can be simultaneously assigned definite values without contradiction. This allows the introduction of a phase-space description of the eigenspace of H_{noncon} [6]. Next we will introduce what this phase-space model is in the context of this work.

A joint value assignment of ± 1 to each operator in $\mathcal{R}$ represents the ontic state of the physical system. Probability distributions corresponding to valid quantum states must obey an uncertainty relation [6, 8]. To enforce these two conditions are sufficient. (1) The commuting generators $\mathcal{G}$ have definite values and (2) the expectation values for the $P_0^{(j)}$ terms form a unit vector [6]. In this frame, our noncontextual state is defined as [6]:

$$(\vec{q},\vec{r}) = (q_0, q_1, \ldots, q_{|\mathcal{G}|-1},\ r_0, r_1, \ldots, r_{N-1}). \tag{S.9}$$

With respect to the phase-space model [6], a valid noncontextual state $(\vec{q},\vec{r})$ sets the expectation value of the operators in $\mathcal{R}$, where $\langle G_i\rangle = q_i = \pm 1$ and $\langle P_0^{(j)}\rangle = r_j$ such that $\left(\sum_{j=0}^{N-1}|r_j|^2\right)^{1/2} = 1$.

It was shown in [6] that probabilities for outcomes G_i and $P_0^{(j)}$ should be obtained as the marginals of:

$$P(p_{j=0}, p_{j=1}, \ldots, p_{j=(N-1)}, g_0, g_1, \ldots, g_{|\mathcal{G}|-1}) = \left(\prod_{i=0}^{|\mathcal{G}|-1}\delta_{g_i,q_i}\right)\left(\prod_{j=0}^{N-1}\frac{1}{2}|p_j + r_j|\right). \tag{S.10}$$

Further analysis is given in [6].

In summary, a noncontextual state is fully defined by $(\vec{q},\vec{r})$ which determines all the expectation values of the operators in $\mathcal{R}$ (Equation S.8), where $\langle G_i\rangle = q_i \in \{-1,+1\}$ and $\langle P_0^{(j)}\rangle = r_j$. We summarise this as:

$$\underbrace{\{\underbrace{\langle P_0^{(0)}\rangle}_{r_0}, \underbrace{\langle P_0^{(1)}\rangle}_{r_1}, \ldots, \underbrace{\langle P_0^{(N-1)}\rangle}_{r_{N-1}}\}}_{\vec{r}} \text{ and } \underbrace{\{\underbrace{\langle G_0\rangle}_{q_0}, \underbrace{\langle G_1\rangle}_{q_1}, \ldots, \underbrace{\langle G_{|\mathcal{G}|-1}\rangle}_{q_{|\mathcal{G}|-1}}\}}_{\vec{q}}. \tag{S.11}$$

The expectation value of all the operators in $\mathcal{S}^{H_{\text{noncon}}}$ are generated from some finite combination of terms in $\mathcal{R}$ under the Jordan product. This by extension will induce the expectation value for H_{noncon}. Explicitly, let $P_i^{\mathcal{Z}} \in \mathcal{Z} \subseteq \mathcal{S}^{H_{\text{noncon}}}$ then if we let $\mathcal{J}^{\mathcal{G}}_{P_i^{\mathcal{Z}}}$ be the set of indices such that $P_i^{\mathcal{Z}} = \prod_{i\in\mathcal{J}^{\mathcal{G}}_{P_i^{\mathcal{Z}}}} G_i$; then [6]:

$$\langle P_i^{\mathcal{Z}}\rangle = \prod_{i\in\mathcal{J}^{\mathcal{G}}_{P_i^{\mathcal{Z}}}}\langle G_i\rangle = \prod_{i\in\mathcal{J}^{\mathcal{G}}_{P_i^{\mathcal{Z}}}} q_i. \tag{S.12}$$

In words, we combine the expectation value of some finite set of Pauli operators in the independent set $\mathcal{G}$ - given by $\mathcal{J}^{\mathcal{G}}_{P_i^{\mathcal{Z}}}$ - to reproduce the expectation value for $\langle P_i^{\mathcal{Z}}\rangle$.

Similarly, the expectation value for each $A_k^{(j)} P_0^{(j)} \in \mathcal{T} \subseteq \mathcal{S}^{H_{\text{noncon}}}$ (Equation S.5) term is given by:

$$\langle A_i^{(j),\in\mathcal{G}} P_0^{(j),\in\mathcal{T}} \rangle = \Big(\prod_{i \in \mathcal{J}^{\mathcal{G}}_{A_i^{(j)}}} \langle G_i \rangle \Big) r_j = \Big(\prod_{i \in \mathcal{J}^{\mathcal{G}}_{A_i^{(j)}}} q_i \Big) r_j, \tag{S.13}$$

where $\mathcal{J}^{\mathcal{G}}_{A_i^{(j)}}$ are the set of indices such that $\langle A_i^{(j)} \rangle = \prod_{i \in \mathcal{J}^{\mathcal{G}}_{A_i^{(j)}}} \langle G_i \rangle$ and $\langle P_0^{(j)} \rangle = r_j$ [6].

We can write the noncontexual Hamiltonian as:

$$H_{\text{noncon}} = \Big(\sum_{i=0}^{|\mathcal{Z}|-1} c_i P_i^{\mathcal{Z}} \Big) + \sum_{j=0}^{N-1} \Bigg[\sum_{k=0}^{|C_j|-1} a_k A_k^{(j)} P_0^{(j)} \Bigg] \tag{S.14}$$

and find the energy by:

$$\langle H_{\text{noncon}} \rangle = E_{noncon}(\vec{q}, \vec{r}) = \Big(\sum_{i=0}^{|\mathcal{Z}|-1} \beta_i \langle P_i^{\mathcal{Z}} \rangle \Big) + \sum_{j=0}^{N-1} \Bigg[\sum_{k=0}^{|C_j|-1} \beta_k \langle A_k^{(j)} P_0^{(j)} \rangle \Bigg], \tag{S.15}$$

where β_i and β_k are real coefficients and each expectation value is given by Equation S.12 and S.13 [6].

4. Solving the Noncontexual Hamiltonian

To find the ground state of H_{noncon}, we minimize Equation S.15 via a brute-force search as described in [6]. Algorithm 2 summarises the steps. This could be done in the work presented here, because $\mathcal{R}$ was small for all molecular systems considered. First a trial $\vec{q}$ is defined. This is a set of ± 1 expectation values for each G_j. An initial guess of the amplitudes r_i of the unit vector $\vec{r}$ is made and the energy (Equation S.15) is minimized over this continuous parameterization of $\vec{r}$ for a fixed trial $\vec{q}$, until the energy converges to a minimum. These steps are repeated for all the $2^{|\mathcal{G}|}$ assignments of $\vec{q}$. The $(\vec{q}, \vec{r})$ combination that gives the lowest overall energy represents the noncontextual ground state of the physical system. In the main text we denote this parameterization as $(\vec{q}_0, \vec{r}_0)$. Note for a fixed $\vec{q}$, we optimize over $\vec{r}$. This can be thought of as optimizing a function defined on a hypersphere. Currently we haven't explored the properties of this function.

It remains an open question for the CS-VQE algorithm if alternate optimization strategies are possible, for example using chemical intuition during optimization. This brute force approach of searching over all $2^{|\mathcal{G}|}$ possibilities for $\vec{q}$ may not be necessary. In the next section, we discuss how to map the contextual problem into a subspace consistent with a defined noncontextual state $(\vec{q}, \vec{r})$.

Algorithm 2 Brute force method to solve noncontexual problem

$\mathcal{Q} \leftarrow \{q_0, q_1, ..., q_{|\mathcal{G}|-1}\}^{2^{|\mathcal{G}|}}$ ▷ Set $\mathcal{Q}$ contains all possible $\vec{q}$ vectors, where $q_i \in \{+1, -1\}$.

$\vec{q}_0 \leftarrow \{\}$
$\vec{r}_0 \leftarrow \{\}$
$E^0_{\text{noncon}} \leftarrow 0$

for $\vec{q}_{\text{test}}$ in $\mathcal{Q}$ **do**
  $\vec{r}_{opt}, E^{opt}_{\text{noncon}} \leftarrow \underset{\vec{r}}{argmin}[E_{\text{noncon}}(\vec{q}_{\text{test}}, \vec{r})]$ ▷ for a given $\vec{q}_{\text{test}}$, minimize the energy (Equation S.15)with respect to $\vec{r}$.
  if $E^{opt}_{\text{noncon}} < E^0_{\text{noncon}}$ **then**
    $\vec{q}_0 \leftarrow \vec{q}_{\text{test}}$
    $\vec{r}_0 \leftarrow \vec{r}_{opt}$
    $E^0_{\text{noncon}} \leftarrow E^{opt}_{\text{noncon}}$
  else
    continue
  end if
end for

return $\vec{q}_0, \vec{r}_0, E^0_{\text{noncon}}$

B. Mapping to contextual subspace

In section II E of the main text, the full Hamiltonian is mapped contextual subspace consistent with the noncontextual ground state by implementing $U_{\mathcal{W}}$ (Equation 10) followed by projecting the rotated Hamiltonian with $Q_{\mathcal{W}}$. However, the definitions for the operators making up $U_{\mathcal{W}}$ were omitted. This subsection gives these details and is split into three parts. The first two parts consider how R is constructed. This is the problem of mapping a linear combination of pairwise anticommuting Pauli operators to a single Pauli operator and is known as unitary partitioning [9, 10]. In the context of this work, we use this to define R such that $RA(\vec{r})R^\dagger \mapsto P_0^{(k)}$. In the original formulation of CS-VQE only the sequence of rotations construction of R is used in the algorithm. We provide an alternative approach using the linear combination of unitaries construction proposed in [9], which results in superior scaling. We show the effect each conjugate rotation R has on the number of terms in a given qubit Hamiltonian. The last subsection gives the unitary rotations required to map a commuting set of Pauli operators to single qubit Pauli Z operators.

In each of these subsections, we use the notation that Pauli operators with multiple indices represent the multiplication of Pauli operators: $P_aP_bP_c = P_{abc}$. These terms will also be Pauli operators up to a complex phase.

1. *Unitary partitioning via a sequence of rotations*

In this subsection, we show how $A(\vec{r}) \mapsto P_0^{(k)} = R_S A(\vec{r}) R_S^\dagger$, where R_S is defined by a sequence of rotations [9, 11]. Given the set of anticommuting operators $A(\vec{r})$ (Equation 5), we can define the following self-inverse operators:

$$\{\mathcal{X}_{kj} = iP_0^{(k)}P_0^{(j)} \;\; \forall P_0^{(j)} \in \mathcal{A} \text{ where } j \neq k\}, \tag{S.16}$$

where $P_0^{(k)} \in \mathcal{A}$. To simplify the notation we drop the subscript 0 (denoting the first operator in a clique) and write each $P_0^{(k)}$, $P_0^{(j)}$ as P_k and P_j respectively.

The adjoint rotation generated by one of these operator $\mathcal{X}_{kj}$ operators will be:

$$\begin{aligned} e^{(-i\frac{\theta_{kj}}{2}\mathcal{X}_{kj})}A(\vec{r})e^{(+i\frac{\theta_{kj}}{2}\mathcal{X}_{kj})} &= R_{S_{kj}}(\theta_{kj})A(\vec{r})R_{S_{kj}}^\dagger(\theta_{kj}) \\ &= \big(r_j\cos\theta_{kj} - r_k\sin\theta_{kj}\big)P_j + \big(\beta_j\sin\theta_{kj} + r_k\cos\theta_{kj}\big)P_k + \sum_{\substack{P_l\in\mathcal{A}\\ \forall l\neq k,j}}\beta_l P_l. \end{aligned} \tag{S.17}$$

The coefficient of P_j can be made to go to 0, by setting $r_j\cos\theta_{kj} = r_k\sin\theta_{kj}$. This approach removes the term with index j and increases the coefficient of P_k from $r_k \mapsto \sqrt{r_k^2 + r_j^2}$ [9]. This process is repeated over all indices excluding $j = k$ until only the P_k term remains. This procedure can be concisely written using the following operator [9]:

$$R_S = \prod_{\substack{j=0\\ \forall j\neq k}}^{|\mathcal{A}|-1} e^{(-i\frac{\theta_{kj}}{2})\mathcal{X}_{kj}} = \prod_{\substack{j=0\\ \forall j\neq k}}^{|\mathcal{A}|-1} R_{S_{kj}}(\theta_{kj}) = \prod_{\substack{j=0\\ \forall j\neq k}}^{|\mathcal{A}|-1} \Big[\cos\big(\frac{\theta_{kj}}{2}\big)\mathcal{I} - i\sin\big(\frac{\theta_{kj}}{2}\big)\mathcal{X}_{kj}\Big], \tag{S.18}$$

which is simply a sequence of rotations. The angle θ_{kj} is defined recursively at each step of the removal process, as the coefficient of P_k increases at each step and thus must be taken into account. The correct solution for θ_{kj} must be chosen given the signs of r_k and r_k [9]. The overall action of this sequence of rotations is:

$$R_S A(\vec{r})R_S^\dagger = P_k. \tag{S.19}$$

Looking at Equation S.18, expanding the product of rotations results in R_S containing $\mathcal{O}(2^{|\mathcal{A}|-1})$ Pauli operators. We write this operator as:

$$R_S = \sum_b^{\mathcal{O}(2^{|\mathcal{A}|-1})} \delta_b P_b. \tag{S.20}$$

The adjoint rotation of R_S on a general Hamiltonian $H_q = \sum_a^{|H_q|} c_a P_a$ is:

$$\begin{aligned} R_S H_q R_S^\dagger &= \Big(\sum_b^{\mathcal{O}(2^{|\mathcal{A}|-1})} \delta_b P_b \Big) \sum_a^{|H_q|} c_a P_a \Big(\sum_c^{\mathcal{O}(2^{|\mathcal{A}|-1})} \delta_c^* P_c \Big) \\ &= \sum_b^{\mathcal{O}(2^{|\mathcal{A}|-1})} \sum_a^{|H_q|} \sum_c^{\mathcal{O}(2^{|\mathcal{A}|-1})} (\delta_b c_a \delta_c^*) P_b P_a P_c. \end{aligned} \tag{S.21}$$

We see that the number of terms increases as $\mathcal{O}(2^{|\mathcal{A}|}|H_q|)$ which was previously shown in [6]. What we show next is additional structure in R_S - due to the X_{kj} operators - means the base of the exponent can be slightly lower; however, it still remains exponential in $|\mathcal{A}|$.

