Supplementary Information for Reducing the Resources Required by ADAPT-VQE Using Coupled Exchange Operators and Improved Subroutines

Mafalda Ramôa,[1,2,3,4,5,*] Panagiotis G. Anastasiou,[1,2] Luis Paulo Santos,[3,4,5]
Nicholas J. Mayhall,[2,6] Edwin Barnes,[1,2] and Sophia E. Economou[1,2]

[1] *Department of Physics, Virginia Tech, Blacksburg, VA, 24061, USA*
[2] *Virginia Tech Center for Quantum Information Science and Engineering, Blacksburg, VA 24061, USA*
[3] *International Iberian Nanotechnology Laboratory (INL), Portugal*
[4] *High-Assurance Software Laboratory (HASLab), Portugal*
[5] *Department of Computer Science, University of Minho, Portugal*
[6] *Department of Chemistry, Virginia Tech, Blacksburg, VA, 24061, USA*

I. VARIANTS OF CEO-ADAPT-VQE

CEO-ADAPT-VQE can be defined as a variant of ADAPT-VQE which uses a pool comprised of coupled exchange operators (CEOs). We defined two types of such operators: OVP- and MVP-CEOs, with one or up to three variational parameters respectively. One key decision in the algorithm is how to choose between these subsets in each iteration. In this appendix, we compare four different decision criteria.

As explained in the main text, in each iteration of our algorithm we use the gradients of OVP-CEOs to select two candidates: $T_n^{(OVP-CEO)}$, the OVP-CEO with the highest gradient, and $T_n^{(MVP-CEO)}$, the MVP-CEO formed from all QEs with nonzero gradients acting on the same spin-orbitals as $T_n^{(OVP-CEO)}$. One of these operators must then be chosen to generate the new ansatz unitary.

We note that often, the gradient of $T_n^{(OVP-CEO)}$ is the same as the sum of the gradient magnitudes of the excitations in $T_n^{(MVP-CEO)}$. This is only not the case when the spin-orbitals are all of the same type, in which case the gradient of $T_n^{(MVP-CEO)}$ may be higher by virtue of including one more operator. Hence, we cannot in general use the gradient to decide between the two operators. One might expect that OVP-CEOs favor hardware-efficiency, since the corresponding evolutions are implemented with 9 CNOTs while MVP-CEOs are implemented with 13. However, the latter offer more variational freedom by implementing up to three independent qubit excitations each.

We use the notation of the main text to define four variants of CEO-ADAPT-VQE which serve the purpose of assessing different criteria for this decision. Step 2 is left unchanged with respect to the one defined in the main text. In what concerns step 3, the definitions of $T_n^{(MVP-CEO)}$ and $T_n^{(OVP-CEO)}$ remain the same, but the criteria of choice between the two vary across variants as indicated below.

