Supplementary Information for: Qubit-excitation-based adaptive variational quantum eigensolver

Yordan S. Yordanov[1,2], V. Armaos[3], Crispin H. W. Barnes[1], and David R. M. Arvidsson-Shukur[2,1]
[1] *Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[2] *Hitachi Cambridge Laboratory, J. J. Thomson Avenue, CB3 0HE, Cambridge, United Kingdom and*
[3] *Laboratory of Atmospheric Physics, Department of Physics, University of Patras, Patras, Greece*

I. SUPPLEMENTARY NOTE 1

The expression for the single parameter energy gradient in equation (23) of the main text, can be derived as follows:

$$\frac{\partial}{\partial\theta_p}E^{(m-1)}(\theta_p) = \frac{\partial}{\partial\theta_p}\langle\psi^{(m-1)}|\tilde{A}_p{}^\dagger(\theta_p)H\tilde{A}_p(\theta_p)|\psi^{(m-1)}\rangle = \frac{\partial}{\partial\theta_p}\langle\psi^{(m-1)}|e^{\theta_p\tilde{T}_p{}^\dagger}He^{\theta_p\tilde{T}_p}|\psi^{(m-1)}\rangle =$$

$$\langle\psi^{(m-1)}|e^{\theta_p\tilde{T}_p{}^\dagger}\tilde{T}_p{}^\dagger He^{\theta_p\tilde{T}_p}|\psi^{(m-1)}\rangle + \langle\psi^{(m-1)}|e^{\theta_p\tilde{T}_p{}^\dagger}H\tilde{T}_p e^{\theta_p\tilde{T}_p}|\psi^{(m-1)}\rangle = \tag{1}$$

$$\langle\psi^{(m-1)}|e^{\theta_p\tilde{T}_p{}^\dagger}[H,\tilde{T}_p]e^{\theta_p\tilde{T}_p}|\psi^{(m-1)}\rangle \tag{2}$$

In going from (1) to (2) we use that $\tilde{T}_p$ is skew-Hermitian, so that $\tilde{T}_p{}^\dagger = -\tilde{T}_p$. In the limit $\theta_p \to 0$ expression (2) becomes equivalent to the expression in equation (23) of the main text.

II. SUPPLEMENTARY NOTE 2

Here we outline the method used to calculate the trial statevectors $|\psi(\boldsymbol{\theta})\rangle$ in the classical numerical simulations used for the results in this paper. Calculating the trial statevectors, is the most time consuming part of the numerical simulations, and optimizing it is vital.

In this work, we are concerned with states of the form

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_0\rangle = \prod_{i=N_U}^{1} e^{\theta_i S_i}|\psi_0\rangle, \tag{3}$$

where N_U is the size (the number of ansatz elements) of the ansatz $U(\boldsymbol{\theta})$, $|\psi_0\rangle$ is the initial reference state, and S_i is a skew-Hermitian operator, which in this paper corresponds to either a qubit excitation operator, a fermionic excitation operator, or a string of Pauli operators. Each of these three types of skew-Hermitian operators, also satisfies the relation

$$S_i^3 = -S_i. \tag{4}$$

To calculate the $2^{N_{\text{MO}}}$-dimensional state-vector representing $|\psi(\boldsymbol{\theta})\rangle$, classically, we need to calculate the N_U exponents $\{e^{\theta_i S_i}\}$ in equation (3), and then multiply them sequentially to $|\psi_0\rangle$. Each S_i is represented by an $2^{N_{\text{MO}}}\times 2^{N_{\text{MO}}}$-dimensional matrix. Hence, the complexity of estimating each exponent $e^{\theta_i S_i}$, directly, is $O(2^{3N_{\text{MO}}})$. However, we can make use of relation (4) and write each exponent in equation (3) as

$$e^{\theta_i S_i} = \sum_{r=0}^{\infty}\frac{\theta_i^r S_i^r}{r!} = I + \sum_{r_1=0}^{\infty}\frac{(-1)^{r_1}\theta_i^{2r_1+1}}{(2r_1+1)!}S_i + \sum_{r_2=1}^{\infty}\frac{(-1)^{r_2}\theta_i^{2r_2}}{(2r_2)!}S_i^2 = I + \sin\theta S_i + (1-\cos\theta)S_i^2. \tag{5}$$

The operators $\{S_i\}$ are fixed throughout a simulation. Therefore, if we compute in advance and store the matrix representations of each S_i and S_i^2, we can evaluate the expression in equation (5) by performing matrix addition only, which has a complexity of $O(2^{2N_{\text{MO}}})$. Hence, the calculation of $|\psi(\boldsymbol{\theta})\rangle$, requires N_U matrix-to-vector multiplications and N_U matrix additions, which gives a total complexity of $O(N_U 2^{2N_{\text{MO}}})$.