Consider the adjoint rotation of a particular X_{kj} in R_S (Equation S.18):

$$\begin{aligned} R_{S_{kj}} &= \cos\big(\frac{\theta_{kj}}{2}\big)\mathcal{I} + \sin\big(\frac{\theta_{kj}}{2}\big) P_{kj}, \\ R_{S_{kj}}^\dagger &= \cos\big(\frac{\theta_{kj}}{2}\big)\mathcal{I} + \sin\big(\frac{\theta_{kj}}{2}\big) P_{jk}. \end{aligned} \tag{S.22}$$

Performing the adjoint rotation on H_q results in the following:

$$\begin{aligned} R_{S_{kj}} H_q R_{S_{kj}}^\dagger &= \Big[\alpha_{kj}\mathcal{I} + \beta_{kj} P_{kj}\Big] \sum_a c_a P_a \Big[\alpha_{kj}\mathcal{I} + \beta_{kj} P_{jk}\Big] \\ &= \sum_a c_a \big(\alpha_{kj} P_a + \beta_{kj} P_{kj} P_a\big) \Big[\alpha_{kj}\mathcal{I} + \beta_{kj} P_{jk}\Big] \\ &= \sum_a c_a \big(\alpha_{kj}^2 P_a + \alpha_{kj}\beta_{kj} P_a P_{jk} + \alpha_{kj}\beta_{kj} \underline{P_{kj}} P_a + \beta_{kj}^2 P_{kj} P_a P_{jk}\big) \\ &= \sum_a c_a \big(\alpha_{kj}^2 P_a + \alpha_{kj}\beta_{kj} \underline{P_a P_{jk}} - \alpha_{kj}\beta_{kj} \underline{P_{jk} P_a} + \beta_{kj}^2 P_{kj} P_a P_{jk}\big) \\ &= \sum_a c_a \big(\alpha_{kj}^2 P_a + \alpha_{kj}\beta_{kj} [P_a, P_{jk}] + \beta_{kj}^2 P_{kj} P_a P_{jk}\big) \\ &= \sum_a c_a \begin{cases} \big(\alpha_{kj}^2 P_a + \beta_{kj}^2 P_{kj} P_a P_{jk}\big), & \text{if } [P_a, P_{jk}] = 0 \\ \big(\alpha_{kj}^2 P_a + 2\alpha_{kj}\beta_{kj} P_a P_{jk} + \beta_{kj}^2 P_{kj} P_a P_{jk}\big), & \text{else } \{P_a, P_{jk}\} = 0 \end{cases} \end{aligned} \tag{S.23}$$

When $[P_a, P_{jk}] = 0$, we get:

$$\begin{aligned} \sum_a c_a \big(\alpha_{kj}^2 P_a + \beta_{kj}^2 P_{kj} \underline{P_a P_{jk}}\big) &= \sum_a c_a \big(\alpha_{kj}^2 P_a + \beta_{kj}^2 \underline{P_{kj} P_{jk}} P_a\big) \\ &= \sum_a c_a \big(\alpha_{kj}^2 P_a + \beta_{kj}^2 P_a\big) \\ &= \sum_a c_a \big(\alpha_{kj}^2 + \beta_{kj}^2\big) P_a \\ &= \sum_a c_a P_a. \end{aligned} \tag{S.24}$$

When $\{P_a, P_{jk}\} = 0$, we find:

$$\begin{aligned}\sum_a c_a\big(\alpha_{kj}^2 P_a + 2\alpha_{kj}\beta_{kj}P_aP_{jk} + \beta_{kj}^2 P_{kj}\underline{P_aP_{jk}}\big) &= \sum_a c_a\big(\alpha_{kj}^2 P_a + 2\alpha_{kj}\beta_{kj}P_aP_{jk} - \beta_{kj}^2\underline{P_{kj}P_{jk}}P_a\big)\\ &= \sum_a c_a\big(\alpha_{kj}^2 P_a + 2\alpha_{kj}\beta_{kj}P_aP_{jk} - \beta_{kj}^2 P_a\big)\\ &= \sum_a c_a\big(\alpha_{kj}^2 P_a + sin(\theta_{kj})P_aP_{jk} - \beta_{kj}^2 P_a\big)\\ &= \sum_a c_a\big((\alpha_{kj}^2 - \beta_{kj}^2)P_a + sin(\theta_{kj})P_aP_{jk}\big)\\ &= \sum_a c_a\big(cos(\theta_{kj})P_a + sin(\theta_{kj})P_aP_{jk}\big).\end{aligned} \tag{S.25}$$

Both cases use the following identities:

$$\begin{aligned}\alpha_{kj}^2 - \beta_{kj}^2 &= \cos^2\big(\frac{\theta_{kj}}{2}\big) - \sin^2\big(\frac{\theta_{kj}}{2}\big) = \cos\big(\theta_{kj}\big),\\ \alpha_{kj}^2 + \beta_{kj}^2 &= 1,\\ 2\alpha_{kj}\beta_{kj} &= 2\cos\big(\frac{\theta_{kj}}{2}\big)\sin\big(\frac{\theta_{kj}}{2}\big) = \sin\big(\theta_{kj}\big),\end{aligned} \tag{S.26}$$

where $\alpha_{kj} = \cos\big(\frac{\theta_{kj}}{2}\big)$ and $\beta_{kj} = \sin\big(\frac{\theta_{kj}}{2}\big)$. Using these results Equation S.23 reduces to:

$$\begin{aligned}R_{S_{kj}}H_qR_{S_{kj}}^\dagger &= \sum_{\substack{a\\ \forall[P_a,P_{jk}]=0}} c_aP_a + \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} c_a\Big(cos(\theta_{kj})P_a + sin(\theta_{kj})P_aP_{jk}\Big)\\ &= \sum_a \eta_aP_a + \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big(P_{jk}P_a\big),\end{aligned} \tag{S.27}$$

where η_a represent the new real coefficients.

Consider the application of the next rotation operator R_{kl} in R_S (note k index represents the same Pauli operator P_k):

$$R_{S_{kl}}R_{S_{kj}}H_qR_{S_{kj}}^\dagger R_{S_{kl}}^\dagger = R_{S_{kl}}\Big(\sum_a \eta_aP_a\Big)R_{S_{kl}}^\dagger + \underline{R_{S_{kl}}\Big(\sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big(P_{jk}P_a\big)\Big)R_{S_{kl}}^\dagger} \tag{S.28}$$

Focusing on the last term in Equation S.28:

$$
\begin{aligned}
R_{S_{kl}}\Bigg(\sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\big(P_{jk}P_a\big)\Bigg)R_{S_{kl}}^{\dagger} &= \Big[\gamma_{kl}\mathcal{I}+\delta_{kl}P_{kl}\Big]\sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\big(P_{jk}P_a\big)\Big[\gamma_{kl}\mathcal{I}+\delta_{kl}P_{lk}\Big]\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}P_{jk}P_a+\delta_{kl}P_{kl}P_{jk}P_a\Big)\Big[\gamma_{kl}\mathcal{I}+\delta_{kl}P_{lk}\Big]\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}^2P_{jk}P_a+\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk}+\gamma_{kl}\delta_{kl}P_{k\underline{l}}P_{\underline{j}k}P_a+\delta_{kl}^2P_{kl}P_{jk}P_aP_{lk}\Big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}^2P_{jk}P_a+\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk}-\gamma_{kl}\delta_{kl}P_{\underline{kj}}P_{lk}P_a+\delta_{kl}^2P_{kl}P_{jk}P_aP_{lk}\Big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}^2P_{jk}P_a+\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk}+\gamma_{kl}\delta_{kl}P_{jk}P_{lk}P_a+\delta_{kl}^2P_{kl}P_{jk}P_aP_{lk}\Big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}^2P_{jk}P_a+\gamma_{kl}\delta_{kl}P_{jk}\{P_a,P_{lk}\}+\delta_{kl}^2P_{kl}P_{jk}P_aP_{lk}\Big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\begin{cases}\Big(\gamma_{kl}^2P_{jk}P_a+2\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk}+\delta_{kl}^2P_{kl}P_{jk}P_aP_{lk}\Big), & \text{if } [P_a,P_{lk}]=0\\ \Big(\gamma_{kl}^2P_{jk}P_a+\delta_{kl}^2P_{kl}P_{jk}P_aP_{lk}\Big), & \text{if } \{P_a,P_{lk}\}=0\end{cases}
\end{aligned}
\tag{S.29}
$$

For the case $\{P_a, P_{lk}\} = 0$:

$$
\begin{aligned}
\sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}^2P_{jk}P_a+\delta_{kl}^2P_{kl}P_{jk}\underline{P_aP_{lk}}\Big) &= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}^2P_{jk}P_a-\delta_{kl}^2\underline{P_{kl}P_{jk}}P_{lk}P_a\Big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}^2P_{jk}P_a+\delta_{kl}^2P_{jk}P_{kl}P_{lk}P_a\Big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(\gamma_{kl}^2P_{jk}P_a+\delta_{kl}^2P_{jk}P_a\Big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}}\eta_a\Big(P_{jk}P_a\Big).
\end{aligned}
\tag{S.30}
$$

We observe that there is no increase in the number of terms and the weight of each Pauli operator changes.

For the case $[P_a, P_{lk}] = 0$:

$$\begin{aligned}
\sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big(\gamma_{kl}^2 P_{jk}P_a + 2\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk} + \delta_{kl}^2 P_{kl}P_{jk}\underline{P_aP_{lk}}\big) &= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big(\gamma_{kl}^2 P_{jk}P_a + 2\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk} + \delta_{kl}^2 \underline{P_{kl}P_{jk}}P_{lk}P_a\big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big(\gamma_{kl}^2 P_{jk}P_a + 2\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk} - \delta_{kl}^2 P_{jk}P_{kl}P_{lk}P_a\big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big(\gamma_{kl}^2 P_{jk}P_a + 2\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk} - \delta_{kl}^2 P_{jk}P_a\big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big((\gamma_{kl}^2 - \delta_{kl}^2)P_{jk}P_a + 2\gamma_{kl}\delta_{kl}P_{jk}P_aP_{lk}\big)\\
&= \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big(\cos\big(\theta_{kl}\big)P_{jk}P_a + \sin\big(\theta_{kl}\big)P_{jk}P_aP_{lk}\big).
\end{aligned} \tag{S.31}$$

The number of terms in the resulting operator has increased for each case where $[P_a, P_{lk}] = 0$. The action of two rotations of R_S on the whole Hamiltonian results in:

$$\begin{aligned}
R_{S_{kl}}R_{S_{kj}}HR_{S_{kj}}^\dagger R_{S_{kl}}^\dagger =& R_{S_{kl}}\Big(\sum_a \eta_a P_a\Big)R_{S_{kl}}^\dagger + R_{S_{kl}}\Bigg(\sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \eta_a\big(P_{jk}P_a\big)\Bigg)R_{S_{kl}}^\dagger\\
=& R_{S_{kl}}\Big(\sum_a \eta_a P_a\Big)R_{S_{kl}}^\dagger + \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0\\ \forall\{P_a,P_{lk}\}=0}} \eta_a P_{jk}P_a +\\
&\sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0\\ [P_a,P_{lk}]=0}} \eta_a\big(\cos\big(\theta_{kl}\big)P_{jk}P_a + \sin\big(\theta_{kl}\big)P_{jk}P_aP_{lk}\big)\\
=& R_{S_{kl}}\Big(\sum_a \eta_a P_a\Big)R_{S_{kl}}^\dagger + \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \mu_a P_{jk}P_a + \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0\\ [P_a,P_{lk}]=0}} \mu_a P_{jk}P_aP_{lk}\\
=& \sum_a \nu_a P_a + \sum_{\substack{a\\ \forall\{P_a,P_{lk}\}=0}} \nu_a\big(P_{lk}P_a\big) + \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0}} \mu_a P_{jk}P_a + \sum_{\substack{a\\ \forall\{P_a,P_{jk}\}=0\\ [P_a,P_{lk}]=0}} \mu_a P_{jk}P_aP_{lk}.
\end{aligned} \tag{S.32}$$

where Greek letters are new coefficients according to the expansion. We use the results of Equations S.30 and S.31 to determine what occurs to the second term of Equation S.32. We have applied the result in Equation S.27 to the first term ($R_{S_{kl}}\Big(\sum_i \eta_i P_i\Big)R_{S_{kl}}^\dagger$) in Equation S.32.

