Supplementary Information for "Variational Quantum Eigensolver with Reduced Circuit Complexity"

Yu Zhang,[1,*] Lukasz Cincio,[1] Christian F. A. Negre,[1] Piotr Czarnik,[1] Patrick J. Coles,[1]
Petr M. Anisimov,[2] Susan M. Mniszewski,[3] Sergei Tretiak,[1,4] and Pavel A. Dub[5]

[1] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA*
[2] *Accelerators and Electrodynamics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[3] *Computer, Computational and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4] *Center for Integrated Nanotechnologies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[5] *Chemistry Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA*

SUPPLEMENTARY SECTION I. CLUSTERING METHODS USED

Three clustering methods were explored for this effort. The details of each are described below.

A. Metis Graph Partitioning on CPU

The MI matrix can be viewed as a graph. This allows for the use of classical graph partitioning methods from the Metis library [1]. The *part_graph* function in the Metis library was used, and the partition weights were also optimized to minimize the cut.

B. MI-based clustering on Quantum Annealer QPU

The MI between qubits can be interpreted as a degree of correlation. Note that when two random variables are fully correlated, their joint entropy is minimal, and the MI is maximal. Moreover, MI can be mapped to what others have coined as the "generalized correlation coefficient", where all higher order correlations are taken into account [2–4]. This generalized correlation coefficient is computed as:

$$r_{ij}^{\mathrm{MI}} = [1 - \mathrm{e}^{-2I_{ij}/d}]^{1/2} \tag{1}$$

Where d is the dimensionality of the problem, and I_{ij} is the Mutual Information matrix elements. Note that, when I_{ij} goes to infinity, r_{ij}^{MI} approaches 1; whereas if I_{ij} tends to zero, r_{ij}^{MI} tends to 0. Taking into account Eq. 2, we see that, by maximizing MI for a constrained set of qubits, we will be maximizing the correlation among that same set. Moreover, maximizing the MI restricted to a certain number of qubits is somehow related to getting the "most central hub," according to node centrality ordering. Albeit for some exceptions, we have verified the latter using the eigenvector centrality metric.

Maximizing the MI restricted to a certain number of qubits can be formulated as a quadratic unconstrained binary optimization (QUBO) problem. This formulation can, in turn, benefit from available alternative computational paradigms tailored to solve these kinds of problems, e.g., Quantum, Digital annealers. The cost function to be maximized can be written as:

$$C = \mathbf{x}(I - \lambda R)\mathbf{x}^t \tag{2}$$

Where the vector $\mathbf{x}$ is a bit-string containing the information of the selected qubits. This is, if the i-th qubit is selected $x_i = 1$, and 0 otherwise. The matrix I contains all the MI elements I_{ij} as defined by Eq. 1 of the main text, and R is the matrix representation of an operator that "restricts" the total number of selected qubits [5]. A penalty constant λ is used to tune the strength of the constraint. Since the annealing methods and devices commonly minimize the cost function, the problem is rather solved by minimizing $-C$ leading to a QUBO matrix equal to $Q = -I + \lambda R$

Matrix R is defined as follows.

$$R_{ij} = \begin{cases} 1 & \text{if} \quad i \neq j \\ 1 - \alpha \times f \times S & \text{if} \quad i = j \end{cases}$$

Here, f is the target fraction of qubits to be selected; S is the total size of the qubits pool; α is a constant that depends on the optimizer (1 or 2 depending on if only the upper triangle vs the full QUBO matrix is used to compute the energy). For all the calculations where this clustering method was used we have set $f = 0.5$, and $\lambda = 2.0$.

An explanation of why the R matrix restricts the set is described next. Consider the case where a general vector $\mathbf{x}$ of length $S = N$ is such that $\sum_k x_k = K$. Now let us compute the energy term of that vector given by matrix R. This is $E_R(\mathbf{x}, \text{such that } ||x||_1 = K) = \mathbf{x}^t R \mathbf{x}$. Then, the off diagonal elements of R will contribute with $K(K-1)$. If we want $f = M/N$, then, the diagonal elements will contribute with $K - \alpha MK$. By summing the two contributions, we have that $E_R(K) = K^2 - \alpha MK = K(K - \alpha M)$. This means that the function will have two minima, the trivial one at $K = 0$, and another at $K = \alpha M$ (αM bits equal to 1).

