

# Adaptive pruning-based optimization of parameterized quantum circuits

Sukin Sim,<sup>1,2,\*</sup> Jonathan Romero,<sup>2</sup> Jérôme F. Gonthier,<sup>2</sup> and Alexander A. Kunitsa<sup>2</sup>

<sup>1</sup>*Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA*

<sup>2</sup>*Zapata Computing, Inc., 100 Federal Street, Boston, MA 02110, USA*

Variational hybrid quantum-classical algorithms are powerful tools to maximize the use of Noisy Intermediate Scale Quantum devices. While past studies have developed powerful and expressive ansatze, their near-term applications have been limited by the difficulty of optimizing in the vast parameter space. In this work, we propose a heuristic optimization strategy for such ansatze used in variational quantum algorithms, which we call “Parameter-Efficient Circuit Training” (PECT). Instead of optimizing all of the ansatz parameters at once, PECT launches a sequence of variational algorithms, in which each iteration of the algorithm activates and optimizes a subset of the total parameter set. To update the parameter subset between iterations, we adapt the *dynamic sparse reparameterization* scheme by Mostafa *et al.* (arXiv:1902.05967). We demonstrate PECT for the Variational Quantum Eigensolver, in which we benchmark unitary coupled-cluster ansatze including UCCSD and  $k$ -UpCCGSD, as well as the low-depth circuit ansatz (LDCA), to estimate ground state energies of molecular systems. We additionally use a layerwise variant of PECT to optimize a hardware-efficient circuit for the Sycamore processor to estimate the ground state energy densities of the one-dimensional Fermi-Hubbard model. From our numerical data, we find that PECT can enable optimizations of certain ansatze that were previously difficult to converge and more generally can improve the performance of variational algorithms by reducing the optimization runtime and/or the depth of circuits that encode the solution candidate(s).

## I. INTRODUCTION

Variational quantum algorithms (VQAs) were developed to maximize the capabilities of Noisy Intermediate-Scale Quantum (NISQ) computers [1]. Two early examples of VQAs are the variational quantum eigensolver [2] and the quantum approximate optimization algorithm [3]. In addition to applications in quantum chemistry and combinatorial optimization, variational algorithms have more recently been applied to address a range of tasks in machine learning, including data classification [4–6], data compression [7], and generative modeling [8–13].

Within the VQA framework, a trial wavefunction is prepared on a quantum device by executing a parameterized quantum circuit. This is followed by repeatedly measuring the state to estimate the expectation value of some Hermitian operator with respect to the current trial wavefunction. These expectation values are then used to evaluate an objective function which a classical optimizer maximizes or minimizes by varying the parameter values of the quantum circuit.

While VQAs are promising candidates for demonstrating advantage of early quantum computers, one of the main challenges of realizing VQAs is effectively optimizing the parameters of the tunable ansatz. In recent years, there have been significant efforts in various directions to better understand and improve this circuit optimization step. Such directions include analyzing the trainability of parameterized quantum circuits (PQCs) [14–16], improving existing or proposing new optimizers [17–25], employing

high-level optimization strategies on top of optimizers (e.g. adiabatically-assisted VQAs, layerwise training) [26–28], devising more intelligent parameter initialization strategies [18, 29, 30], and running comprehensive optimizer benchmarks [31]. Despite the rapid progress, optimization of PQCs remains a significant challenge especially in the NISQ era: we ideally want a *low-depth* ansatz that is able to *accurately* describe some quantum system-of-interest and is also relatively *easy to optimize*.

To approach this problem, in this work we propose an optimization strategy that leverages past developments of powerful ansatze, which we call “Parameter-Efficient Circuit Training” (PECT). That is, numerous past studies have developed and analyzed ansatze that have the potential to express or well-describe quantum systems-of-interest, e.g. ground states of (strongly correlated) fermionic systems [32–38]. Examples include unitary coupled-cluster with singles and doubles excitations (UCCSD),  $k$ -layered generalized variant of UCC wavefunction ( $k$ -UpCCGSD), and the low-depth circuit ansatz (LDCA) [36, 37, 39]. However, applications of these ansatze are often inhibited by the large numbers of variational parameters to tune. We leverage the fact that these parameterized ansatze often contain superfluous parameters, i.e. parameters that are redundant and thus are not strictly necessary for preparing a state that achieves the objective(s). Our optimization strategy exploits this characteristic of PQCs, without prior knowledge of the target state, by limiting the optimization task to a subset of the total pool of ansatz parameters and refining the parameter subset by updating both the corresponding gate composition and parameter values.