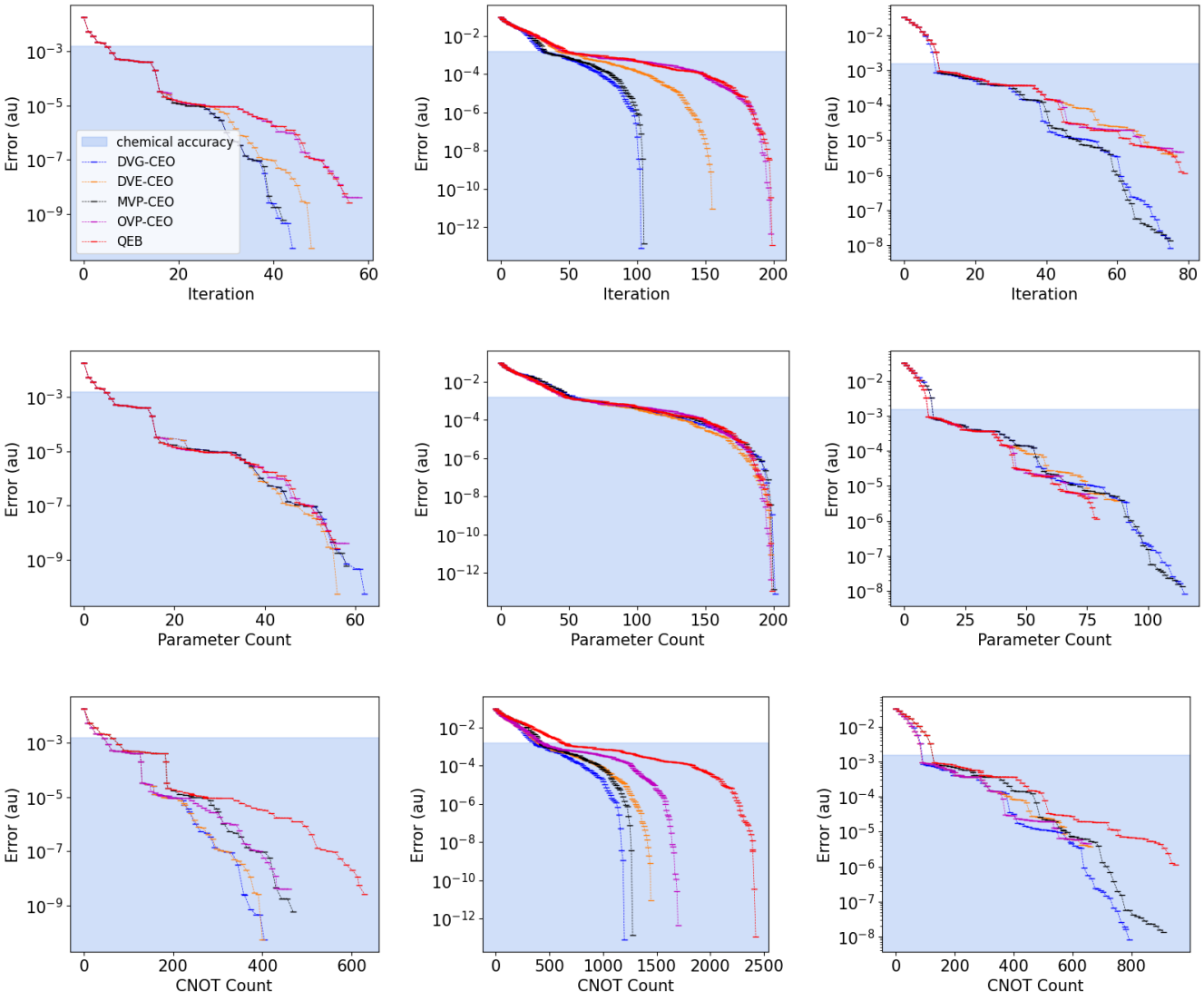
- OVP-CEO-ADAPT-VQE: Add $e^{T_n^{(OVP-CEO)}}$ to the ansatz.
- MVP-CEO-ADAPT-VQE: Add $e^{T_n^{(MVP-CEO)}}$ to the ansatz.
- Decision via gradient (DVG)-CEO-ADAPT-VQE: Add $e^{T_n^{(OVP-CEO)}}$ to the ansatz if $\#M_{\neq 0}^{(QE)}(T_n^{(OVP-CEO)}) = 1$. Otherwise, add $e^{T_n^{(MVP-CEO)}}$.
- Decision via energy (DVE)-CEO-ADAPT-VQE: Add $e^{T_n^{(OVP-CEO)}}$ to the ansatz if $\#M_{\neq 0}^{(QE)}(T_n^{(OVP-CEO)}) = 1$. Otherwise, obtain ΔE_{OVP} and ΔE_{MVP}, the energy changes produced by adding to the ansatz $e^{T_n^{(OVP-CEO)}}$ and $e^{T_n^{(MVP-CEO)}}$ (respectively) and performing a full optimization. Add the former unitary if $\frac{\Delta E_{MVP}}{13} > \frac{\Delta E_{OVP}}{9}$, and the latter otherwise [1].

In essence, we have two algorithms which use OVP-CEOs or MVP-CEOs exclusively, and two algorithms which combine the two operator types. DVG-CEO-ADAPT-VQE is the algorithm defined in the main text, where the decision is gradient-based: $T_n^{(MVP-CEO)}$ is chosen unless there is only one operator with nonzero gradient. In DVE-CEO-ADAPT-VQE, the decision is energy-based: We perform two independent optimizations of the ansatz with each of the two candidate unitaries, and effectively select the one which leads to the highest absolute value of the energy change per unit CNOT. The aim is to maximize the impact on the energy with respect to the added CNOT count in each iteration. Naturally, we could consider other hardware-related criteria, such as the CNOT depth. Note that this decision criterion roughly doubles the number of optimization required per ADAPT-VQE iteration.

Figure 1 compares these four algorithms. QEB-ADAPT-VQE is also plotted for reference. We verify that all variants of CEO-ADAPT-VQE improve upon QEB-ADAPT-VQE in terms of the number of CNOTs required for a given error, while being roughly matched in terms of the parameter count.