* zhy@lanl.gov

C. Quantum Community Detection on Quantum Annealer QPU

Classical community detection as graph clustering maximizes modularity to split the MI matrix into clusters that minimize correlation or connectivity between clusters [6]. In quantum community detection (QCD), the problem is represented as a QUBO that can be run on a D-Wave quantum annealer as an NP-complete combinatorial optimization problem.

The symmetric MI matrix is viewed as a graph. The adjacency matrix A is calculated as follows.

$$A_{ij} = \begin{cases} 0 & i = j \\ |I_{ij}| & i \neq j \end{cases}$$

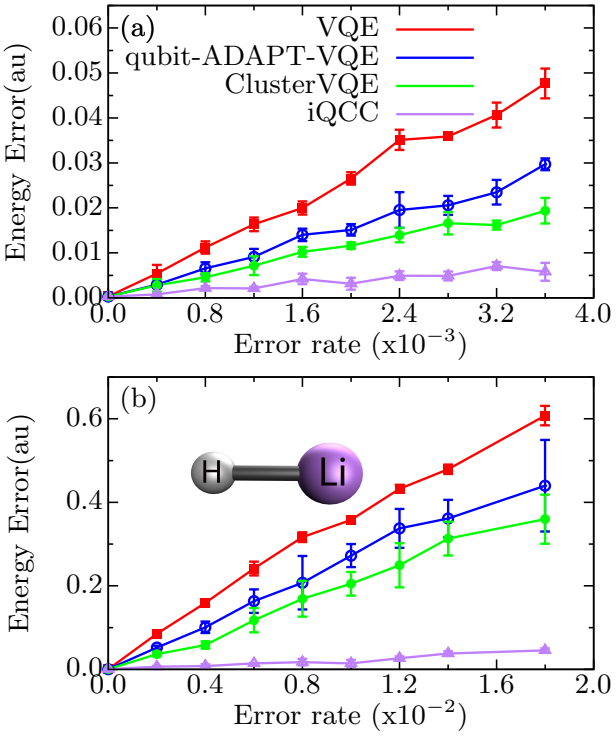
The modularity matrix B is calculated from the adjacency matrix A and the degree of each node (number of connected nodes) d_i.

$$B_{ij} = A_{ij} - \frac{d_i d_j}{2m} = A_{ij} - \frac{d_i d_j}{\sum_i d_i} \tag{3}$$

While QCD can cluster a graph into 2 or more clusters [7–9], in this work, we are interested in splitting only into 2 clusters. The modularity matrix in QUBO form is run on the D-Wave 2000Q quantum annealer:

$$Q(x) = max(x^T B x) \tag{4}$$

The resulting x is a string of zeros and ones that maps each of the MI nodes (representing qubits) to one of two clusters that can be used in the ClusterVQE algorithm.



Supplementary Figure 1. Energy error (ΔE) in the presence of noise. (a) and (b) plot the energy error as a function of the error rates of single and two-qubit gates, respectively. The two and single-qubit gate errors in (a) and (b) are fixed at 0.01 and 0.001, respectively. The single-qubit gate on real quantum devices is usually one order of magnitude smaller than that of two-qubit gates. Hence, the single qubit error has a smaller impact on the results.