From these results we can infer how the terms in H_q will scale for a general sequence of rotations of size $|R_S|$ (Equation S.18), which in general change as:

$$|H_q|\sum_{g=0}^{|R_S|}\binom{|R_S|}{g} = 2^{|R_S|}|H_q| \tag{S.33}$$

This operation increases the number of terms in H_q to $\mathcal{O}\big(2^{(|\mathcal{A}|-1)}|H_q|\,\big)$. However, the structure of the sequence of rotation operator actually requires 2^g commuting/anticommuting conditions to be met for new Pauli operators to be generated by subsequent rotations. We therefore need to consider the probability that a given Pauli operator will either commute or anticommute with another. For the case of single qubit Pauli matrices $\sigma_a, \sigma_b \in \{I, X, Y, Z\}$ by a simple counting argument $P([\sigma_a, \sigma_b] = 0) = \frac{5}{8}$ and $P(\{\sigma_a, \sigma_b\} = 0) = \frac{3}{8}$, for Pauli matrices selected uniformly at random. Generalising this to tensor products of Pauli matrices on n qubits, for a Pauli operator to anticommute with another there needs to be an odd number of anticommuting tensor factors. First consider the binomial distribution:

$$P(x) = \binom{n}{x} p^x q^{n-x}, \tag{S.34}$$

where n is the number of trials (repeated experiments), p is the probability of success - here the probability a single Pauli matrix anticommutes with another ($p = \frac{3}{8}$) - and q is the probability of failure - here the probability a single Pauli matrix commutes with another ($q = \frac{5}{8}$). Under these conditions, $P(x)$ gives the probability that two n-fold Pauli operators, selected uniformly at random, anticommute in x-many tensor factors. Therefore, the probability of two uniformly random Pauli operators anticommuting (commuting) is given as a sum over odd (even) values of $x \leq n$:

$$P\big(\{P_a, P_b\} = 0\big) = \sum_{c=1}^{\lceil n/2 \rceil} P(2c-1). \tag{S.35}$$

Now, the binomial theorem states

$$(p+q)^n = \sum_{c=0}^{n} \binom{n}{c} p^c q^{n-c} \tag{S.36}$$

for any $p, q \in \mathbb{R}$ and define the following difference:

$$(p+q)^n - (-p+q)^n = \sum_{c=0}^{n} \binom{n}{c} \underbrace{\big[1-(-1)^c\big]}_{=\begin{cases} 2, & \text{if } c \text{ odd} \\ 0, & \text{if } c \text{ even} \end{cases}} p^c q^{n-c} = 2 \sum_{c=1}^{\lceil n/2 \rceil} \binom{n}{2c-1} p^{2c-1} q^{n-(2c-1)}. \tag{S.37}$$

Overall we find the probability that two n-fold Pauli operators anticommute to be:

$$\begin{aligned} P\big(\{P_a, P_b\} = 0\big) &= \sum_{c=1}^{\lceil n/2 \rceil} P(2c-1) \\ &= \sum_{c=1}^{\lceil n/2 \rceil} \binom{n}{2c-1} \cdot \left(\frac{3}{8}\right)^{2c-1} \cdot \left(\frac{5}{8}\right)^{n-(2c-1)} \\ &= \frac{1}{2}\left[\left(\frac{3}{8} + \frac{5}{8}\right)^n - \left(-\frac{3}{8} + \frac{5}{8}\right)^n \right] \\ &= \frac{1}{2}\left[1 - \left(\frac{1}{4}\right)^n\right], \end{aligned} \tag{S.38}$$

when each operator P_a, P_b is chosen uniformly at random. The n choose $2c-1$ term in equation S.38 counts all the possible ways an odd number of single qubit pairs of Pauli tensor factors can differ on n qubits, the first fraction gives the probability that there are $2c-1$ anticommuting terms on each pair of qubits and the final fraction gives the probability that the remaining $n-(2c-1)$ qubit positions pairwise commute on each qubit. The penultimate line of equation S.38 uses the definition in S.37, with the factor of two taken into account. Through equation S.38, it can be seen that the probability of two n-fold Pauli operators anticommuting quickly converges to 0.5 as the number of qubits n increases. The motivation for S.37 arises from observing that the quantity we subtract, $(1/4)^n$, is the probability of obtaining an n-fold identity operator, which has the unique property of commuting universally. The complement $1-(1/4)^n$ therefore corresponds with the probability of selecting uniformly at random a Pauli operator with at least one non-trivial tensor factor. After discounting identity operators from consideration, the probabilities of anticommuting or commuting coincide, hence each occurs half of the time, explaining the $1/2$ factor in S.38; the probability bias towards commutation is a consequence of the identity operator commuting universally, whereas there is no such operator that can anticommute universally.

If we consider how the number of terms in H_q changes upon the sequence of rotations transformation: $H_q \mapsto R_S H_q R_S^\dagger$ where terms either commute or anticommute with a probability of 0.5, then the scaling is as follows:

$$\sum_{g=0}^{|R_S|} \frac{|H_q|}{2^g} \binom{|R_S|}{g} = \left(\frac{3}{2}\right)^{|R_S|} |H_q|. \tag{S.39}$$

Equation S.33 is modified to have a constant factor of 2^{-g}, where g represents the number of commuting or anticommuting conditions required for operators in H_q to obey in order to increase the number of terms upon a rotation of R_S. Here each condition is assumed to occur with a probability of 0.5. This operation increases the number of terms in H_q to $\mathcal{O}\big(1.5^{(|\mathcal{A}|-1)}|H_q|\,\big)$. Note $|R_S| = |\mathcal{A}| - 1$. In general, the scaling will be $\mathcal{O}\big(x^{(|\mathcal{A}|-1)}|H_q|\,\big)$ where $1 \leq x \leq 2$, depending on how each rotation in the sequence of rotations commutes with terms in H_q. The $x = 1$ case occurs if each rotation in R_S commutes with the whole Hamiltonian. Apart from this special case, the number of terms in H_q will increase exponentially with the size of $\mathcal{A}$ or equivalently with the number of qubits n (as $|\mathcal{A}| \leq 2n + 1$ [6]) when R is defined by a sequence of rotations.

2. *Unitary partitioning via a linear combination of unitaries*

Here we show how $A(\vec{r}) \mapsto P_0^{(k)} = R_{LCU} A(\vec{r}) R_{LCU}^\dagger$, where R_{LCU} is defined by a linear combination of Pauli operators. We consider the set of anticommuting Pauli operators making up $A(\vec{r})$ (Equation 5). We can re-write this Equation, with the term we are reducing to ($r_k P_0^{(k)}$) outside the sum:

$$A(\vec{r}) = r_k P_0^{(k)} + \sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} r_j P_0^{(j)}. \tag{S.40}$$

To simplify the notation we drop the subscript 0 (denoting the first operator in a clique) and write each $P_0^{(k)}$, $P_0^{(j)}$ as P_k and P_j respectively.

A re-normalization can be performed on the remaining sum yielding:

$$\begin{aligned} A(\vec{r}) &= r_k P_k + \Omega \sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} \delta_j P_j \\ &= r_k P_k + \Omega H_{\mathcal{A}\setminus\{r_k P_k\}}, \end{aligned} \tag{S.41}$$

where:

$$\sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} |\delta_j|^2 = 1, \tag{S.42a}$$

$$r_j = \Omega \delta_j, \tag{S.42b}$$

$$H_{\mathcal{A}\setminus\{r_k P_k\}} = \sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} \delta_j P_j. \tag{S.42c}$$

Using the Pythagorean trigonometric identity: $\sin^2(x) + \cos^2(x) = 1$, $A(\vec{r})$ can be re-written as:

$$\begin{aligned} A(\vec{r}) &= \cos(\phi_k) P_k + \sin(\phi_k) \sum_{\substack{j=0 \\ \forall j \neq k}}^{N-1} \delta_i P_j \\ &= \cos(\phi_k) P_k + \sin(\phi_k) H_{\mathcal{A}\setminus\{r_k P_k\}}. \end{aligned} \tag{S.43}$$

Comparing Equations S.41 and S.43, it is clear that $\cos(\phi_k) = r_k$ and $\sin(\phi_k) = \Omega$.

It was shown in [9] that one can consider rotations of $A(\vec{r})$ around an axis that is Hilbert-Schmidt orthogonal to both $H_{\mathcal{A}\setminus\{r_k P_k\}}$ and P_k :

$$\mathcal{X} = \frac{i}{2}\big[H_{\mathcal{A}\setminus\{r_k P_k\}}, P_k\big] = i \sum_{\substack{j=0 \\ \forall j \neq k}}^{|H_{\mathcal{A}\setminus\{r_k P_k\}}|-1} \delta_j P_j P_k. \tag{S.44}$$

$\mathcal{X}$ anticommutes with $\mathcal{A}$ and is self-inverse [9]:

$$\mathcal{X}^2 = \left(i \sum_{j=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j P_j P_k\right)\left(i \sum_{l=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_l P_l P_k\right), \tag{S.45a}$$

$$= - \sum_{j=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \sum_{l=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j\delta_l P_j P_k P_l P_k, \tag{S.45b}$$

$$= - \sum_{k=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \sum_{\substack{l=0 \\ \forall l=j}}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j\delta_j P_j P_k P_j P_k - \sum_{j=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \sum_{\substack{l=0 \\ \forall l\neq j}}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j\delta_l P_j P_k P_l P_k, \tag{S.45c}$$

$$= + \sum_{k=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j^2 \underbrace{P_k P_j}_{\text{order change}} P_j P_k - \sum_{j=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \sum_{l>j}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j\delta_l \underbrace{\{P_jP_k, P_lP_k\}}_{=0 \text{ when } j\neq l}, \tag{S.45d}$$

$$= + \sum_{j=0}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j^2 \mathcal{I}, \tag{S.45e}$$

$$= \mathcal{I} \tag{S.45f}$$

and has the following action [9]:

$$\mathcal{X}A(\vec{r}) = i\big(-\sin\phi_k P_k + \cos\phi_k H_{\mathcal{A}\setminus\{r_kP_k\}}\big). \tag{S.46}$$

One can also define the rotation [9, 11]:

$$R_{LCU} = e^{\left(-i\frac{\alpha}{2}\mathcal{X}\right)} = \cos\left(\frac{\alpha}{2}\right)\mathcal{I} - i\sin\left(\frac{\alpha}{2}\right)\mathcal{X} \tag{S.47a}$$

$$= \cos\left(\frac{\alpha}{2}\right)\mathcal{I} - i\sin\left(\frac{\alpha}{2}\right)\left(i \sum_{\substack{j=0 \\ \forall j\neq k}}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j P_j P_k\right) \tag{S.47b}$$

$$= \cos\left(\frac{\alpha}{2}\right)\mathcal{I} + \sin\left(\frac{\alpha}{2}\right) \sum_{\substack{j=0 \\ \forall j\neq k}}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j P_{jk}, \tag{S.47c}$$

$$= \delta_{\mathcal{I}}\mathcal{I} + \sum_{\substack{j=0 \\ \forall j\neq k}}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j P_{jk}. \tag{S.47d}$$

The conjugate rotation will be:

$$R_{LCU}^\dagger = \cos\left(\frac{\alpha}{2}\right)\mathcal{I} + i\sin\left(\frac{\alpha}{2}\right) i \sum_{\substack{j=0 \\ \forall j\neq k}}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j P_j P_k, \tag{S.48a}$$

$$= \cos\left(\frac{\alpha}{2}\right)\mathcal{I} + \sin\left(\frac{\alpha}{2}\right) \sum_{\substack{j=0 \\ \forall j\neq k}}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j \underbrace{P_k P_j}_{\text{order change}}, \tag{S.48b}$$

$$= \delta_{\mathcal{I}}\mathcal{I} + \sum_{\substack{j=0 \\ \forall j\neq k}}^{|H_{\mathcal{A}\setminus\{r_kP_k\}}|-1} \delta_j P_{kj}. \tag{S.48c}$$

Note the different order of j and k for R_{LCU} and $R_{LCU}^\dagger$. The adjoint action of R_{LCU} on $A(\vec{r})$ is:

$$R_{LCU}A(\vec{r})R_{LCU}^\dagger = \cos\big(\phi_k - \alpha\big)P_k + \sin\big(\phi_k - \alpha\big)H_{\mathcal{A}\setminus\{r_k P_k\}}. \tag{S.49}$$

By choosing $\alpha = \phi_k$, the following transformation occurs $R_{LCU}A(\vec{r})R_{LCU}^\dagger = P_k$ [9, 11]. This fully defines the R_{LCU} operator required by unitary partitioning. Next we need to consider the use of this operator in CS-VQE.

The adjoint action of R_{LCU} on a general Hamiltonian $H_q = \sum_i^{|H_q|} c_i P_i$ is:

$$\begin{aligned}
R_{LCU}H_qR_{LCU}^\dagger =& \left(\delta_{\mathcal{I}}\mathcal{I} + \sum_j^{|R_{LCU}|-1}\delta_j P_{jk}\right)\sum_i^{|H_q|} c_i P_i \left(\delta_{\mathcal{I}}\mathcal{I} + \sum_l^{|R_{LCU}|-1}\delta_l P_{kl}\right) && \text{(S.50a)}\\
=& \left(\delta_{\mathcal{I}}\sum_i^{|H_q|} c_i P_i + \sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\delta_j c_i P_{jk}P_i\right)\left(\delta_{\mathcal{I}}\mathcal{I} + \sum_l^{|R_{LCU}|-1}\delta_l P_{kl}\right) && \text{(S.50b)}\\
=& \ \delta_{\mathcal{I}}^2\sum_i^{|H_q|} c_i P_i && \text{(S.50c)}\\
&+ \sum_l^{|R_{LCU}|-1}\sum_i^{|H_q|}\delta_{\mathcal{I}} c_i\delta_l P_i P_{kl} + \sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\delta_{\mathcal{I}}\delta_j c_i P_{jk}P_i && \text{(S.50d)}\\
&+ \sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\sum_l^{|R_{LCU}|-1}\delta_j c_i\delta_l P_{jk}P_i P_{kl} && \text{(S.50e)}
\end{aligned}$$

We can rewrite the final term (Equation S.50e) as:

$$\begin{aligned}
\sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\sum_l^{|R_{LCU}|-1}\delta_j c_i\delta_l P_{jk}P_iP_{kl} =& \sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\sum_{\substack{l=j\\[P_{jk},P_i]=0}}^{|R_{LCU}|-1}(c_i\delta_j\delta_j^*)P_i && \text{(S.51a)}\\
&+ \sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\sum_{\substack{l=j\\\{P_{jk},P_i\}=0}}^{|R_{LCU}|-1}(-c_i\delta_j\delta_j^*)P_i && \text{(S.51b)}\\
&+ \sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\sum_{l\neq j}^{|R_{LCU}|-1}(\delta_j c_i\delta_l)P_{jk}P_iP_{kl}. && \text{(S.51c)}
\end{aligned}$$

Here we have applied the identity of conjugating a Pauli operator P_u with another Pauli operator P_v resulting in two cases:

$$P_vP_uP_v = \begin{cases} P_u, & \text{if } [P_v, P_u] = 0\\ -P_u, & \text{otherwise } \{P_v, P_u\} = 0\end{cases} \tag{S.52}$$