The decision via gradient performs the best among all variants of CEO-ADAPT-VQE, outperforming all oth-

* mafalda@vt.edu



(a) LiH at 1.5Å (b) H_6 at 1Å (c) BeH_2 at 1.3Å

FIG. 1. Comparison of CEO-ADAPT-VQE Variants at Equilibrium Bond Distances

The subfigures depict the convergence of different variants of the CEO-ADAPT-VQE algorithm for **a** LiH, **b** H_6, **c** BeH_2, at bond distances close to equilibrium. The QEB-ADAPT-VQE algorithm is also included for reference. The error is plotted against the iteration number (top), parameter count (middle) and CNOT count (bottom). The convergence criterion is a gradient threshold of 10^{-6} and 10^{-5} on the 12 qubits and 14 qubit molecules, respectively.

ers for all systems. This is remarkable, given the fact that the decision via energy is locally optimal (with respect to the CNOT count). In fact, in the first iteration where DVG- and DVE-CEO-ADAPT-VQE diverge, the latter is sure to have a lower error, because the optimization of the former is essentially a restricted version of the optimization of the latter. Yet, as the iterations proceed, the DVE variant lags behind. This can only be explained as a nonlocal effect. We hypothesize that the fact that DVG-CEO-ADAPT-VQE always includes all independently parameterized QEs with nonzero gradients leads to a higher variational flexibility which becomes more advantageous as the iterations proceed. Certain operators may have a low impact on the energy upon being added, but be beneficial later on — perhaps they introduce important Slater determinants into the superposition state, or they interact favorably with operators added after them.

In general, the dynamic strategies (DVG and DVE), which decide between CEO types on a case-by-case basis, seem to outperform the predetermined ones. This is expected, since the latter make no effort to optimize the choice of operator type. However, there is a notable exception: MVP- beats DVE-CEO-ADAPT-VQE for H_6. Once again, this suggests that privileging extra variational freedom can be rewarding in the long term,

and particularly so for highly correlated system—which strengthens the conjecture in the paragraph above.

Figure 2 shows similar plots for stretched bond distances. We observe the same trends: QEB-ADAPT-VQE is outperformed by all variants of CEO-ADAPT-VQE, with the leading one being DVG. The parameter count is roughly equivalent for all five algorithms.

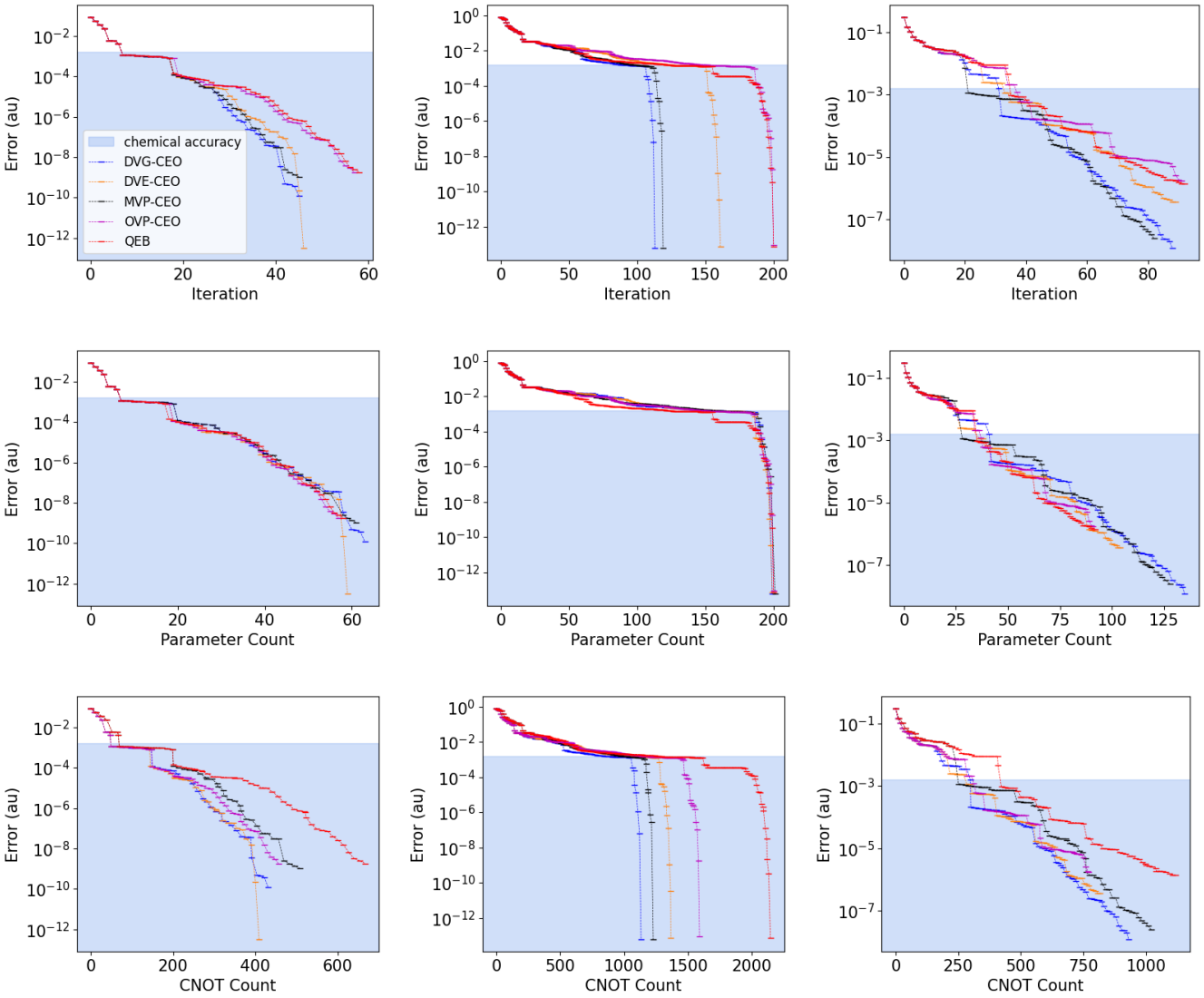
It has become evident that the decision via gradient is the optimal choice in terms of the CNOT count and number of iterations, despite requiring no extra optimizations as compared to the canonical ADAPT-VQE (unlike the decision via energy). Therefore, we take DVG-CEO-ADAPT-VQE as the standard ADAPT-VQE algorithm with CEOs and refer to it simply as CEO-ADAPT-VQE in the main text.