SUPPLEMENTARY SECTION II. INTRODUCTION TO UCCSD ANSATZ AND QUBIT-UCCSD

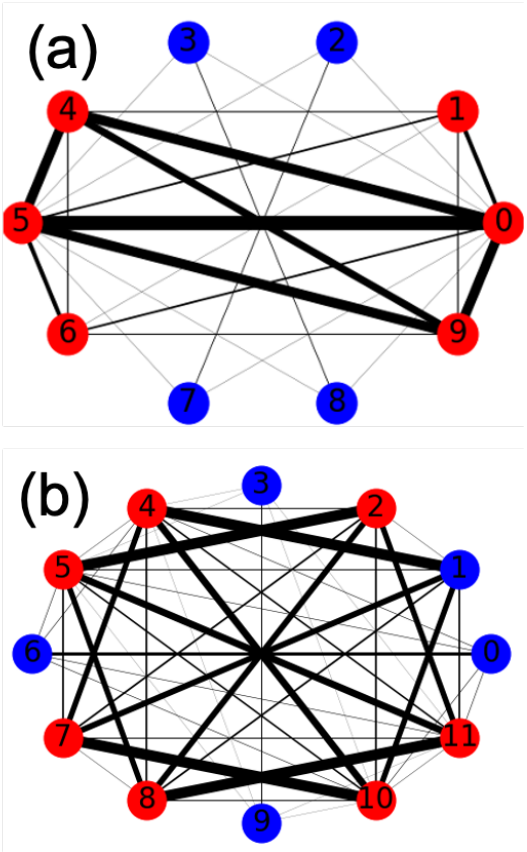
Within the UCCSD ansatz, the correlation beyond the Hartree-Fock reference state is introduced via the single and double excitation (and de-excitation) operators: $|\Phi(\theta)\rangle = e^{\hat{T}(\theta)-\hat{T}^\dagger(\theta)}|\Phi_r\rangle$, where $|\Phi_r\rangle$ is the reference state being the HF state in this work. $\hat{T}(\theta) = \hat{T}_1(\theta_1) + \hat{T}_2(\theta_2)$ is the excitation operator with $\hat{T}_1(\theta_1)$ and $\hat{T}_2(\theta_2)$ being the parameterized single and double excitation operators, respectively,

$$\begin{aligned} \hat{T}_1(\theta) =& \sum_{ia} \theta_i^a \hat{c}_a^\dagger \hat{c}_i \\ \hat{T}_2(\theta) =& \sum_{ij,ab} \theta_{ij}^{ab} \hat{c}_a^\dagger \hat{c}_b^\dagger \hat{c}_i \hat{c}_j. \end{aligned} \tag{5}$$

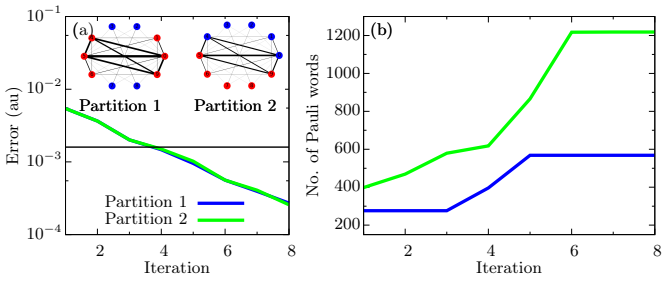
For simplicity, indices $i, j, \cdots$ are used to represent the occupied molecular orbitals (MOs). $a, b, \cdots$ represent the unoccupied MOs. Because the NISQ devices only have gates acting on a few qubits at a time, the UCCSD operator must be broken up into a sequence of single operators (either single or double excitation) via the Trotter expansion of an exponent operator [10], i.e., $e^{A+B} = \lim_{n\to\infty}\left(e^{A/n}e^{B/n}\right)^n$. Even though the truncated Trotter expansion is an approximation to the UCCSD ansatz, the variational procedure is able to absorb the error according to a recent report [11]. Consequently, a finite order of Trotter expansion is able to reproduce the FCI energies due to the fact that the many-body interaction can be decomposed as the tensor product of one- and two-body interactions [12]. In this way, the exact FCI state can be written as the tensor product of single and double excitation operators, $|\Phi(\theta)\rangle = \prod_{k=1}^{D} e^{\hat{t}_k(\theta)}|\Phi_r\rangle$, where $\hat{t}_k(\theta)$ represents either single or double excitation operators and their anti-Hermitian forms ($\hat{t}_k^\dagger = -\hat{t}_k$). Consequently, the GS energy is readily obtained by minimizing the parameterized ansatz, i.e.,

$$\begin{aligned} E_{\rm GS} =& \min_\theta \langle\Phi(\theta)|\hat{H}|\Phi(\theta)\rangle \\ =& \min_\theta \langle\Phi_r| \prod_{k=D}^{1} e^{-\hat{t}_k(\theta_k)} \hat{H} \prod_{k=1}^{D} e^{\hat{t}_k(\theta_k)} |\Phi_r\rangle \end{aligned} \tag{6}$$