Focusing on the last term of Equation S.51, we can simplify S.51c as j and l run over the same indices we can re-write each $l \neq j$ sum as $l > j$ and expand into two terms:

$$\sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\sum_{l\neq j}^{|R_{LCU}|-1}(\delta_j c_i\delta_l)P_{jk}P_iP_{kl} = \sum_j^{|R_{LCU}|-1}\sum_i^{|H_q|}\sum_{l>j}^{|R_{LCU}|-1}(\delta_j c_i\delta_l)\Big(P_{jk}P_iP_{kl} + P_{lk}P_iP_{kj}\Big) \tag{S.53}$$

We can expand then expand this equation into the four cases for when:

1. $[P_{jk}, P_i] = 0$ and $[P_{lk}, P_i] = 0$

2. $[P_{jk}, P_i] = 0$ and $\{P_{lk}, P_i\} = 0$

3. $\{P_{jk}, P_i\} = 0$ and $[P_{lk}, P_i] = 0$

4. $\{P_{jk}, P_i\} = 0$ and $\{P_{lk}, P_i\} = 0$

For the first case and last case:

$$\begin{aligned}\sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\sum_{l>j}^{|R_{LCU}|-1}(\delta_j c_i \delta_l)\Big(\underline{P_{jk}P_i}P_{kl} + \underline{P_{lk}P_i}P_{kj}\Big) &= \sum_{j}\sum_{i}^{|H_q|}\sum_{l>j}(\delta_j c_i \delta_l)\Big(\pm P_i\underline{P_{jk}P_{kl}} \pm P_i\underline{P_{lk}P_{kj}}\Big)\\ &= \sum_{j}\sum_{i}^{|H_q|}\sum_{l>j}(\delta_j c_i \delta_l)\Big(\pm P_iP_jP_l \pm P_iP_lP_j\Big)\\ &= \sum_{j}\sum_{i}^{|H_q|}\sum_{l>j}(\delta_j c_i \delta_l) \pm P_i\{P_j, P_l\}\\ &= 0\end{aligned} \tag{S.54}$$

Whereas, for the second and third cases:

$$\begin{aligned}\sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\sum_{l>j}^{|R_{LCU}|-1}(\delta_j c_i \delta_l)\Big(\underline{P_{jk}P_i}P_{kl} + \underline{P_{lk}P_i}P_{kj}\Big) &= \sum_{j}\sum_{i}^{|H_q|}\sum_{l>j}(\delta_j c_i \delta_l)\Big(\pm P_i\underline{P_{jk}P_{kl}} \mp P_i\underline{P_{lk}P_{kj}}\Big)\\ &= \sum_{j}\sum_{i}^{|H_q|}\sum_{l>j}(\delta_j c_i \delta_l)\Big(\pm P_iP_jP_l \mp P_i\underline{P_lP_j}\Big)\\ &= \sum_{j}\sum_{i}^{|H_q|}\sum_{l>j}(\delta_j c_i \delta_l)\Big(\pm P_iP_jP_l \pm P_iP_jP_l\Big)\\ &= \sum_{j}\sum_{i}^{|H_q|}\sum_{l>j}(\delta_j c_i \delta_l) \pm 2P_iP_jP_l\end{aligned} \tag{S.55}$$

We can rewrite Equation S.51 using this result:

$$\begin{aligned}\sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\sum_{l}^{|R_{LCU}|-1}\delta_j c_i \delta_l P_{jk}P_iP_{kl} =& \sum_{j}\sum_{i}^{|H_q|}\sum_{\substack{l=j\\[P_{jk},P_i]=0}}(c_i\delta_j\delta_j^*)P_i + \sum_{j}\sum_{i}^{|H_q|}\sum_{\substack{l=j\\\{P_{jk},P_i\}=0}}(-c_i\delta_j\delta_j^*)P_i +\\ &\sum_{j}\sum_{i}^{|H_q|}\sum_{\substack{l>j\\\forall[P_{jk},P_i]=0\\\{P_{lk},P_i\}=0}}(\delta_j c_i \delta_l)2P_iP_jP_l - \sum_{j}\sum_{i}^{|H_q|}\sum_{\substack{l>j\\\forall\{P_{jk},P_i\}=0\\[P_{lk},P_i]=0}}(\delta_j c_i \delta_l)2P_iP_jP_l\\ =& \sum_{i}^{|H_q|}\nu_iP_i + \sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\sum_{\substack{l>j\\\forall\{P_i,P_jP_l\}=0}}^{|R_{LCU}|-1}\nu_{ijl}P_iP_jP_l\end{aligned} \tag{S.56}$$

where we have combined the second and third conditions into a single condition of $\{P_a, P_{jk}P_{kl}\} = \{P_a, P_jP_l\} = 0$ and combined the new coefficients into one coefficient denoted ν.

Next consider the S.50d term of equation S.50. One can use the fact that j and l run over the same indices:

$$\underbrace{\sum_{l}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\delta_{\mathcal{I}}c_i\delta_l P_i P_{kl}}_{\text{re-write using } l=j}+\sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\delta_{\mathcal{I}}\delta_j c_i P_{jk}P_i = \sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\delta_{\mathcal{I}}c_i\delta_j P_i P_{kj}+\sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\delta_{\mathcal{I}}\delta_j c_i P_{jk}P_i$$

$$\begin{aligned}
&= \sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|} -\delta_{\mathcal{I}}c_i\delta_j P_i P_{jk}+\sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\delta_{\mathcal{I}}\delta_j c_i P_{jk}P_i \\
&= \sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\delta_{\mathcal{I}}c_i\delta_j\big(P_{jk}P_i - P_i P_{jk}\big) \\
&= \sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\delta_{\mathcal{I}}c_i\delta_j[P_{jk},P_i] \\
&= \sum_{j}^{|R_{LCU}|-1}\sum_{\substack{i \\ \forall\{P_{jk},P_i\}=0}}^{|H_q|} 2\delta_{\mathcal{I}}c_i\delta_j P_{jk}P_i
\end{aligned} \tag{S.57}$$

Overall we can re-write equation S.50 using these results, yielding:

$$\begin{aligned}
R_{LCU}H_qR_{LCU}^\dagger = &\underbrace{\delta_{\mathcal{I}}^2\sum_{i}^{|H_q|}c_iP_i}_{S.50c}+\underbrace{\sum_{j}^{|R_{LCU}|-1}\sum_{\substack{i \\ \forall\{P_{jk},P_i\}=0}}^{|H_q|}2\delta_{\mathcal{I}}c_i\delta_jP_{jk}P_i}_{S.50d \text{ using } S.57}+ \\
&\underbrace{\sum_{i}^{|H_q|}\nu_iP_i+\sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\sum_{\substack{l>j \\ \forall\{P_i,P_jP_l\}=0}}^{|R_{LCU}|-1}\nu_{ijl}P_iP_jP_l}_{S.50e \text{ using } S.56} \\
= &\sum_{i}^{|H_q|}(\delta_{\mathcal{I}}^2c_i+\nu_i)P_i+\sum_{j}^{|R_{LCU}|-1}\sum_{\substack{i \\ \forall\{P_{jk},P_i\}=0}}^{|H_q|}2\delta_{\mathcal{I}}c_i\delta_jP_{jk}P_i+ \\
&\sum_{j}^{|R_{LCU}|-1}\sum_{i}^{|H_q|}\sum_{\substack{l>j \\ \forall\{P_i,P_jP_l\}=0}}^{|R_{LCU}|-1}\nu_{ijl}P_iP_jP_l
\end{aligned} \tag{S.58}$$

We observe that the number of terms in $R_{LCU}H_qR_{LCU}^\dagger$ at worst scales as $|H_q|+|H_q|\cdot(|R_{LCU}|-1)+|H_q|\big(\frac{(|R_{LCU}|-1)(|R_{LCU}|-2)}{2}\big)$ or $\mathcal{O}\big(\ |H_q|\cdot|\mathcal{A}|^2\ \big)$. The total number of qubits n bounds the size of $|\mathcal{A}|\leq 2n+1$ [6], and thus the number of terms in H_q will increase quadratically with the size of $\mathcal{A}$ or number of qubits n when R is defined by a linear combination of unitaries.

3. Mapping Pauli operators to single-qubit Pauli Z operators

In Appendix A of [1] a proof is given on how to map a completely commuting set of Pauli operators to a single qubit Pauli Z operator. We summarise the operation required and omit the proof. We denote a given Pauli operator on n qubits as: $P=\bigotimes_{i=0}^{n-1}\sigma_i^P$, where σ_i are a single qubit Pauli operators. There are two cases we need to consider (diagonal and non-diagonal), with the goal to reduce the operators in $\mathcal{R}^{'}\equiv\{P_0^{(k)}\}\cup\mathcal{G}$ (Equation 7) to single-qubit Z Pauli operators.

For a non-diagonal Pauli operators $P_a\in\mathcal{R}'$, there must be at least one single qubit Pauli operator indexed by qubit k such that: $\sigma_k^{P_a}\in\{X,Y\}$. We can use this to define operator P_b that must anticommute with P_a:

$$\begin{aligned} P_a &= \left(\bigotimes_{i=0}^{k-1} \sigma_i^{P_a}\right) \otimes \sigma_k^{P_a} \otimes \left(\bigotimes_{i=k+1}^{n-1} \sigma_i^{P_a}\right) \\ P_b &= \left(\bigotimes_{i=0}^{k-1} \sigma_i^{P_a}\right) \otimes \sigma_k' \otimes \left(\bigotimes_{i=k+1}^{n-1} \sigma_i^{P_a}\right) \end{aligned} \text{ where } \{P_a, P_b\} = 0 \Leftrightarrow \sigma_k' = \begin{cases} X, & \text{if } \sigma_k^{P_a} = Y \\ Y, & \text{if } \sigma_k^{P_a} = X \end{cases} \tag{S.59}$$

These two Pauli operators differ by exactly one Pauli operator on qubit index k. We can define the rotation:

$$B = \exp\left(i\frac{\pi}{4}P_b\right) \tag{S.60}$$

Conjugating P_a with this operator results in:

$$BP_aB^\dagger = \pm 1\left(\bigotimes_{i=0}^{k-1} I_i\right) \otimes Z_k \otimes \left(\bigotimes_{i=k+1}^{n-1} I_i\right) = P_a', \tag{S.61}$$

and P_a has been mapped to a single qubit Pauli Z operator.

For diagonal operators $P_c \in \mathcal{R}'$, all the n-fold tensor products of single qubit Pauli operators must be either Z or I: $P_c = \bigotimes_{i=0}^{n-1} \sigma_i^{P_c}$ where $\sigma_i^{P_c} \in \{I, Z\}\ \forall i$. Since $\mathcal{R}'$ is an independent set, for all the rotated P_a' there must be at least one index l such that $\sigma_l^{P_a'} = I$ and $\sigma_l^{P_c} = Z$. We denote this operator P_c. We also define a new operator P_d from this, which only acts non-trivially on the l-th qubit with a single qubit Y. To summarise:

$$\begin{aligned} P_c &= \left(\bigotimes_{i=0}^{l-1} \sigma_i^{P_c}\right) \otimes Z_l \otimes \left(\bigotimes_{l=k+1}^{n-1} \sigma_i^{P_c}\right) \text{ where } \sigma_i^{P_c} \in \{I, Z\}\ \forall i \\ P_d &= \left(\bigotimes_{i=0}^{l-1} I_i\right) \otimes Y_l \otimes \left(\bigotimes_{i=l+1}^{n-1} I_i\right) \\ P_a' &= \left(\bigotimes_{i=0}^{l-1} \sigma_i^{P_a'}\right) \otimes I_l \otimes \left(\bigotimes_{i=l+1}^{n-1} \sigma_i^{P_a'}\right) \text{ where } [P_d, P_a'] = 0 \text{ and } \sigma_i^{P_a'} \in \{I, Z\} \forall\ i \end{aligned} \tag{S.62}$$

We can define the rotation:

$$D = \exp\left(i\frac{\pi}{4}P_d\right) \tag{S.63}$$

Conjugating P_c with this operator results in:

$$DP_cD^\dagger = \pm 1\left(\bigotimes_{i=0}^{l-1} \sigma_i^{P_c}\right) \otimes X_l \otimes \left(\bigotimes_{i=l+1}^{n-1} \sigma_i^{P_c}\right) = P_c'. \tag{S.64}$$

P_c' is now a non-diagonal Pauli operator (contains a single qubit X acting on qubit l). This operator P_c' can now be mapped to a single qubit Z operator using a further $\frac{\pi}{2}$-rotation following the previously given procedure for diagonal Pauli operators.

The operators V_i in the main text (equation 10) are defined by these $\frac{\pi}{2}$-rotations, such that each q_iG_i and $P_0^{(k)}$ is mapped to a single qubit Pauli Z term. At worst, two $\frac{\pi}{2}$-rotations are needed for every operator in $\mathcal{R}'$ (Equation 7), which occurs when all operators in $\mathcal{R}'$ are diagonal.