II. COMBINING CEO-ADAPT-VQE WITH TRANSVERSAL PROPOSALS

In this appendix we investigate the individual impact of each of the strategies consolidated to create the CEO-ADAPT-VQE* algorithm: TETRIS [2], optimized gradient measurements (OGM) [3] and Hessian recycling (HR) [4]. We refer to the main text for details about the methods.

Figure 3 compares the convergence of CEO-ADAPT-VQE with no improvements, with all improvements, and with each improvement individually. The uppermost panels show that, as would be expected, only TETRIS relevantly affects the iteration count. The remaining strategies affect the total number of measurements per iteration, but leave the iteration count unchanged. The CNOT count (second row of panels) is similar for all curves, while the CNOT depth (third row) is predictably lowered by TETRIS while being roughly unchanged by the the other proposals. Finally, the last row of panels shows that OGM and HR contribute to decreasing the measurement costs of the algorithm. Interestingly, the relative impact of the two is system-dependent. While OGM is more impactful than HR for LiH and, in earlier iterations, for BeH_2, the reverse happens for H_6 and BeH_2 in later iterations. We attribute this to the complexity of the optimizations. As an example, in the case of H_6, not only are the optimizations higher dimensional on average, but they also tend to require more cost function evaluations than optimizations for other molecules (even for matched parameter counts). As such, in this case, the cost of the energy measurements throughout the optimizations prevails over the cost of the gradient measurements, such that a strategy which tackles the cost of the optimization (HR) is more beneficial than one which tackles the cost of the gradient measurement round (OGM). We note that while its focus is decreasing circuit depth, TETRIS also offers a slight reduction in measurement costs by virtue of requiring a lower number of iterations (and thus fewer gradient measurement rounds and fewer optimizations in total).