In the UCCSD ansatz, the SD sequences are directly generated from the Trotter expansion of the $\hat{T}(\theta) - \hat{T}^\dagger(\theta)$

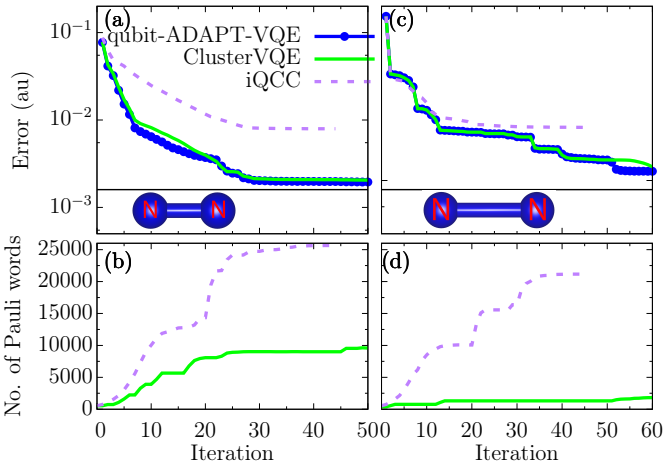


Supplementary Figure 2. Qubit clustering (marked by different colors) of LiH and N_2 molecules. (a) and (b) are for LiH and N_2, respectively. The thickness of edges between different nodes (qubits) denotes the mutual information.



Supplementary Figure 3. The effect of different partitions on the performance of ClusterVQE. Different partitions have tiny effect on the energy, but have large impact on the Hamiltonian size. Less optimal partition (Partition 2 in this example) significantly increases the Hamiltonian size.

operator. In the recently developed ADAPT-VQE algorithm [12], the sequence of SD operators in the ansatz is chosen adaptively by measuring their contribution to the GS energy weighted by the energy gradients with respect to the parameters. With the adaptive method, a more compact sequence of the unitary operators is generated, and a smaller number of parameters is optimized at the cost of more VQE simulations. Because the UCCSD ansatz is not hardware-efficient, a qubit version of the UCCSD ansatz was developed recently in which the ansatz is broken up into a series of Pauli words, and a small portion of the Pauli words is chosen to adaptively grow the ansatz. Eq. 6 in the main text indicates the product $\hat{P}_{kj} = [\hat{P}_k, \hat{P}_j]$, that only contains the $\hat{Z}$ operators, has non-vanishing derivatives because $\langle\Phi_r|\,\hat{\sigma}_i\,|\Phi_r\rangle = 0$ if $\hat{\sigma}_i = \hat{X}, \hat{Y}$. Therefore, the Pauli word $\hat{P}_j$ that contains $\hat{Z}$ has no contribution to the derivative. And two Pauli words with the same flip indices (indices of $\hat{X}/\hat{Y}$) have the same contribution. Hence, after the fermion-qubit mapping of the UCCSD excitation operators, we only include those Pauli words with unique flip indices. Thus, only one of these Pauli words obtained from the fermion-qubit mapping is included in the unitary operators $\{\hat{U}_j\}$. In this way, the number of operators and circuit depth can be reduced. The resulting ansatz is named as QUCCSD ansatz. However, we found that, for strongly correlated molecules, QUCCSD ansatz may result in slow convergence or trapped in a local minimum. To improve the convergence, one more adjoint Pauli word from each UCCSD excitation operator could be added, i.e., the resulting ansatz is named QUCCSD2.



Supplementary Figure 4. Simulation of N_2 molecule with QUCCSD ansatz. (a, c) GS Energy errors and (b, d) number of Pauli words per each iteration for the N_2 molecule with different bond lengths, 1.09 Åand 1.6 Å, obtained from the qubit-ADAPT-VQE (blue), iQCC (purple) and ClusterVQE (green) methods.