II. UNITARY PARTITIONING MEASUREMENT REDUCTION

In unitary partitioning, the Hamiltonian is partitioned into groups of operators that's linear combination are unitary Hermitian operators. This is done by forming normalized groups of Pauli operators that pairwise anticommute. We can write this as:

$$\begin{aligned} H &= \sum_i c_i P_i \\ &= \sum_j \gamma_j C_j \\ &= \sum_j^{N_C} \gamma_j \Bigg(\sum_{\substack{k \\ \{P_a, P_b\}=0 \\ \forall P_a, P_b \in C_j \\ a \neq b}}^{|C_j|} \frac{c_k}{\gamma_j} P_k \Bigg), \end{aligned} \tag{S.65}$$

where $\gamma_j = \left(\sum_k^{C_j} c_k^2\right)^{0.5}$. The complete approach is provided in [9–11]. We follow the analysis of Crawford *et al.* to determine the measurement cost to determine $\langle H \rangle$ to a certain precision [12]. The measurement requirement for measuring the Hamiltonian in terms of grouped terms to precision ϵ is [12]:

$$M_g = \frac{1}{\epsilon^2_{\langle H \rangle}} \left(\sum_j^{N_C} \sqrt{Var[C_j]} \right)^2, \tag{S.66}$$

We can use this to determine the number of measurements when no partitioning has being done [12]:

$$M_u = \frac{1}{\epsilon^2_{\langle H \rangle}} \left(\sum_j^{N_C} \left[\sum_{k=0}^{|C_j|-1} |c_k^{(j)}| \sqrt{Var[P_k^{(j)}]} \right] \right)^2. \tag{S.67}$$

This can be thought of as each clique is of size one. The subscript u is to denote no grouping. A natural metric to evaluate the the measurement cost of a particular grouping of Pauli operators is therefore given by the ratio R of these two terms:

$$R = \frac{M_u}{M_g} = \left(\frac{\sum_j^{N_C} \left[\sum_{k=0}^{|C_j|-1} |c_k^{(j)}| \sqrt{Var[P_k^{(j)}]} \right]}{\sum_j^{N_C} \sqrt{Var[C_j]}} \right)^2 \tag{S.68}$$

where the greater the value of R, the better the measurement saving is by assembling these operators into a particular group.

Next, our analysis diverges from Crawford *et al.*, where we consider groups of anticommuting operators (rather than commuting operators)[12]. First we consider the covariance of two anticommuting Pauli operators.

The amount two random variables vary together (co-vary) is measured by their covariance. Consider the results of random variables x and y, one can obtain a set of M paired measurements:

$$\begin{aligned} \{&(x_0, y_0), \\ &(x_1, y_1), \\ &\ldots, \\ &(x_{M-1}, y_{M-1})\}. \end{aligned} \tag{S.69}$$

A positive covariance indicates that higher than average values of one variable tend to be paired with higher than average values of the other variable. A negative covariance indicates that a higher than average value of one variable tend to be paired with lower than average values of the other. If two random variables are independent, then their

covariance will be zero. However, a covariance of zero does not mean two random variables are independent, as nonlinear relationships can result in a covariance of zero.

In the context of measuring a quantum state in the Pauli basis on a quantum computer this would be a set of paired single shot samples $\{s_i^a, s_i^b | i = 0, 1, \ldots, N-1\}$, where $s_i^a, s_i^b \in \{-1, +1\}$. Experimentally, each pair is the (single shot) measurement outcome for P_a followed by the (single shot) measurement outcome for P_b. Taking simultaneous projective measurements, without re-preparing the quantum state is a meaningful operation if the operators share a common eigenbasis. The order of measurement does not effect measurement outcomes, but the paired samples will be statistically correlated and have a certain covariance. However, for anticommuting Pauli operators this is not the case, as these operators do not share a common eigenbasis. Projective measurement means the expectation value of these operators cannot be known simultaneously. We consider the covariance in this scenario.

Consider the the spectral decomposition of two anticommuting Pauli operators $\{P_a, P_b\}$:

$$P_a = +1 \left|\kappa_0\right\rangle \left\langle\kappa_0\right| - 1 \left|\kappa_1\right\rangle \left\langle\kappa_1\right|, \tag{S.70}$$

and

$$P_b = +1 \left|\Omega_0\right\rangle \left\langle\Omega_0\right| - 1 \left|\Omega_1\right\rangle \left\langle\Omega_1\right|, \tag{S.71}$$

where for Pauli operators:

$$\left\langle\kappa_0 \middle| \kappa_0\right\rangle = \left\langle\kappa_1 \middle| \kappa_1\right\rangle = \left\langle\Omega_0 \middle| \Omega_0\right\rangle = \left\langle\Omega_1 \middle| \Omega_1\right\rangle = 1, \tag{S.72a}$$

$$\left\langle\kappa_0 \middle| \kappa_1\right\rangle = \left\langle\Omega_0 \middle| \Omega_1\right\rangle = 0, \tag{S.72b}$$

$$\left|\left\langle\kappa_0 \middle| \Omega_0\right\rangle\right|^2 = \left|\left\langle\kappa_0 \middle| \Omega_1\right\rangle\right|^2 = \left|\left\langle\kappa_1 \middle| \Omega_0\right\rangle\right|^2 = \left|\left\langle\kappa_1 \middle| \Omega_1\right\rangle\right|^2 = 0.5. \tag{S.72c}$$

Without loss of generality, assume P_a is measured first on a general normalized quantum state $|\psi\rangle = \gamma |\kappa_0\rangle + \delta |\kappa_1\rangle$. The only possible post measurement outcomes are $|\kappa_0\rangle$ or $|\kappa_1\rangle$, with probabilities $|\gamma|^2$ or $|\delta|^2$ respectively. Consider the result of subsequently measuring P_b. The expectation value in each scenario will be:

$$\begin{aligned}
\left\langle\kappa_0\right| P_b \left|\kappa_0\right\rangle &= \left\langle\kappa_0\right| \big(\left|\Omega_0\right\rangle \left\langle\Omega_0\right| - \left|\Omega_1\right\rangle \left\langle\Omega_1\right| \big) \left|\kappa_0\right\rangle \\
&= \left\langle\kappa_0 \middle| \Omega_0\right\rangle \left\langle\Omega_0 \middle| \kappa_0\right\rangle - \left\langle\kappa_0 \middle| \Omega_1\right\rangle \left\langle\Omega_1 \middle| \kappa_0\right\rangle \\
&= \left|\left\langle\kappa_0 \middle| \Omega_0\right\rangle\right|^2 - \left|\left\langle\kappa_0 \middle| \Omega_1\right\rangle\right|^2 \\
&= \underbrace{0.5}_{\mathbb{P}(\Omega_0|\kappa_0)} - \underbrace{0.5}_{\mathbb{P}(\Omega_1|\kappa_0)} = 0
\end{aligned} \tag{S.73a}$$

$$\begin{aligned}
\left\langle\kappa_1\right| P_b \left|\kappa_1\right\rangle &= \left\langle\kappa_1\right| \big(\left|\Omega_0\right\rangle \left\langle\Omega_0\right| - \left|\Omega_1\right\rangle \left\langle\Omega_1\right| \big) \left|\kappa_1\right\rangle \\
&= \left\langle\kappa_1 \middle| \Omega_0\right\rangle \left\langle\Omega_0 \middle| \kappa_1\right\rangle - \left\langle\kappa_1 \middle| \Omega_1\right\rangle \left\langle\Omega_1 \middle| \kappa_1\right\rangle \\
&= \left|\left\langle\kappa_1 \middle| \Omega_0\right\rangle\right|^2 - \left|\left\langle\kappa_1 \middle| \Omega_1\right\rangle\right|^2 \\
&= \underbrace{0.5}_{\mathbb{P}(\Omega_0|\kappa_1)} - \underbrace{0.5}_{\mathbb{P}(\Omega_1|\kappa_1)} = 0
\end{aligned} \tag{S.73b}$$

Overall, we find the probabilities of all possible combinations of measurement outcomes to be:

$$\begin{aligned}
\mathbb{P}\big(P_b = |\Omega_0\rangle \,|P_a = |\kappa_0\rangle\big) = \mathbb{P}\big(P_b = |\Omega_1\rangle \,|P_a = |\kappa_0\rangle\big) = 0.5 \\
\mathbb{P}\big(P_b = |\Omega_0\rangle \,|P_a = |\kappa_1\rangle\big) = \mathbb{P}\big(P_b = |\Omega_1\rangle \,|P_a = |\kappa_1\rangle\big) = 0.5
\end{aligned} \tag{S.74}$$

This result shows that the probability of obtaining $|\Omega_0\rangle$ or $|\Omega_1\rangle$ is not affected by the probability of obtaining $|\kappa_0\rangle$ or $|\kappa_1\rangle$ in the first measurement. The variables are therefore statistically independent [1]. We find the covariance of P_a and P_b, where $\{P_a, P_b\} = 0$, to be:

---

[1] This analysis is strictly for the case of subsequent measurement of anticommuting Pauli operators.

$$
\begin{aligned}
Cov[P_a, P_b] &= \mathbb{E}\left[\Big(p_a - \langle P_a\rangle\Big)\Big(p_b - \langle P_b\rangle\Big)\right] \\
&= \mathbb{E}\left[\Big(p_a p_b - p_a\langle P_b\rangle - \langle P_a\rangle p_b + \langle P_a\rangle\langle P_b\rangle\Big)\right] \\
&= \Big(\mathbb{E}[p_a p_b] - \mathbb{E}[p_a\langle P_b\rangle] - \mathbb{E}[\langle P_a\rangle p_b] + \mathbb{E}[\langle P_a\rangle\langle P_b\rangle]\Big) \\
&= \langle P_a P_b\rangle - \langle P_a\rangle\langle P_b\rangle - \langle P_a\rangle\langle P_b\rangle + \langle P_a\rangle\langle P_b\rangle \\
&= \langle P_a P_b\rangle - \langle P_a\rangle\langle P_b\rangle \\
&= \langle P_a\rangle\langle P_b\rangle - \langle P_a\rangle\langle P_b\rangle = 0
\end{aligned}
\tag{S.75}
$$

where under independence: $\langle P_a P_b\rangle = \langle P_a\rangle\langle P_b\rangle$. Intuitively, this result makes sense. The projective measurement of the first Pauli operator maximally randomizes the expectation value of the other Pauli operator and thus the covariance will be zero. Interestingly, the projective measurement causes the underlying distribution of the quantum state to change and so subsequent measurements generating paired samples are not well defined in this setting (for anticommuting operators). This phenomenon is not present in classical experiments. However, the same statistical analysis can be done if we just take pairs of subsequent measurements and only do a statistical analysis on these random variables. We note that our analysis did not have to account for $\langle P_a P_b\rangle$ not being a valid observable, as for anticommuting Pauli operators this operator is not Hermitian.

Given the covariance of two anticommuting Pauli operators is zero, we find the variance of a normalized anticommuting clique $\gamma_j C_j$ to be:

$$
\begin{aligned}
Var[\gamma_j C_j] = \gamma_j^2 Var[C_j] = \gamma_j^2 Var[\sum_i^{|C_j|}\frac{c_i}{\gamma_j}P_i] &= \gamma_j^2\sum_i^{|C_j|}\sum_k^{|C_j|} Cov[\frac{c_i}{\gamma_j}P_i, \frac{c_k}{\gamma_j}P_k] \\
&= \gamma_j^2\sum_i^{|C_j|}\frac{c_i^2}{\gamma_j^2}Var[P_i] + \gamma_j^2\sum_i^{|C_j|}\sum_{\substack{k\\ \forall k\neq i}}^{|C_j|}\frac{c_i}{\gamma_j}\frac{c_k}{\gamma_j}\underbrace{Cov[P_i,P_k]}_{=0} \\
&= \sum_i^{|C_j|} c_i^2 Var[P_i]
\end{aligned}
\tag{S.76}
$$

We use this to obtain the following R ratio (equation S.68):

$$
\begin{aligned}
R = \frac{M_u}{M_g} &= \left(\frac{\sum_j^{N_C}\left[\sum_{k=0}^{|C_j|-1}\sqrt{Var[c_k^{(j)}P_k^{(j)}]}\right]}{\sum_j^{N_C}\sqrt{Var[\gamma_j C_j]}}\right)^2 \\
&= \left(\frac{\sum_j^{N_C}\left[\sum_{k=0}^{|C_j|-1}|c_k^{(j)}|\sqrt{Var[P_k^{(j)}]}\right]}{\sum_j^{N_C}\sqrt{\sum_{k=0}^{|C_j|-1}|c_k^{(j)}|^2 Var[P_k^{(j)}]}}\right)^2 \\
&= \left(\frac{\sum_j^{N_C}\left[\sum_{k=0}^{|C_j|-1}|x_k^{(j)}|\right]}{\sum_j^{N_C}\sqrt{\sum_{k=0}^{|C_j|-1}|x_k^{(j)}|^2}}\right)^2 \\
&= \left(\frac{\sum_j^{N_C}\|\vec{x}_j\|_1}{\sum_j^{N_C}\|\vec{x}_j\|_2}\right)^2
\end{aligned}
\tag{S.77}
$$

where $x_k^{(j)} = |c_k^{(j)}|\sqrt{Var[P_k^{(j)}]}$ and $\vec{x}_j = (x_0^{(j)}, x_1^{(j)}, \ldots, x_{|C_j|-1}^{(j)})$. Minkowski inequality ensures $\|\vec{x}_j\|_2 \leq \|\vec{x}_j\|_1$. At worst unitary partitioning will achieve the same number of measurements as no grouping and will more often achieve an improvement. However, we can actually bound the improvement in general as:

$$\|u\|_1 = \sum_i^n |u_i| = \sum_i^n |u_i| \cdot 1 \leq \left(\sum_i^n |u_i|^2 \right)^{0.5} \cdot \left(\sum_i^n 1^2 \right)^{0.5} = \sqrt{n} \|u\|_2 \tag{S.78}$$

where the Cauchy-Schwarz inequality has been utilized. Overall, we find $\|u\|_2 \leq \|u\|_1 \leq \sqrt{n}\|u\|_2$ and thus:

$$1 \leq R = \frac{M_u}{M_g} = \left(\frac{\sum_j^{N_C} \|\vec{x}_j\|_1}{\sum_j^{N_C} \|\vec{x}_j\|_2} \right)^2 \leq \left(\frac{\sum_j^{N_C} \sqrt{|C_j|} \cdot \|\vec{x}_j\|_2}{\sum_j^{N_C} \|\vec{x}_j\|_2} \right)^2. \tag{S.79}$$

III. NUMERICAL DETAILS OF THE TOY EXAMPLE

This section provides all the details for the Toy problem described in Section III B. The full noncontextual ground state is:

$$(\underbrace{-1, +1, -1}_{\vec{q}_0}, \underbrace{0.25318483, -0.65828059, -0.70891756}_{\vec{r}_0}). \tag{S.80}$$

This defines the $A(\vec{r}_0)$:

$$A(\vec{r}_0) = 0.25318483\, YXYI - 0.65828059\, XYXI - 0.70891756\, XZXI. \tag{S.81}$$

The operators to map $A(\vec{r}_0)$ to a single Pauli operator are:

$$R_S = \; e^{+1i \cdot -0.7879622757719398 \cdot ZYZI} \cdot e^{+1i \cdot 1.2036225088338255 \cdot ZZZI}, \tag{S.82}$$

and

$$R_{LCU} = \; 0.79157591\, IIII + 0.41580383i\, ZZZI - 0.44778874i\, ZYZI. \tag{S.83}$$

Their action results in: $R_S A(\vec{r}_0) R_S^\dagger = R_{LCU} A(\vec{r}_0) R_{LCU}^\dagger = YXYI$.