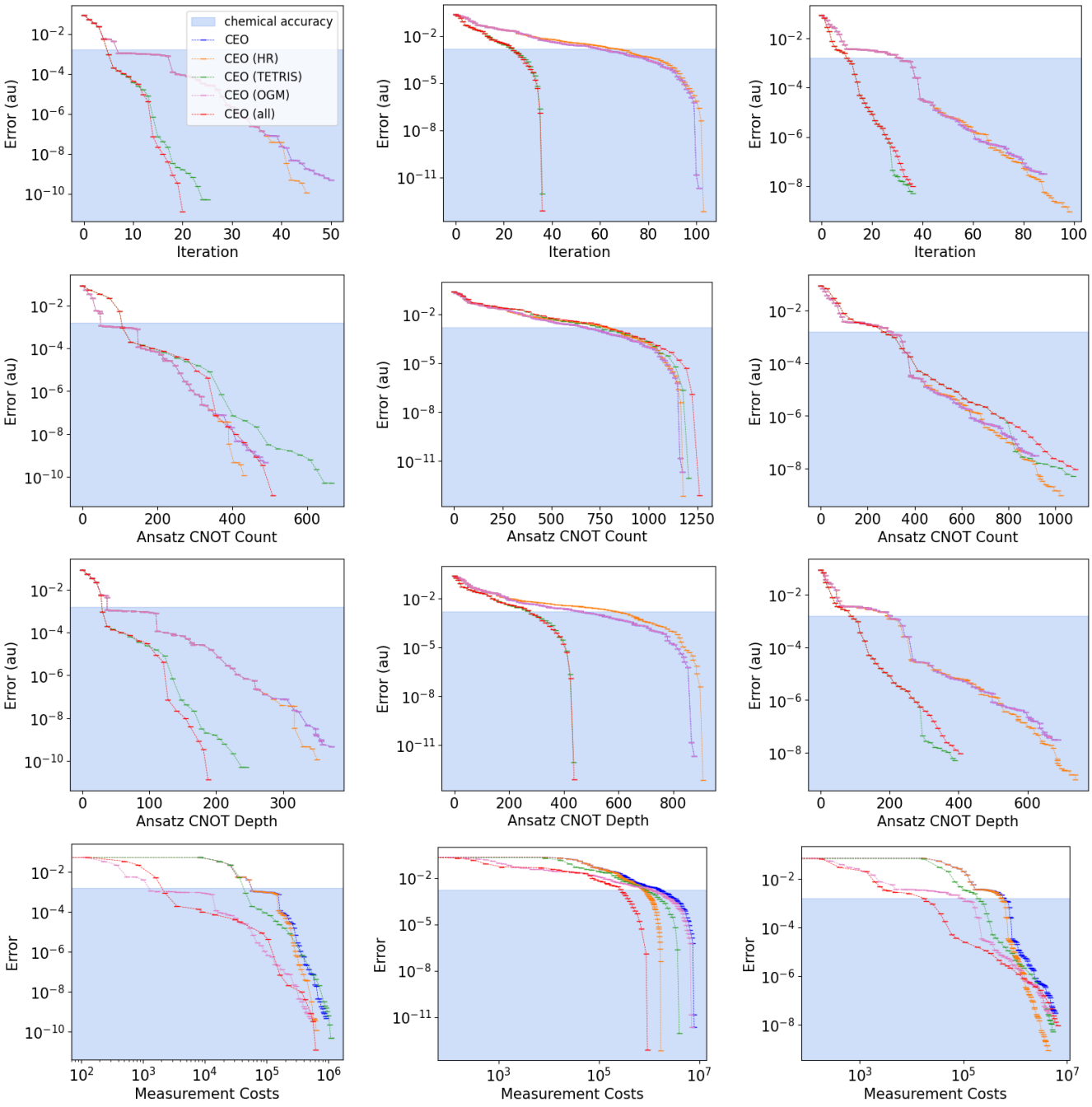
Finally, we remark that the benefits of HR are expected to take more iterations to be harvested when we employ strategies which increase the number of new variational parameters per iteration, such as MVP-CEOs and TETRIS, due to the higher number of cold-started entries in the inverse Hessian. In the case of the 12 qubit molecules, TETRIS-CEO-ADAPT-VQE can add up up to three MVP-CEOs per iteration. Up to two of them may act on spatial orbitals of the same type, and thus have 3 variational parameters. In total, this may lead to up to 8 variational parameters, and the count will be higher for larger molecules. In early iterations, a large number of new parameters represents a significant number of second derivatives about which the recycled Hessian contains no information. Therefore, we can expect HR to become more impactful for larger and more strongly correlated systems, where the number of old parameters far outweighs the number of freshly added parameters in later iterations.



(a) LiH at 3Å (b) H_6 at 3Å (c) BeH_2 at 3Å

FIG. 2. Comparison of CEO-ADAPT-VQE Variants at Stretched Bond Distances

Th subfigures depict the convergence of different variants of the CEO-ADAPT-VQE algorithm for **a** LiH, **b** H_6, **c** BeH_2, at stretched bond distances. The QEB-ADAPT-VQE algorithm is also included for reference. The error is plotted against the iteration number (top), parameter count (middle) and CNOT count (bottom). The convergence criterion is the same as in Fig. 1.