The drawback of the method outlined above is that we need to store the matrices for all S_i and S_i^2 operators. For example, the most memory demanding simulation in this work, running the qubit-ADAPT-VQE for BeH_2, required

around 2GB of RAM to store the matrices for all Pauli string operators, which define the ansatz element pool of the qubit-ADAPT-VQE, and their respective squares. However, in this case, a speed-up of nearly a factor of 20 was achieved, in comparison to calculating $|\psi(\boldsymbol{\theta})\rangle$ with the general IBM's Qiskit statevector simulator.

III. SUPPLEMENTARY NOTE 3

When using a gradient-descent minimizer, e.g. the BFGS, we have the option to supply a function that returns the gradient vector of the minimized function. If we are close to the global minimum, supplying a gradient vector function guarantees a faster optimization of the variational parameters. In the case of minimizing the Hamiltonian expectation value $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle$, the i^{th} component of the energy gradient vector, $\boldsymbol{\nabla} E(\boldsymbol{\theta})$, is given by

$$\boldsymbol{\nabla}_i E(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i}\langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle = \frac{\partial}{\partial\theta_i}\langle\psi_0|U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})|\psi_0\rangle = \frac{\partial}{\partial\theta_i}\langle\psi_0|\prod_{1}^{k_1=N_U} e^{\theta_{k_1}S_{k_1}^\dagger} H \prod_{k_2=N_U}^{1} e^{\theta_{k_2}S_{k_2}}|\psi_0\rangle = \tag{6}$$

$$\langle\psi(\boldsymbol{\theta})|H\prod_{k_1=N_u}^{i+1} e^{\theta_{k_1}S_{k_1}} S_i \prod_{k_1=i}^{1} e^{\theta_{k_2}S_{k_2}}|\psi_0\rangle + \langle\psi_0|\prod_{k_1=1}^{i} e^{\theta_{k_1}S_{k_1}^\dagger} S_i^\dagger \prod_{k_1=i+1}^{N_U} e^{\theta_{k_2}S_{k_2}^\dagger}|H|\psi(\boldsymbol{\theta})\rangle = 2\langle\alpha_i(\boldsymbol{\theta})|S_i|\beta_i(\boldsymbol{\theta})\rangle, \tag{7}$$

where

$$|\beta_i(\boldsymbol{\theta})\rangle = \prod_{k=i}^{1} e^{\theta_k S_k}|\psi_0\rangle \quad \text{and} \tag{8}$$

$$|\alpha_i(\boldsymbol{\theta})\rangle = \prod_{k=i+1}^{N_u} e^{\theta_k S_k^\dagger} H|\psi(\boldsymbol{\theta})\rangle, \tag{9}$$

and N_U is the size (the number of ansatz elements) of the ansatz $U(\boldsymbol{\theta})$.

For the numerical simulations presented in this paper, the N_U components of $\boldsymbol{\nabla} E(\boldsymbol{\theta})$ can be calculated with minimum number of matrix multiplications by updating $|\beta_i(\boldsymbol{\theta})\rangle$ and $|\alpha_i(\boldsymbol{\theta})\rangle$ in the following way:

1. For $i = N_U$, initiate

$$|\alpha_{N_U}(\boldsymbol{\theta})\rangle = H|\psi(\boldsymbol{\theta})\rangle \quad \text{and} \tag{10}$$

$$|\beta_{N_U}(\boldsymbol{\theta})\rangle = |\psi(\boldsymbol{\theta})\rangle \tag{11}$$

2. For $1 < i < N_U$, update

$$|\alpha_{i-1}(\boldsymbol{\theta})\rangle = e^{\theta_i S_i^\dagger}|\alpha_i(\boldsymbol{\theta})\rangle \quad \text{and} \tag{12}$$

$$|\beta_{i-1}(\boldsymbol{\theta})\rangle = (e^{\theta_i S_i})^{-1}|\beta_i(\boldsymbol{\theta})\rangle = e^{\theta_i S_i^\dagger}|\beta_i(\boldsymbol{\theta})\rangle, \tag{13}$$

where in equation (13), we use that S_i is skew-Hermitian, so $S_i^\dagger = -S_i$.

Assuming that we have already computed and stored the matrices of the exponentials $\{e^{\theta_i S_i}\}$, when calculating $|\psi(\boldsymbol{\theta})\rangle$, to calculate each component of $\boldsymbol{\nabla} E(\boldsymbol{\theta})$ we need to perform 3 matrix-to-vector multiplications. Thus, overall to calculate $\boldsymbol{\nabla} E(\boldsymbol{\theta})$ we need to perform $3N_U$ matrix-to-vector multiplications, resulting in a total cost of $O(3N_U 2^{2N_{\text{MO}}})$ operations.