Using this strategy, we show that we can efficiently

\* ssim@g.harvard.edu

prepare ground states of various molecular systems using ansatze that have high parameter counts and are redundant in parameterization. For instance, using 30 out of 60 parameters in the 2-UpCCGSD ansatz to estimate ground states of lithium hydride, we observed a 28% reduction in circuit depth and 42% reduction in two-qubit gate count on average when using PECT to optimize the parameterized ansatz. We additionally estimate optimization runtimes using circuit depths and numbers of function evaluations. We show that using PECT with ansatze such as  $k$ -UpCCGSD and LDCA, one can significantly reduce the optimization runtime estimates. For instance, averaged over bond lengths considered in VQE calculations, using PECT to optimize LDCA circuits modeling ground states of lithium hydride led to a 73% reduction in estimated optimization runtime, which in this case was achieved by reducing both the circuit depths and numbers of function calls.

The rest of the paper is organized as follows. We first introduce our optimization strategy in detail in Section II. Our proposed strategy, PECT, is adapted from the “dynamic sparse reparameterization” (DSR) method that has been applied to train deep convolutional neural networks [40]. We then describe the PECT method applied in the context of the Variational Quantum Eigensolver (VQE) as an example of a VQA. In Section III, we numerically demonstrate the utility of PECT by optimizing VQE circuits for lithium hydride and the symmetric stretching of O-H bonds in a water molecule using UCCSD,  $k$ -UpCCGSD, and LDCA. While PECT can provide advantages in energy accuracy, circuit depths, and optimization runtimes, we show that certain ansatze are better suited for PECT than others. In Section IV, we further analyze our simulations and discuss our observations on the robustness of parameter initialization for LDCA, prunability of the three ansatze, reductions in circuit resources due to PECT, and estimation of optimization runtimes. We additionally propose a method to combine PECT with a layerwise circuit optimization strategy. Using the layerwise PECT approach, we show that we can efficiently optimize an ansatz that has recently been proposed to estimate the ground state energy densities of the one-dimensional Fermi-Hubbard model at various chain lengths [41]. Lastly, we comment on the observed parameter dynamics in the optimization of multi-layered PQCs, in which parameters of earlier layers appear to change less, on average, throughout the optimization compared to parameters of later layers. We conclude with a summary and outlook of future directions for this work.

## II. PARAMETER-EFFICIENT CIRCUIT TRAINING (PECT)

In the traditional implementation of a variational quantum algorithm, there is typically a single optimization, in which the classical optimization routine is tasked

with tuning every angle of the PQC to find parameter values that correspond to a minimum (or maximum) of an objective function. While this strategy is viable for a limited number of parameters, as we scale the algorithm to larger systems, larger circuits with more variational parameters may be needed. As a direct result, the classical optimizers expend a substantially larger number of function evaluations from the quantum computer and become over-burdened when attempting to search over the vast parameter space.

The strategy employed by PECT turns this large optimization problem into a series of optimizations in which, at each iteration, an optimization is performed only on a subset of the whole parameter set. We show the overall flow diagram of PECT in Figure 1. In the following description of PECT, we adopt notations used in Ref. [40] to provide an analogous description of their training method in the context and language of parameterized quantum circuits.

To initialize PECT, we first define the variational quantum algorithm and the hyperparameters for PECT. The variational algorithm can be defined by choosing an objective function and an  $L$ -layered parameterized quantum circuit. Many variational quantum algorithms make use of multi-layered parameterized quantum circuits, in which layers of a circuit template are repeated with the goal of increasing the expressiveness of circuits [42]. We consider a case in which a PQC is described by the set of its circuit parameters, grouped by their corresponding layers:  $\{\theta_l\}$ , where  $l \in [L]$  index the layers. We let  $N$  be the total number of parameters in the PQC, with  $N_l$  denoting the number of parameters in each circuit layer  $l$ . Before implementing PECT, there are two assumptions for the PQC-at-hand:

1. The ansatz  $|\psi(\theta)\rangle$  is able to express the solution space/state and the ansatz is “prunable,” i.e. contains parameterized gate operations that are not necessary to generate the solution space. This may be due to redundant parameterization. We claim such redundancy is a valid assumption for many PQC designs with some commonly used patterns of gates, e.g. single-qubit rotations on every qubit. In this case, it was recently shown that not all single-qubit rotations are necessary for a PQC to be highly expressive [43].
2. Each gate operation of the PQC is constructed such that setting its corresponding parameter value to 0 corresponds to implementing the trivial (identity) gate operation. This is a valid assumption for a wide class of parameterized gates of the form  $G(\theta) = e^{-ia\theta P}$  for some Hermitian operator  $P$  and constant  $a$ .