(a) LiH at 3Å (b) H_6 at 1.5Å (c) BeH_2 at 2Å

FIG. 3. Combining CEO-ADAPT-VQE with Transversal Improvements

The subfigures depict the convergence of the CEO-ADAPT-VQE algorithm for **a** LiH, **b** H_6, **c** BeH_2, at various bond distances. The algorithm is implemented *in tandem* with other recent proposals: Hessian recycling (HR) [4], TETRIS [2], optimized gradient measurements (OGM)[3], and all three (all). The baseline case, which uses the CEO pool while following the original ADAPT-VQE protocol [5] with a vanilla measurement strategy, is also included for reference. The error is plotted against the iteration number, CNOT count, CNOT depth, and measurement costs. The region shaded blue is the region of chemical accuracy (error below 1kcal/mol). The convergence criterion is a gradient threshold of 10^{-6} and 10^{-5} on the 12 qubits and 14 qubit molecules, respectively. The curves for CEO and CEO (OGM) overlap on all plots except those pertaining to measurement costs.

III. ORBITAL OPTIMIZATION

Reference [6] proposed ADAPT-VQE-SCF, an algorithm that merges orbital optimization techniques with ADAPT-VQE.

Given a variational state $|\psi(\boldsymbol{\theta})\rangle$, we can obtain an orbital-optimized version $\left|\tilde{\psi}(\boldsymbol{\kappa},\boldsymbol{\theta})\right\rangle$ as

$$\left|\tilde{\psi}(\boldsymbol{\kappa},\boldsymbol{\theta})\right\rangle = e^{\boldsymbol{\kappa}} |\psi(\boldsymbol{\theta})\rangle . \tag{1}$$

Here, $\boldsymbol{\kappa}$ is an anti-hermitian operator defined by

$$\boldsymbol{\kappa} = \sum_{p>q} \kappa_{pq}(\hat{E}_{pq} - \hat{E}_{qp}), \tag{2}$$

where the $\hat{E}_{pq}$ are spin-adapted one-body operators,

$$\hat{E}_{pq} = \hat{a}^{\dagger}_{p\alpha}\hat{a}_{q\alpha} + \hat{a}^{\dagger}_{p\beta}\hat{a}_{q\beta}, \tag{3}$$

and the κ_{pq} are variational parameters that can be optimized to decrease the energy via orbital rotations. These rotations can be implemented efficiently in a classical computer by producing a new Hamiltonian with updated molecular integrals, and therefore do not imply additional circuit costs.

ADAPT-VQE-SCF optimizes the molecular orbital basis along with the ansatz parameters (i.e., it carries out a simultaneous optimization of $\boldsymbol{\kappa}$ and $\boldsymbol{\theta}$) in each iteration). The gradients of the orbital rotation coefficients κ_{pq} can be obtained from the two-particle reduced density matrices, which are measured when evaluating the energy itself; therefore, these gradients are provided to us 'for free' during the optimization. However, $\boldsymbol{\kappa}$ comes with an additional $\frac{N_s(N_s-1)}{2}$ variational parameters (with N_s the number of spatial orbitals). This results in a higher dimensional optimization, which might increase the total number of energy and ansatz gradient measurements required for the optimizer to converge and, therefore, indirectly lead to additional measurement costs.

Figure 4 shows the result of merging orbital optimization (OO) with the CEO-ADAPT-VQE* algorithm. We can see that the impact on the relevant markers is negligible: the CNOT count and depth required to reach a given error are only very slightly decreased, while the measurement costs are slightly increased. We note that ADAPT-VQE-SCF was developed to enable simulations with large atomic orbital basis sets; it is not surprising that it has a subpar performance in the case of minimal basis sets.

While we recognize that orbital optimization may be an important enhancement to ADAPT-VQE in the case of larger molecules and/or basis sets, we exclude it from the algorithms in the main text due to its minor impact on the energy under the circumstances we consider.