We then defined U depending on which generators we wish to fix. We found the optimal ordering of stabilizers (supplied in Equation 24) to fix via a brute force search over all $\sum_{i=1}^{|\mathcal{W}_{all}|} \binom{|\mathcal{W}_{all}|}{i} = 2^4 - 1 = 15$ possibilities for $\mathcal{W}$. The following optimal ordering was obtained:

1. $\{-1\, IIIZ\}$
2. $\{+1\, IXYI, -1\, IIIZ\}$
3. $\{+1\, IXYI, -1\, IIIZ, +1\, \mathcal{A}(\vec{r}_0)\}$
4. $\{-1\, YIYI, +1\, IXYI, -1\, IIIZ, +1\, \mathcal{A}(\vec{r}_0)\}$.

This defines all the information required to implement CS-VQE. Table S.1 summarises the stabilizers fixed, the rotation $U_{\mathcal{W}}$, required projection $Q_{\mathcal{W}}$ and final projected Hamiltonian $Q_{\mathcal{W}}^\dagger U_{\mathcal{W}}^\dagger H U_{\mathcal{W}} Q_{\mathcal{W}}$ for this ordering.

The old approach of applying $U_{\mathcal{W}_{all}}^\dagger H U_{\mathcal{W}_{all}}$ and then fixing certain stabilizer eigenvalues are summarised in Table S.2. It can be seen from these results, that always implementing the unitary partitioning rotation R can unnecessarily increase the number of terms in the Hamiltonian and thus should only be applied if the eigenvalue for $\langle A(\vec{r}) \rangle$ is fixed.

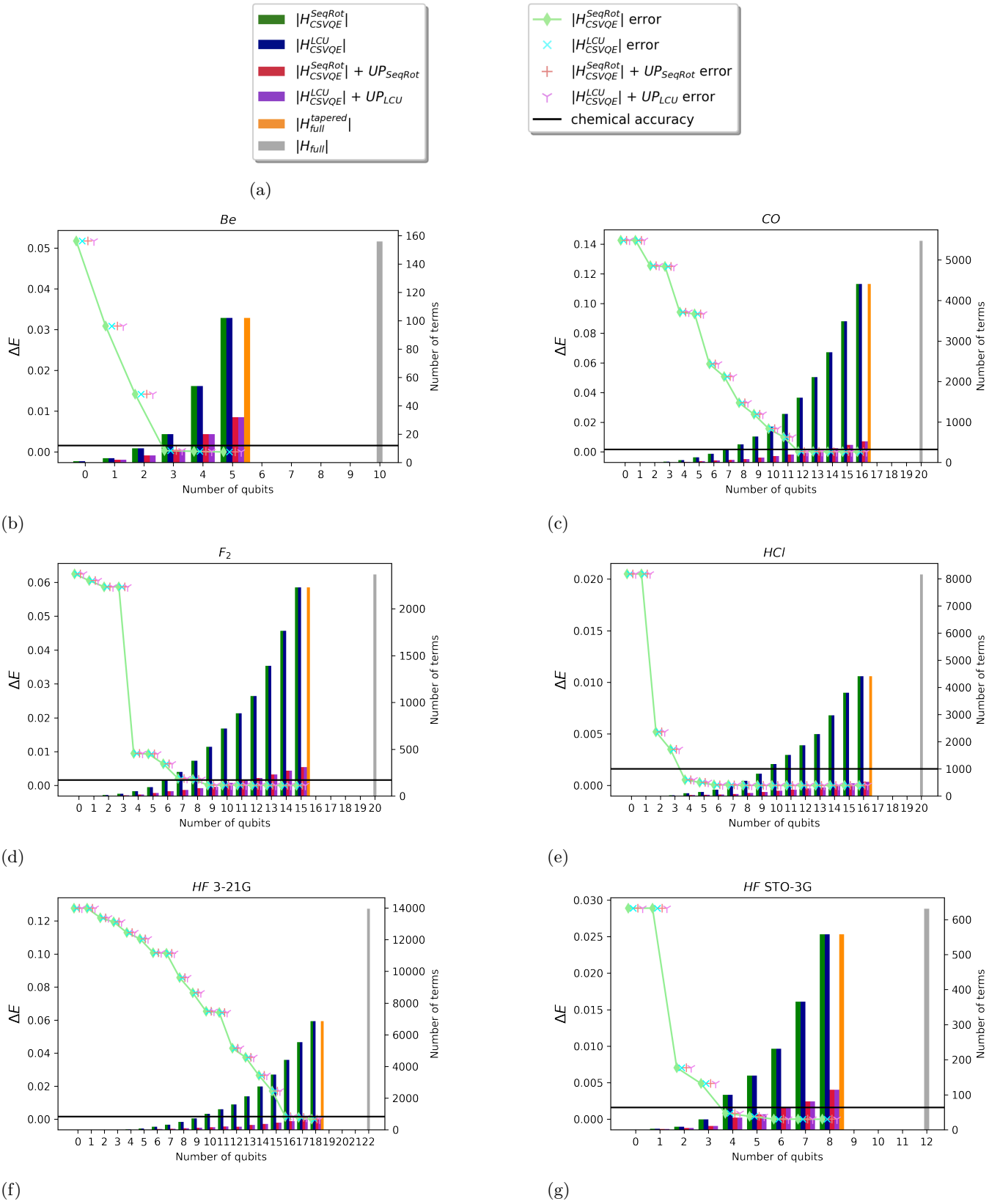
$\mathcal{W}$	$U_{\mathcal{W}}^{\dagger}$	$\mathcal{W}^Z = U_{\mathcal{W}}^{\dagger}\mathcal{W}U_{\mathcal{W}}$	$Q_{\mathcal{W}}$	$Q_{\mathcal{W}}^{\dagger}U_{\mathcal{W}}^{\dagger}HU_{\mathcal{W}}Q_{\mathcal{W}}$
-1 YIYI +1 IXYI -1 IIIZ $\mathcal{A}(\vec{r}_0)$	$e^{1i\frac{\pi}{2}XIYI}e^{1i\frac{\pi}{2}IYYI}R_{S/LCU}$	- ZIII + IZII - IIIZ + IIZI	$\|1\rangle\langle 1\| \otimes \|0\rangle\langle 0\| \otimes \|1\rangle\langle 1\| \otimes \|0\rangle\langle 0\|$	-2.475+0.000j
+1 IXYI -1 IIIZ $\mathcal{A}(\vec{r}_0)$	$e^{1i\frac{\pi}{2}IYYI}R_{S/LCU}$	+ IZII - IIIZ + IIZI	$I \otimes \|0\rangle\langle 0\| \otimes \|0\rangle\langle 0\| \otimes \|1\rangle\langle 1\|$	SeqRot: -1.827+0.000j I + -0.198+0.000j X + -0.467+0.000j Z + 0.648+0.000j Y LCU: -1.827+0.000j I + -0.414+0.000j X + -0.292+0.000j Z + 0.648+0.000j Y
+1 IXYI -1 IIIZ	$e^{1i\frac{\pi}{2}IYYI}$	+ IZII - IIIZ	$I \otimes \|0\rangle\langle 0\| \otimes I \otimes \|1\rangle\langle 1\|$	-0.500+0.000j II + 0.500+0.000j XI + 0.700+0.000j XX + 0.100+0.000j YI + -0.100+0.000j YX + 1.300+0.000j XZ + 0.600+0.000j IY + 0.700+0.000j ZZ
-1 IIIZ	$IIII$	-IIIZ	$I \otimes I \otimes I \otimes \|1\rangle\langle 1\|$	-0.500+0.000j III + 0.100+0.000j XXX + 0.200+0.000j YXX + 0.700+0.000j XZX + 0.700+0.000j XYX + 0.100+0.000j YZX + 0.200+0.000j XXZ + 0.600+0.000j IIY + 0.500+0.000j XXY + 0.100+0.000j YXY + 0.600+0.000j XZZ + 0.700+0.000j ZZZ + 0.200+0.000j YYZ + 0.100+0.000j ZYY

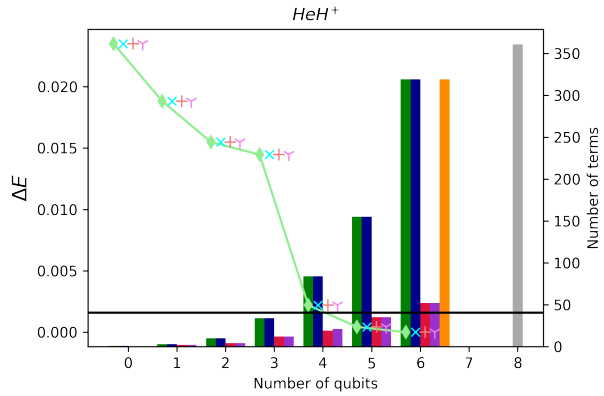
TABLE S.1: Different contextual subspace Hamiltonians defined from H (Equation 15). R_S and R_{LCU} are defined in Equations S.82 and S.83.

$H_{SeqRot} = U^{\dagger,SeqRot}_{\mathcal{W}_{all}} H U^{SeqRot}_{\mathcal{W}_{all}}$	$H_{LCU} = U^{\dagger,LCU}_{\mathcal{W}_{all}} H U^{LCU}_{\mathcal{W}_{all}}$	$\mathcal{W}^Z$	$Q_{\mathcal{W}}$	$Q^\dagger_{\mathcal{W}} H_{SeqRot} Q_{\mathcal{W}}$	$Q^\dagger_{\mathcal{W}} H_{LCU} Q_{\mathcal{W}}$
0.932-0.000j ZIII + -0.056+0.000j YIII + -0.025+0.000j ZXII + -0.025+0.000j YXII + 0.057-0.000j ZIXI + -0.197+0.000j YIXI + -0.051+0.000j ZXXI + 0.051+0.000j YXXI + 0.560-0.000j XZII + 0.395+0.000j ZZII + 0.397-0.000j YZII + 0.141-0.000j ZZXI + 0.142-0.000j YZXI + 0.345-0.000j XIZI + 0.093-0.000j XXZI + 0.467-0.000j IIYI + -0.187+0.000j IXYI + -0.496+0.000j ZIZI + 0.494-0.000j YIZI + 0.215+0.000j XZZI + -0.200+0.000j IYZI + -0.152+0.000j ZZZI + -0.153+0.000j YZZI + 0.071+0.000j ZYYI + 0.071-0.000j YYYI + -0.500+0.000j IIIZ	0.261-0.000j XIII + 0.932-0.000j ZIII + -0.230+0.000j YIII + -0.025+0.000j ZXII + -0.071+0.000j YXII + 0.295-0.000j ZIXI + -0.197+0.000j YIXI + -0.142+0.000j ZXXI + 0.051+0.000j YXXI + 0.395-0.000j XZII + 0.037-0.000j IYXI + 0.395-0.000j ZZII + 0.223-0.000j YZII + 0.120-0.000j ZZXI + 0.142-0.000j YZXI + 0.263-0.000j XIZI + 0.066-0.000j XXZI + 0.366-0.000j IIYI + -0.132+0.000j IXYI + -0.496+0.000j ZIZI + 0.419-0.000j YIZI + 0.393+0.000j XZZI + -0.200+0.000j IYZI + -0.074+0.000j IZYI + -0.152+0.000j ZZZI + -0.425+0.000j YZZI + 0.060+0.000j ZYYI + 0.071-0.000j YYYI + -0.500+0.000j IIIZ	- ZIII + IZII - IIIZ + IIZI	$\|1\rangle\langle 1\| \otimes \|0\rangle\langle 0\| \otimes \|1\rangle\langle 1\| \otimes \|0\rangle\langle 0\|$	-2.475+0.000j	-2.475+0.000j
		+ IZII - IIIZ + IIZI	$I \otimes \|0\rangle\langle 0\| \otimes \|0\rangle\langle 0\| \otimes \|1\rangle\langle 1\|$	-1.827+0.000j I + -0.198+0.000j X + 0.648+0.000j Z + 0.467+0.000j Y	-1.827+0.000j I + -0.414+0.000j X + 0.648+0.000j Z + 0.292+0.000j Y
		+ IZII - IIIZ	$I \otimes \|0\rangle\langle 0\| \otimes I \otimes \|1\rangle\langle 1\|$	-0.500+0.000j II + 0.560+0.000j XI + 1.327+0.000j ZI + 0.341+0.000j YI + 0.198+0.000j ZX + -0.056+0.000j YX + 0.560+0.000j XZ + 0.467+0.000j IY + -0.648+0.000j ZZ + 0.341+0.000j YZ	-0.500+0.000j II + 0.656+0.000j XI + 1.327+0.000j ZI + -0.006+0.000j YI + 0.414+0.000j ZX + -0.056+0.000j YX + 0.656+0.000j XZ + 0.292+0.000j IY + -0.648+0.000j ZZ + -0.006+0.000j YZ
		-IIIZ	$I \otimes I \otimes I \otimes \|1\rangle\langle 1\|$	-0.500+0.000j III + 0.932+0.000j ZII + -0.056+0.000j YII + -0.025+0.000j ZXI + -0.025+0.000j YXI + 0.057+0.000j ZIX + -0.197+0.000j YIX + -0.051+0.000j ZXX + 0.051+0.000j YXX + 0.560+0.000j XZI + 0.395+0.000j ZZI + 0.397+0.000j YZI + 0.141+0.000j ZZX + 0.142+0.000j YZX + 0.345+0.000j XIZ + 0.093+0.000j XXZ + 0.467+0.000j IIY + -0.187+0.000j IXY + -0.496+0.000j ZIZ + 0.494+0.000j YIZ + 0.215+0.000j XZZ + -0.200+0.000j IYZ + -0.152+0.000j ZZZ + -0.153+0.000j YZZ + 0.071+0.000j ZYY + 0.071+0.000j YYY	-0.500+0.000j III + 0.261+0.000j XII + 0.932+0.000j ZII + -0.230+0.000j YII + -0.025+0.000j ZXI + -0.071+0.000j YXI + 0.295+0.000j ZIX + -0.197+0.000j YIX + -0.142+0.000j ZXX + 0.051+0.000j YXX + 0.395+0.000j XZI + 0.037+0.000j IYX + 0.395+0.000j ZZI + 0.223+0.000j YZI + 0.120+0.000j ZZX + 0.142+0.000j YZX + 0.263+0.000j XIZ + 0.066+0.000j XXZ + 0.366+0.000j IIY + -0.132+0.000j IXY + -0.496+0.000j ZIZ + 0.419+0.000j YIZ + 0.393+0.000j XZZ + -0.200+0.000j IYZ + -0.074+0.000j IZY + -0.152+0.000j ZZZ + -0.425+0.000j YZZ + 0.060+0.000j ZYY + 0.071+0.000j YYY