SUPPLEMENTARY SECTION III. ANALYTICAL GRADIENT IN VQE

The energy gradient is used in parameter optimization on classical computers. The gradient can be obtained by the finite-difference method or analytical gradients. According to our numerical experiment on either a noisy simulator or real quantum devices, the analytical gradi-

ent can significantly improve the performance of VQE in the presence of noise. In this work, the analytical gradient is obtained by the measurements of quantum circuits. The gradient of a parameter is,

$$\frac{\partial E}{\partial \theta_i} = 2\mathrm{Im}\left\{ \langle \Phi_r | \hat{U}^\dagger \hat{H} \frac{\partial \hat{U}}{\partial \theta_i} | \Phi_r \rangle \right\}. \tag{7}$$

Suppose there are M operators in the ansatz, i.e, $\hat{U} = \prod_{k=1}^{M} \hat{U}_k$, the derivative of $\hat{U}$ with respect to θ_i reads (note each entangler is $\hat{U}_k = e^{i\theta_k \hat{P}_k}$ and $\hat{U}_k^\dagger \hat{U}_k = 1$)

$$\begin{aligned}
\frac{\partial \hat{U}}{\partial \theta_i} =& \prod_{k=i+1}^{M} \hat{U}_k \left(\frac{\partial \hat{U}_i}{\partial \theta_i} \right) \prod_{k=1}^{i} \hat{U}_k \\
=& i \prod_{k=i+1}^{M} \hat{U}_k \hat{P}_i \prod_{k=1}^{i} \hat{U}_k \\
=& i \prod_{k=i+1}^{M} \hat{U}_k \hat{P}_i \left(\prod_{k=M}^{i+1} \hat{U}_k^\dagger \prod_{k=i+1}^{M} \hat{U}_k \right) \prod_{k=1}^{i} \hat{U}_k \\
=& i \left(\prod_{k=i+1}^{M} \hat{U}_k \hat{P}_i \prod_{k=M}^{i+1} \hat{U}_k^\dagger \right) \prod_{k=1}^{M} \hat{U}_k \\
\equiv& i \hat{P}_i^{\mathrm{eff}} \hat{U},
\end{aligned} \tag{8}$$

where $\hat{P}_i^{\mathrm{eff}} = \left(\prod_{k=i+1}^{M} \hat{U}_k \hat{P}_i \prod_{k=M}^{i+1} \hat{U}_k^\dagger \right)$ is the dressed operator. Consequently, the energy gradient can be obtained by

$$\frac{\partial E}{\partial \theta_i} = 2\mathrm{Im}\left\{ \langle \Phi_r | \hat{U}^\dagger \hat{H} \hat{P}_i^{\mathrm{eff}} \hat{U} | \Phi_r \rangle \right\}. \tag{9}$$

In this way, calculation of the dressed operator $\hat{P}_i^{\mathrm{eff}}$ on classical computers is required, while no ancillary qubit on the quantum circuit is needed.

[1] https://metis.readthedocs.io/en/latest, "Metis for python," (2021).
[2] O. F. Lange and H. Grubmüller, Proteins **70**, 1294 (2008).
[3] O. F. Lange and H. Grubmüller, "Generalized correlation for biomolecular dynamics," (2005).
[4] I. Rivalta, M. M. Sultan, N.-S. Lee, G. A. Manley, J. P. Loria, and V. S. Batista, PNAS **109**, E1428 (2012).
[5] M. L. Rogers and R. L. Singleton, Frontiers in Physics **8** (2020).
[6] M. E. J. Newman, PNAS **103**, 8577 (2006).
[7] H. Ushijima-Mwesigwa, C. F. A. Negre, and S. M. Mniszewski, in *Proceedings of the Second International Workshop on Post Moores Era Supercomputing (PMES'17)* (ACM New York, NY, USA, 2017) pp. 22–29.
[8] C. F. A. Negre, H. Ushijima-Mwesigwa, and S. M. Mniszewski, PLOS ONE **15**, e0227538 (2020).
[9] S. M. Mniszewski, P. A. Dub, S. Tretiak, P. M. Anisimov, Y. Zhang, and C. F. A. Negre, Sci. Rep. **11**, 4099 (2021).
[10] M. Suzuki, Phys. Lett. A **146**, 319 (1990).
[11] 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).
[12] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, Nat. Commun. **10**, 3007 (2019).