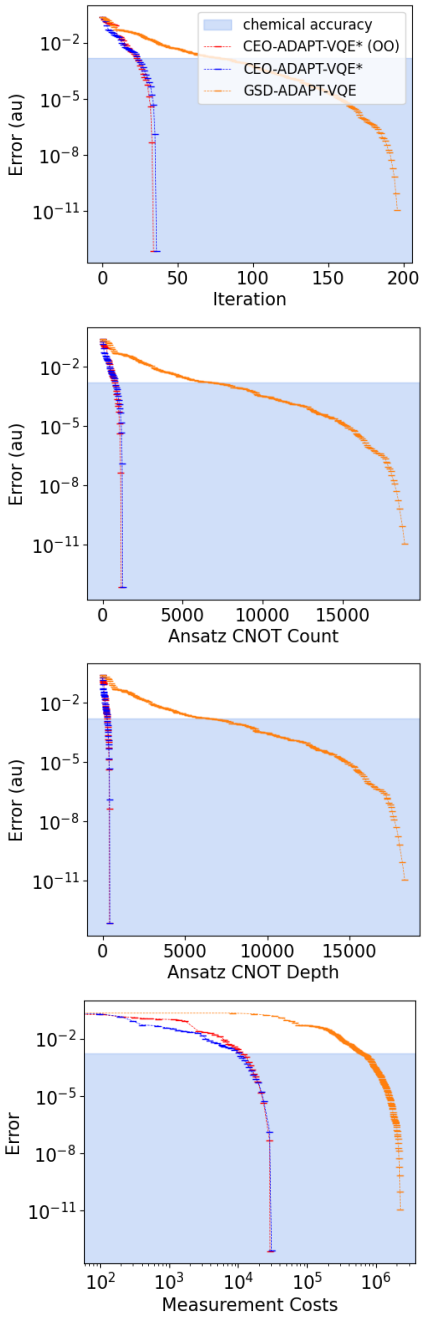


FIG. 4. Combining CEO-ADAPT-VQE with Orbital Optimization

The subfigures depict the convergence of CEO-ADAPT-VQE*, as defined in the main text, with and without orbital optimization. GSD-ADAPT-VQE is also included for reference. We consider H_6 at an interatomic distance of 1.5Å. CEO-ADAPT-VQE*, GSD-ADAPT-VQE and UCCSD-VQE are defined in the main text. The error is plotted against the iteration number, CNOT count, CNOT depth, and measurement costs. The measurement costs are given as multipliers for the cost of one energy measurement. The region shaded blue is the region of chemical accuracy (error below 1kcal/mol). The convergence criterion is a gradient threshold of 10^{-6} and 10^{-5} on the 12-qubit and 14-qubit molecules, respectively.

IV. HAMILTONIAN GROUPING

In part of the results included in the main text we applied the k-commutativity grouping [7] to the molecular Hamiltonians, and took the $\hat{R}$ metric as an approximation to the ratio M_u/M_g between measurement costs without (M_u) and with (M_g) grouping [8]. As discussed, we chose $k = n$ (general commutativity) in light of the expectation that the depth of the measurement circuit will not be significant relative to the depth of the state preparation circuit. In this case, it is convenient to consider full commutativity to maximize the reduction in measurement costs at the expense of negligible additional circuit depth. However, in general, choosing the optimal k involves a compromise between circuit depth and measurement costs. As such, it is interesting to analyze the evolution of the savings in measurement costs against k.

Figure 5 showcases this evolution for the systems we consider in this paper. We observe that for LiH_2 and BeH_2, the case $k = 1$ (qubit-wise commutativity) offers the greatest improvement in measurement costs as compared to other unit increments of k. However, in the case of H_6, the molecule where the grouping achieves the greatest savings, the improvement resulting from incrementing k is roughly consistent throughout the whole plot.

Unlike Ref. [7], we do not observe $\hat{R}$ to stall after a given value k^* - in fact, $\hat{R}$ steadily increases until reaching the maximum value for the highest value of k. We conjecture that this difference may stem from the following facts: (i) the systems we consider are distinct from those where such behavior was encountered (Fermi-Hubbard vs molecular Hamiltonians), and (ii) the number of qubits of our systems is significantly lower. The stabilization of $\hat{R}$ at a maximum value for $k < n$ could happen for larger n. In that case, the measurement cost reduction of general commutativity can be achieved with shallower measurement circuits.

[1] Note that this energy change is expected to be negative.

[2] P. G. Anastasiou, Y. Chen, N. J. Mayhall, E. Barnes, and S. E. Economou, Tetris-adapt-vqe: An adaptive algorithm that yields shallower, denser circuit ansätze (2022), arXiv:2209.10562 [quant-ph].

[3] P. G. Anastasiou, N. J. Mayhall, E. Barnes, and S. E. Economou, How to really measure operator gradients in adapt-vqe (2023), arXiv:2306.03227 [quant-ph].