The cost of calculating $\boldsymbol{\nabla} E(\boldsymbol{\theta})$ is about 3 times the cost of calculating $|\psi(\boldsymbol{\theta})\rangle$. However, we find that using the energy gradient vector in the optimization subroutine of the VQE, reduces the number of VQE iterations by at least an order of magnitude, which justifies the use of the gradient vector.

IV. SUPPLEMENTARY NOTE 4

Here, we consider the computational complexity in terms of number of quantum computer measurements, and total run time. The computational complexity of the QEB-ADAPT-VQE is determined by steps 2 and 3.

Given that the electronic Hamiltonian, H, is represented by up to $O(N_{MO}^4)$ Pauli strings (see equation (7) of the main text), calculating each gradient in step 2 would require $O(N_{MO}^4)$ quantum computer measurements. Since $|\mathbb{P}(\tilde{A}, N_{MO})| \propto N_{MO}^4$, the complexity of step 2, in terms of quantum computer measurements is $O(N_{MO}^8)$. Step 2 is completely parallelizable so if multiple quantum computers are available, its time complexity can be arbitrarily reduced down to the time required to evaluate the expectation value of a single Pauli string term, which is proportional to the ansatz circuit depth, scaling as $O(m/N_{MO})$ (a N_{MO}-qubits circuit of m qubit excitation evolutions), where m is the iteration number of the QEB-ADAPT-VQE.

Using the BFGS minimizer, optimizing ansatz $U^{(m)}\big(\vec{\theta}^{(m)}\big)$, which has $O(m)$ variational parameters, would require $O(m^2)$ VQE energy evaluations. Therefore, each VQE run in step 3 would require $O(m^2 N_{MO}^4)$ quantum computer measurement. Hence, the overall complexity of step 3 in terms of measurements would be $O(nm^2 N_{MO}^4)$. This complexity is a worst case estimate, assuming that at each iteration, all parameters $\vec{\theta}^{(m)}$ are initialized at zero. In fact, we initiate $\vec{\theta}^{(m)}$ as $\vec{\theta}^{(m-1)} \cup 0$, so we will need fewer VQE energy evaluations to optimize the new ansatz, $U^{(m)}\big(\vec{\theta}^{(m)}\big)$. However, the complexity can also be higher if we use a direct search minimizer, like the Nelder-Mead, which is likely to be the case in practice, when noisy quantum hardware is used. Again, if multiple quantum devices are available, each of the n VQE runs can be executed in parallel. Hence, the time complexity would be lower bounded by the run-time of a single VQE run, $O(m^3/N_{MO})$ (the ansatz circuit depth is $O(m/N_{MO})$ and we need to perform $O(m^2)$ VQE energy evaluations).

Overall, the QEB-ADAPT-VQE would require $O\big(N_U(N_{MO}^8 + nN_U^2 N_{MO}^4)\big)$ quantum computer measurements, and its run-time complexity would be lower bounded by $O(N_U^4/N_{MO})$. The size of the ansatz, N_U, depends on the desired accuracy, and also is problem specific. Therefore, it is difficult to predict how it would scale with N_{MO}. For strongly correlated states, achieving chemical accuracy might require an ansatz that consist of as many as $O(N_{MO}^4)$ qubit excitation evolutions. However, for weakly correlated states, the scaling of N_U with N_{MO} is likely to be lower. Assuming the worst case scenario, the time complexity of the QEB-ADAPT-VQE will be lower-bounded by $O(N_{MO}^{15})$ and it will require $O(nN_{MO}^{16})$ quantum computer measurements. For comparison, the UCCSD-VQE has a time complexity of $O(N_{MO}^{11})$, assuming maximum parallelization, and requires $O(N_{MO}^{12})$ quantum computer measurements.

V. SUPPLEMENTARY NOTE 5

Here, we investigate the performance of the QEB-ADAPT-VQE for different values of n, the number of qubit excitation evolutions considered in step 3 of the QEB-ADAPT-VQE. As we increase n, we increase the chance to pick at each iteration the qubit excitation evolution that, added to the ansatz, achieves largest energy reduction. Following this greedy strategy is no guarantee for an optimal ansatz, since qubit evolutions do not commute in general. Nevertheless, we do expect, on average, to construct a more circuit-efficient ansatz by increasing n up to some saturation value.