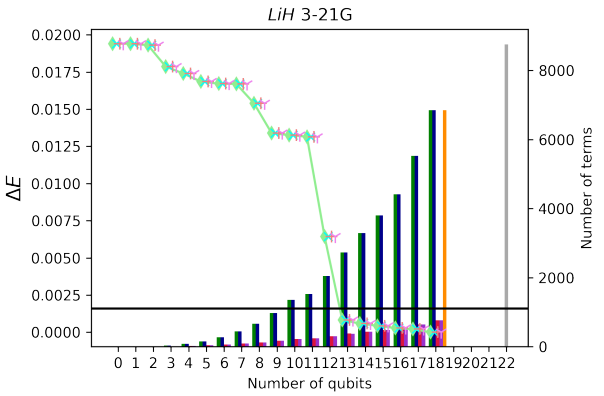
TABLE S.2: Different contextual subspace Hamiltonians defined from H (Equation 15). Here $\mathcal{W}$ has been set to $\mathcal{W}_{all}$, which defines $U^\dagger_{\mathcal{W}_{all}} = e^{1i\frac{\pi}{2}XIYI} e^{1i\frac{\pi}{2}IYYI} R_{S/LCU}$. R_S and R_{LCU} are defined in Equations S.82 and S.83. The two left columns (H_{SeqRot} and H_{LCU}) give H rotated by $U_{\mathcal{W}}$. Each projected Hamiltonian is generated from these, where the eigenvalue of certain stabilizers are fixed according to the projector $Q_{\mathcal{W}}$. For the last two rows, the eigenvalue of $\mathcal{A}(\vec{r}_0)$ has not been fixed, but the non-Clifford operator $R_{S/LCU}$ is still included within $U^\dagger_{\mathcal{W}_{all}}$. This leads to an unnecessary increase in the number of Pauli operators for these two cases, as these transformed operators are isospectral with associated Hamiltonians in Table S.1.

IV. GRAPHICAL RESULTS FOR CS-VQE SIMULATION OF EACH MOLECULAR HAMILTONIAN

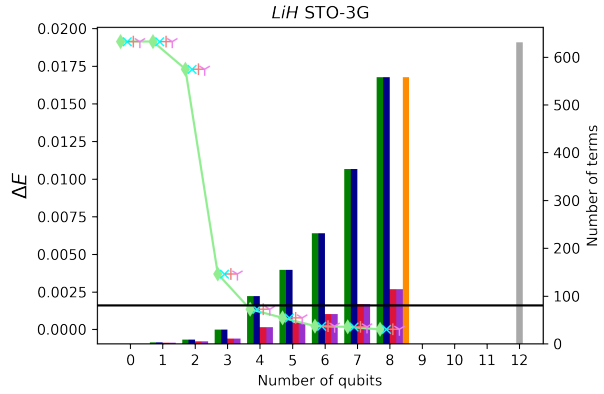



HeH+
ΔE
Number of terms
Number of qubits

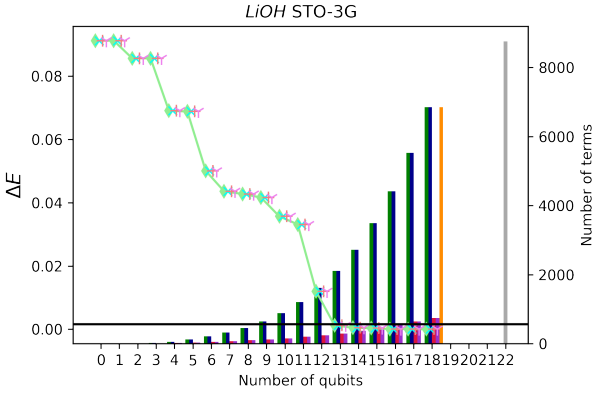
(h)


LiH 3-21G
ΔE
Number of terms
Number of qubits

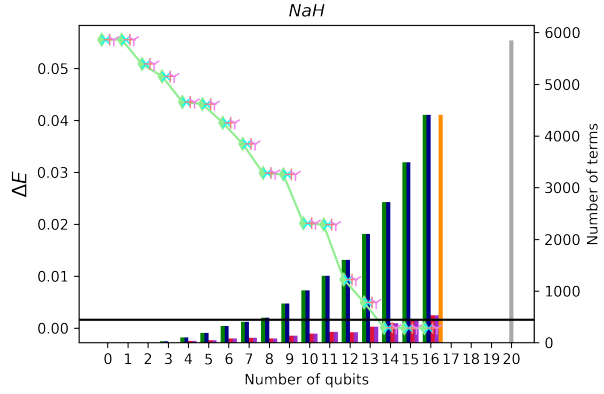
(i)


LiH STO-3G
ΔE
Number of terms
Number of qubits

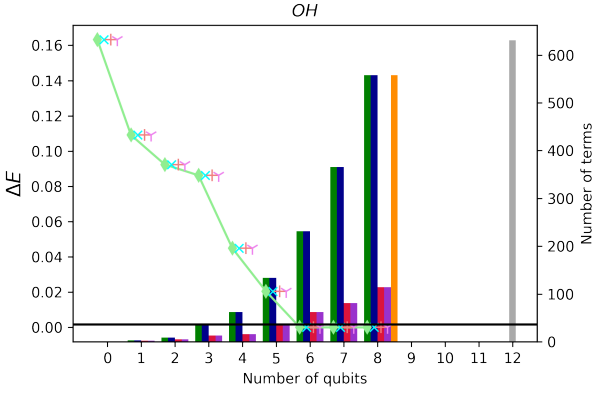
(j)


LiOH STO-3G
ΔE
Number of terms
Number of qubits

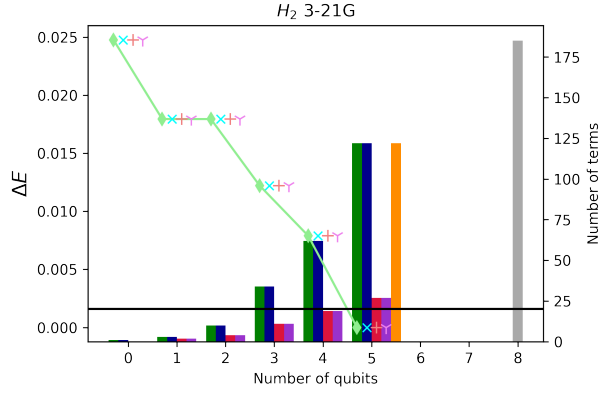
(k)


NaH
ΔE
Number of terms
Number of qubits

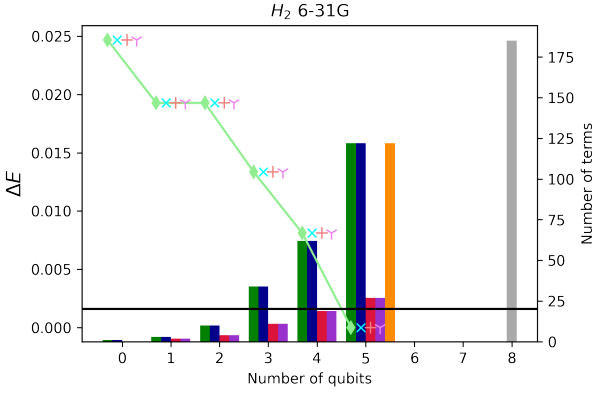
(l)


OH
ΔE
Number of terms
Number of qubits

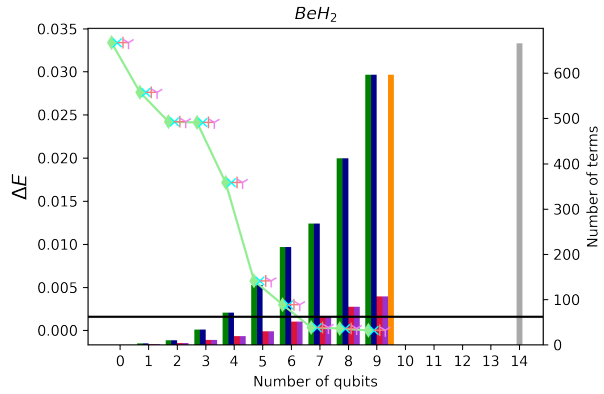
(m)


H2 3-21G
ΔE
Number of terms
Number of qubits

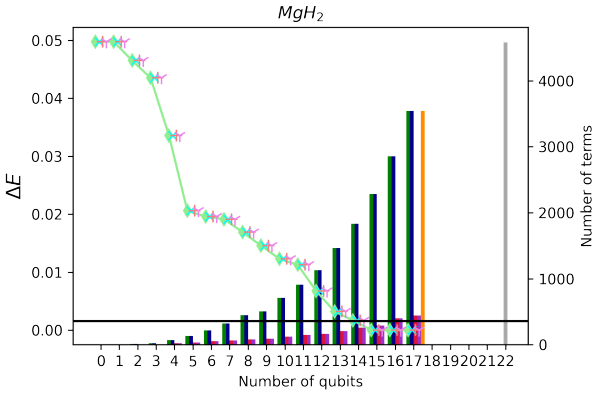
(n)


H2 6-31G
ΔE
Number of terms
Number of qubits

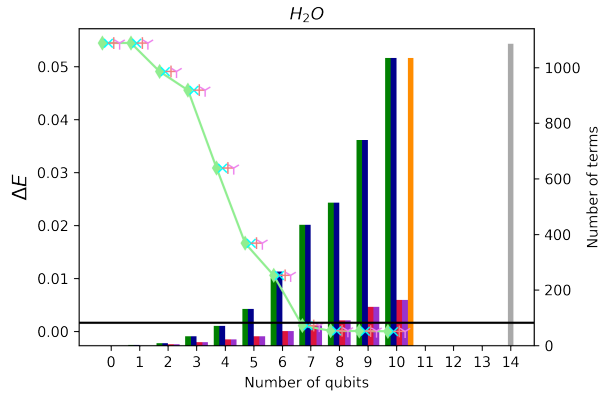
(o)


BeH_2
ΔE
Number of terms
Number of qubits

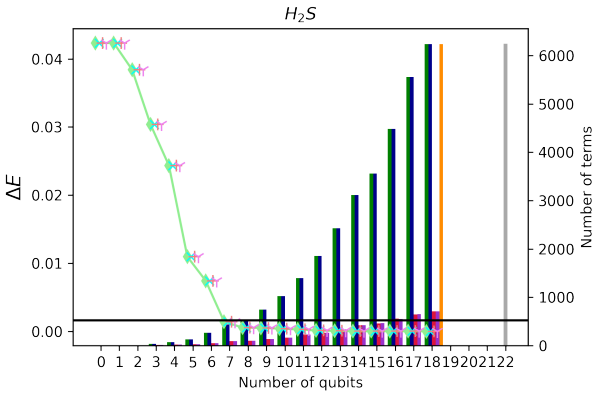
(p)


MgH_2
ΔE
Number of terms
Number of qubits

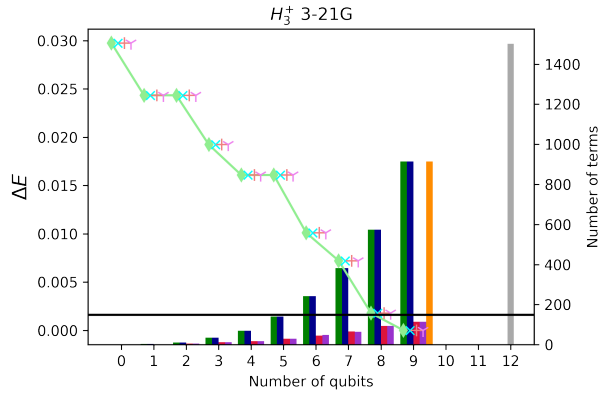
(q)


H_2O
ΔE
Number of terms
Number of qubits

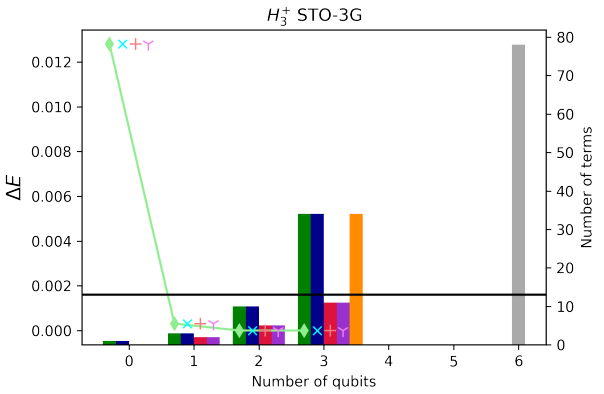
(r)