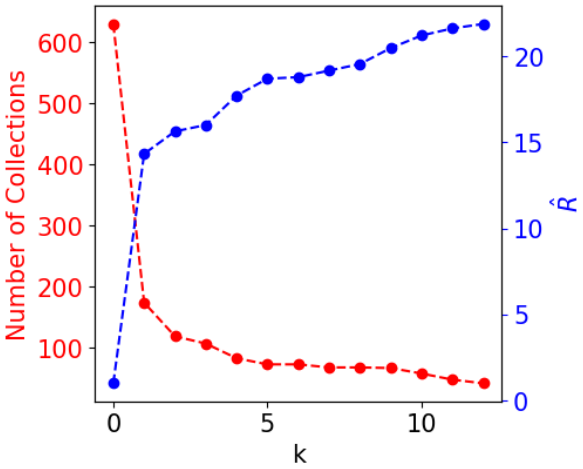
[4] M. Ramôa, L. P. Santos, N. J. Mayhall, E. Barnes, and S. E. Economou, Reducing measurement costs by recycling the hessian in adaptive variational quantum algorithms (2024), arXiv:2401.05172 [quant-ph].

[5] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nature Communications **10**, 10.1038/s41467-019-10988-2 (2019).

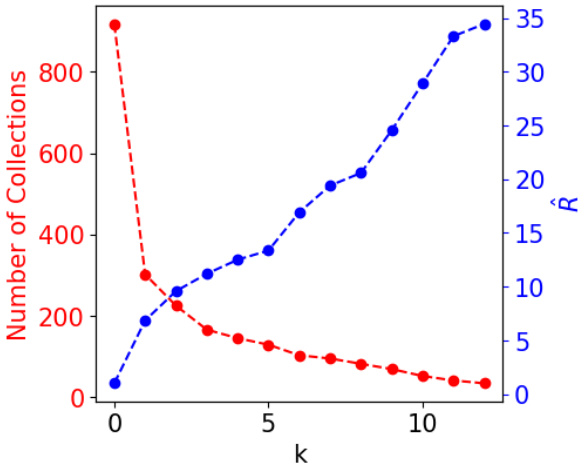
[6] A. Fitzpatrick, A. Nykänen, N. W. Talarico, A. Lunghi, S. Maniscalco, G. García-Pérez, and S. Knecht, Self-consistent field approach for the variational quantum eigensolver: Orbital optimization goes adaptive, The Journal of Physical Chemistry A **128**, 2843–2856 (2024).

[7] B. DalFavero, R. Sarkar, D. Camps, N. Sawaya, and R. LaRose, k-commutativity and measurement reduction for expectation values (2024), arXiv:2312.11840 [quant-ph].

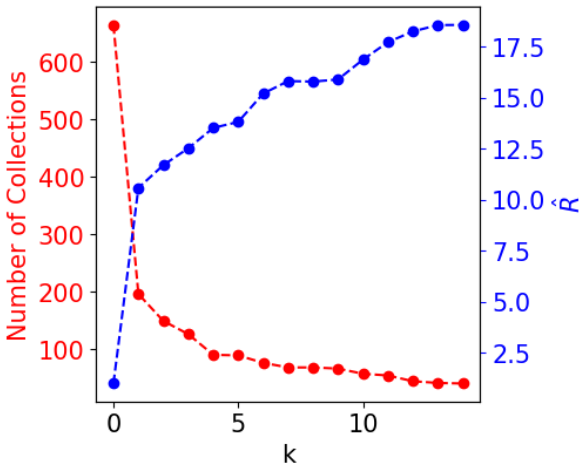
[8] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measurement of pauli operators in the presence of finite sampling error, Quantum **5**, 385 (2021).



(a) LiH at 3Å



(b) H_6 at 1.5Å



(c) BeH_2 at 2Å

FIG. 5. Impact of k on the k-Commutativity Grouping

Evolution of the number of Pauli string collections and $\hat{R}$ against k, which defines the granularity of the commutativity considered in the observable grouping [7], for **a** LiH, **b** H_6, **c** BeH_2. While it is expected that fewer collections will be associated with lower measurement costs, covariances and varying coefficient magnitudes result in a nonlinear relation between the two. The metric $\hat{R}$ [8] approximates the ratio of measurement costs with and without grouping. The point $k = 0$ represents no grouping, such that $\hat{R}$ is one and the number of collections is the total number of Pauli strings in the Hamiltonian.