To test this presumption we perform classical numerical simulations to obtain energy convergence plots for the ground states of LiH, H_6 and BeH_2 in the STO-3G basis. The simulations for the three molecules are performed for bond distances $r_{Li-H} = 3$Å, $r_{H-H} = 3$Å and $r_{Be-H} = 3$Å, away from equilibrium configurations, where correlation effects are stronger, and the effect of increasing n should be more evident. The simulation results are presented in Supplementary figure 1.

The table below summarizes the average (over number of qubit excitation evolutions) $CNOT$ count reductions, with respect to $n = 1$, for each molecule and different value of n:

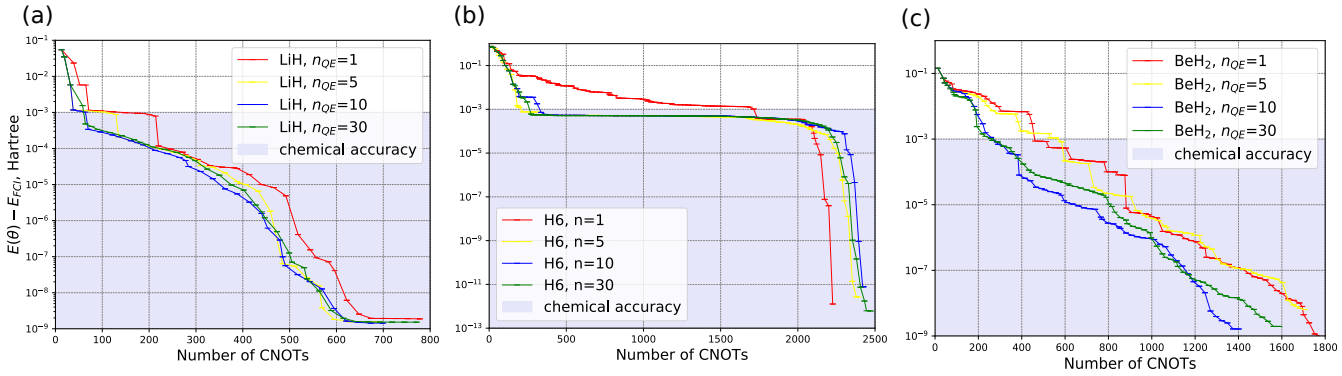
	$n = 5$	$n = 10$	$n = 30$
LiH	16%	20%	16%
BeH_2	3%	26%	22%
H_6	15%	12%	13%

TABLE I: Average (over number of qubit excitation evolutions) $CNOT$ count reduction for QEB-ADAPT-VQE($n_{qe} > 1$) as compared to QEB-ADAPT-VQE($n_{qe} = 1$).

For LiH (Supplementary figure 1.a), the QEB-ADAPT-VQE clearly constructs ansatz circuits with fewer $CNOT$s as n is increased above 1. For BeH_2 (Supplementary figure 1.c), a significant $CNOT$ count reduction is obtained for

$n = 10$ and $n = 30$, but not for $n = 5$. For H_6 (Supplementary figure 1.b), the average $CNOT$ reduction is about the same for $n = 5$, $n = 10$ and $n = 30$, but strangely the ansatz constructed by the QEB-ADAPT-VQE for $n = 1$ is the most $CNOT$-efficient for accuracies higher than 10^{-4} Hartree. Also, for all three molecules we observe no further $CNOT$ reduction for $n = 30$ as compared to $n = 10$. Actually for $n = 30$ the $CNOT$ reduction is a bit lower. As noted above these inconsistencies can be explained by the fact that the greedy strategy to obtain the lowest estimate for $E(\vec{\theta})$ at each iteration is no guarantee for constructing an optimal ansatz, because qubit excitation evolutions do not commute.

Nonetheless, there is a clear advantage in terms of $CNOT$ count, in performing step 3 of the QEB-ADAPT-VQE for $n > 1$. Despite the associated overhead in the number of quantum computer measurements with increasing n, this is justified as long as the bottleneck of NISQ computers is the quantum gate fidelity. Furthermore, we can expect the $CNOT$ count reduction for $n > 1$ to increase for larger molecules, because the QEB-ADAPT-VQE will have to consider a larger ansatz element pool.



Supplementary figure. 1: Performance of the QEB-ADAPT-VQE for different values of the protocol parameter n. The subfigures above present energy convergence plots obtained with the QEB-ADAPT-VQE with different values of n for the ground states of LiH (**a**), H_6 (**b**) and BeH_2 (**c**) in the SO-3G basis, at bond distances of $r_{Li-H} = 3$Å, $r_{H-H} = 3$Å and $r_{Be-H} = 3$Å. The plots are terminated at $\epsilon = 10^{-12}$ Hartree.