H_2S
ΔE
Number of terms
Number of qubits

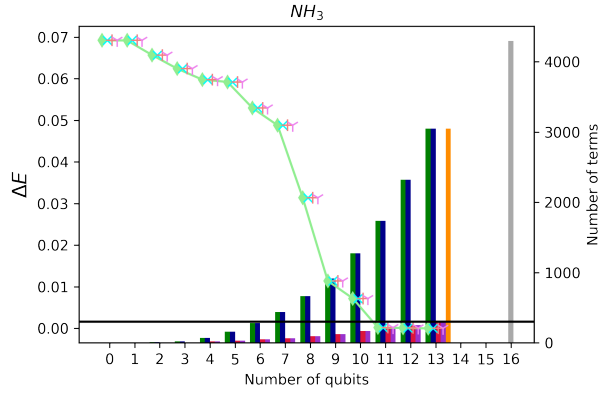
(s)


H_3^+ 3-21G
ΔE
Number of terms
Number of qubits

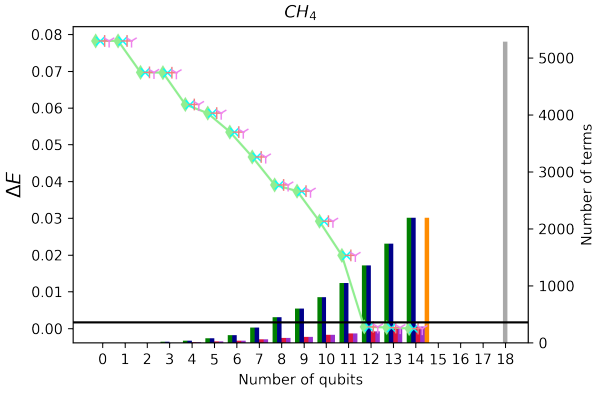
(t)


H_3^+ STO-3G
ΔE
Number of terms
Number of qubits

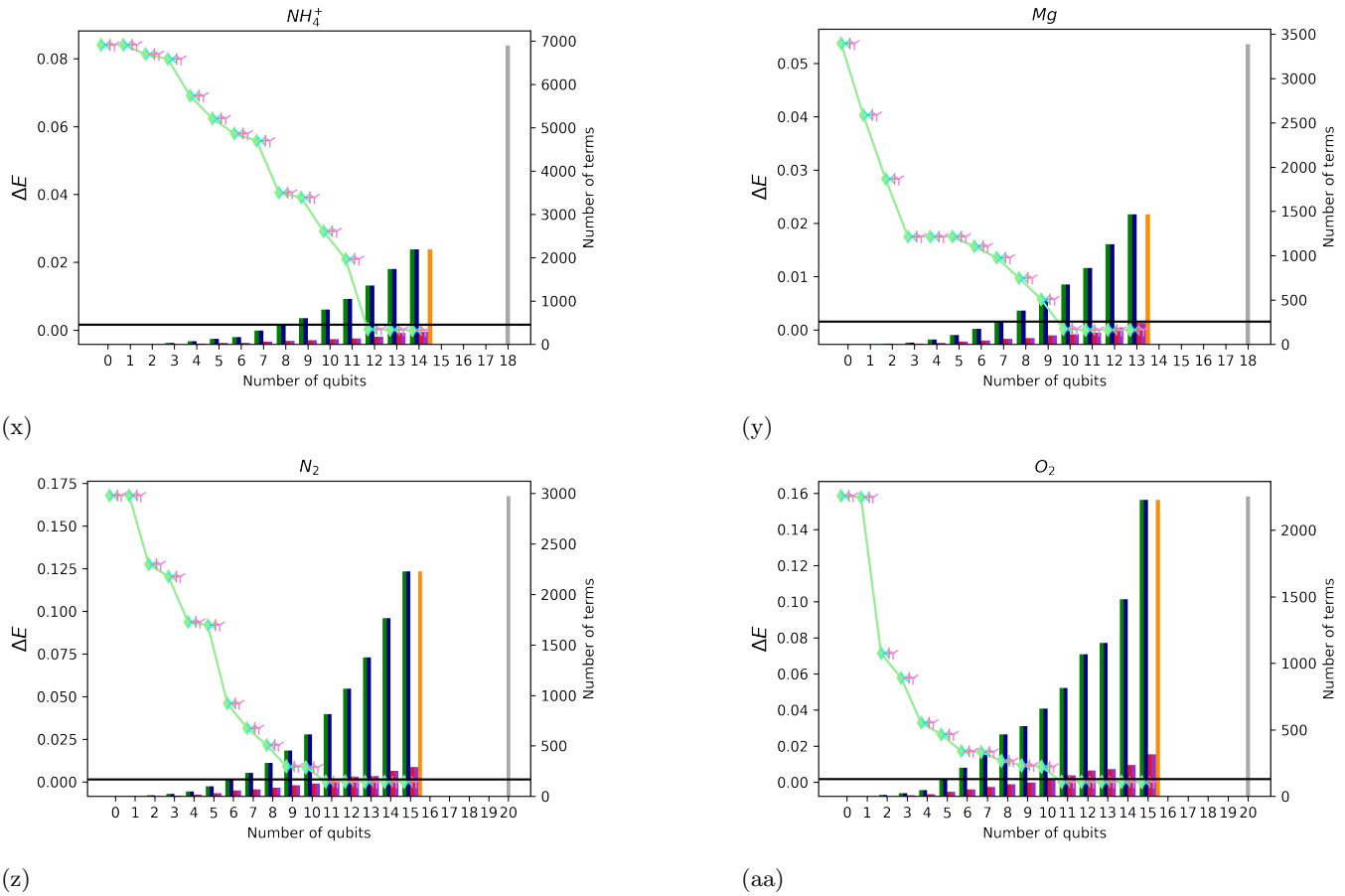
(u)


NH_3
ΔE
Number of terms
Number of qubits

(v)


CH_4
ΔE
Number of terms
Number of qubits

(w)



(x) (y)

(z) (aa)

FIG. S.1: CS-VQE approximation errors ΔE versus number of qubits used on the quantum computer (scatter plot). The horizontal solid black lines indicate chemical accuracy. The number of terms in each Hamiltonian is given by the bar chart.

All the subplots in Figure S.1 give the simulation results of each molecular Hamiltonian at different levels of noncontextual approximations. This is equivalent to how many contextual stabilizers $\mathcal{W}$ eigenvalues are fixed. In each plot, the leftmost data represents the case when all the noncontextual stabilizer eigenvalues are fixed and is the case for the full noncontextual approximation to a given problem [6]. Moving right, we remove a single stabilizer from $\mathcal{W}$ and thus don't fix the eigenvalue of that stabilizer. This reintroduces a qubits worth degree of freedom into the problem. At the limit that no stabilizer eigenvalues are fixed ($\mathcal{W} = \{\}$) we return to standard VQE over the full problem and no noncontextual approximation is made. In each plot this scenario is represented by the far right data point (excluding the data for the full non tapered Hamiltonian that is supplied for reference only). The raw data for these results is supplied in the Supplemental Material (see the zipped file). We include data beyond Hamiltonians achieving chemical accuracy, to show the different possible approximations, rather than stopping once chemical accuracy was achieved.

V. TABULATED RESULTS OF SIMULATION

Table S.3 summarises the numerical results of Figures 2 and 3.

[1] W. M. Kirby, A. Tranter, and P. J. Love, Contextual subspace variational quantum eigensolver, Quantum **5**, 456 (2021).
[2] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurel, Phase-space-simulation method for quantum computation with magic states on qubits, Physical Review A **101**, 012350 (2020).

molecule	basis	$H_{\text{CS-VQE}}$	$H_{\text{CS-VQE}} + UP^{(LCU)}$	$H_{\text{CS-VQE}} + UP^{(SeqRot)}$	H_{tapered}	$RH_{tapered}R^{\dagger}$	H_{full}
BeH_2	STO-3G	(7, 268)	(7, 61)	(7, 61)	(9, 596)	(9, 614)	(14, 666)
Mg	STO-3G	(10, 675)	(10, 114)	(10, 114)	(13, 1465)	(13, 1465)	(18, 3388)
$H_3{}^+$	3-21G	(9, 914)	(9, 115)	(9, 115)	(9, 914)	(9, 786)	(12, 1501)
O_2	STO-3G	(11, 815)	(11, 157)	(11, 157)	(15, 2229)	(15, 2374)	(20, 2255)
OH	STO-3G	(6, 231)	(6, 62)	(6, 62)	(8, 558)	(8, 558)	(12, 631)
CH_4	STO-3G	(12, 1359)	(12, 203)	(12, 203)	(14, 2194)	(14, 2194)	(18, 5288)
Be	STO-3G	(3, 20)	(3, 9)	(3, 9)	(5, 102)	(5, 108)	(10, 156)
NH_3	STO-3G	(11, 1733)	(11, 200)	(11, 200)	(13, 3048)	(13, 2738)	(16, 4293)
H_2S	STO-3G	(7, 435)	(7, 92)	(7, 92)	(18, 6237)	(18, 6237)	(22, 6246)
H_2	3-21G	(5, 122)	(5, 27)	(5, 27)	(5, 122)	(5, 124)	(8, 185)
HF	3-21G	(17, 5530)	(17, 648)	(17, 648)	(18, 6852)	(18, 6852)	(22, 13958)
F_2	STO-3G	(9, 527)	(9, 99)	(9, 99)	(15, 2229)	(15, 2229)	(20, 2367)
HCl	STO-3G	(4, 100)	(4, 35)	(4, 35)	(16, 4409)	(16, 4409)	(20, 8159)
HeH^+	3-21G	(5, 155)	(5, 35)	(5, 35)	(6, 319)	(6, 319)	(8, 361)
MgH_2	STO-3G	(15, 2285)	(15, 289)	(15, 289)	(17, 3540)	(17, 3540)	(22, 4582)
CO	STO-3G	(12, 1599)	(12, 241)	(12, 241)	(16, 4409)	(16, 4409)	(20, 5475)
LiH	STO-3G	(4, 100)	(4, 35)	(4, 35)	(8, 558)	(8, 586)	(12, 631)
N_2	STO-3G	(11, 815)	(11, 153)	(11, 153)	(15, 2229)	(15, 2229)	(20, 2975)
NaH	STO-3G	(14, 2722)	(14, 375)	(14, 375)	(16, 4409)	(16, 4409)	(20, 5851)
H_2O	STO-3G	(7, 435)	(7, 73)	(7, 73)	(10, 1035)	(10, 1035)	(14, 1086)
$H_3{}^+$	STO-3G	(1, 3)	(1, 2)	(1, 2)	(3, 34)	(3, 35)	(6, 78)
LiOH	STO-3G	(13, 2104)	(13, 296)	(13, 296)	(18, 6852)	(18, 6852)	(22, 8758)
LiH	3-21G	(13, 2732)	(13, 375)	(13, 383)	(18, 6852)	(18, 6852)	(22, 8758)
H_2	6-31G	(5, 122)	(5, 27)	(5, 27)	(5, 122)	(5, 124)	(8, 185)
$NH_4{}^+$	STO-3G	(12, 1359)	(12, 176)	(12, 176)	(14, 2194)	(14, 2194)	(18, 6892)
HF	STO-3G	(4, 100)	(4, 35)	(4, 35)	(8, 558)	(8, 558)	(12, 631)

TABLE S.3: Different resource requirements to study different electronic structure Hamiltonians required to achieve chemical accuracy. Each round bracket tuple reports $(n, |H|)$ and gives the number of qubits and terms for each Hamiltonian considered. $RH_{tapered}R^{\dagger}$ describes the effect of the CS-VQE unitary partitioning rotation on the problem Hamiltonian and $H_{\text{CS-VQE}} = Q_{\mathcal{W}}U_{\mathcal{W}}^{\dagger}H_{\text{full}}U_{\mathcal{W}}Q_{\mathcal{W}}^{\dagger}$. The square tuple gives the upper bound of single qubit gates (SQG) and CNOT gates $[SQG, CNOT]$ required to perform R as a sequence of rotations in the unitary partitioning measurement reduction step, based on the largest anticommuting clique - representing the largest possible circuit for R_S. The size of the Hamiltonian for LiH (3-21G singlet) with measurement reduction applied is different for the sequence of rotations and LCU unitary partitioning methods. This is an artifact of the graph colour heuristic finding different anticommuting cliques in the CS-VQE Hamiltonian.

[3] W. M. Kirby and P. J. Love, Contextuality test of the nonclassicality of variational quantum eigensolvers, Physical review letters **123**, 200501 (2019).

[4] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, *et al.*, Openfermion: the electronic structure package for quantum computers, Quantum Science and Technology **5**, 034014 (2020).

[5] D. Wagner, K. Weihe, E. Ihler, D. Wagner, F. Wagner, J. Gustedt, and M. Morvan, A linear time algorithm for edge-disjoint paths in, in *Proc. of the First Annual European Symposium on Algorithms*, Vol. 8 (Springer) pp. 625–712.

[6] W. M. Kirby and P. J. Love, Classical simulation of noncontextual pauli hamiltonians, Physical Review A **102**, 032418 (2020).

[7] M. A. Nielsen and I. L. Chuang, Quantum computation and quantum information: 10th anniversary edition (Cambridge University Press, 2011) pp. 454–459.

[8] R. W. Spekkens, Quasi-quantization: classical statistical theories with an epistemic restriction, in *Quantum Theory: Informational Foundations and Foils* (Springer, 2016) pp. 83–135.

[9] A. Zhao, A. Tranter, W. M. Kirby, S. F. Ung, A. Miyake, and P. J. Love, Measurement reduction in variational quantum algorithms, Physical Review A **101**, 062322 (2020).

[10] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method, Journal of Chemical Theory and Computation **16**, 190 (2020).

[11] A. Ralli, P. J. Love, A. Tranter, and P. V. Coveney, Implementation of measurement reduction for the variational quantum eigensolver, Physical Review Research **3**, 033195 (2021).

[12] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measurement of pauli operators in the presence of finite sampling error, Quantum **5**, 385 (2021).