While there are several hyperparameters in PECT, we introduce one particular hyperparameter, *global sparsity*, here and reserve describing the others in the following subsection, where they are more relevant. The global

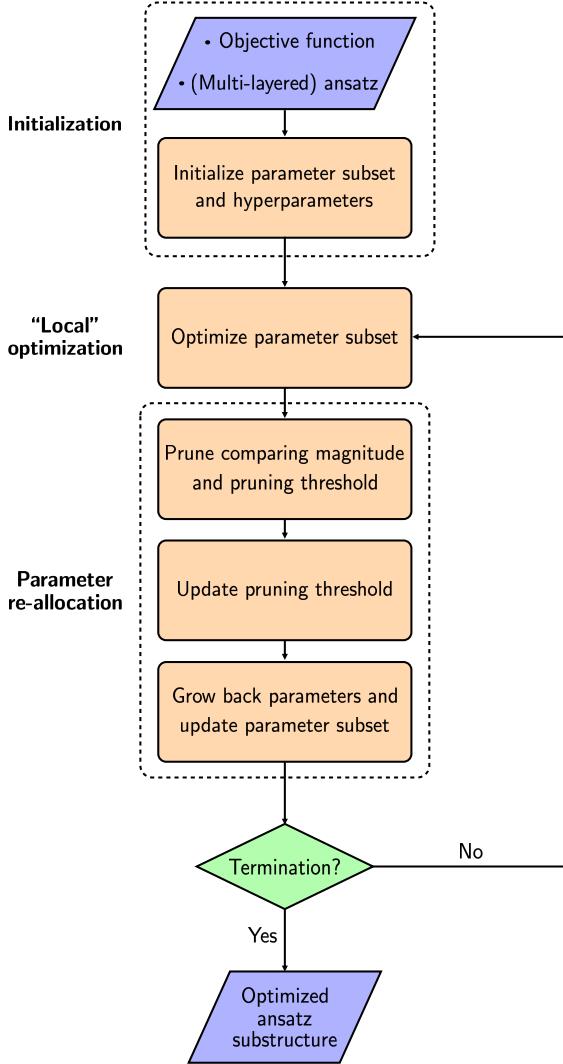


Figure 1. Flow diagram outlining main steps of the “parameter-efficient circuit training” (PECT) method. In addition to inputs of a variational quantum algorithm, e.g. objective function and some (multi-layered) ansatz, several hyperparameters for PECT are provided by the user, e.g. global sparsity  $s$  of the parameterized ansatz and the target number of parameters to prune at each iteration  $N_p$ . After initializing the parameter subset, corresponding to an ansatz substructure, parameters within the subset are tuned using a classical optimizer, e.g. L-BFGS-B. We call this step a “local” optimization. Once the parameter values of the subset have been tuned, the parameter subset is further updated in the parameter re-allocation step. The re-allocation step consists of magnitude-based pruning of parameters/gates based on a threshold followed by updates of the pruning threshold for the next iteration and of the parameter subset by growing or adding the same number of parameters pruned. The local optimization and parameter re-allocation steps are repeated until some termination criteria are satisfied.

sparsity, denoted by  $s$ , is defined as  $s = 1 - \frac{M}{N}$ , where  $M$  is the size of the parameter subset to be optimized. Thus, for example, a low value for  $s$  where  $s \in (0, 1)$  implies a larger fraction of the total parameter set PECT optimizes in each iteration. After choosing the global sparsity, we define the parameter subset for the first iteration. Gates with parameters that are not included in the subset are replaced by the identity operation, for example by setting angles to 0 in the case of Pauli rotations, resulting in shorter circuits both in depth and number of operations

after compilation.

To define the parameter subsets, we utilize the sparse parameter representation in Ref. [40], in which a sparse parameter vector for layer  $l$ , or  $\boldsymbol{\theta}_l$ , can be described using two vectors  $\psi_l$  and  $\phi_l$ . Elements of  $\psi_l$  are the indices of  $\boldsymbol{\theta}_l$  that correspond to parameters with nonzero values, and elements of  $\phi_l$  correspond to those nonzero values. We let  $M_l$  be the size of  $\psi_l$  and  $\phi_l$ , or the number of nonzero parameters in  $\boldsymbol{\theta}_l$ . The total number of nonzero parameters in the full parameter vector  $\boldsymbol{\theta}$  is thus

$M = \sum_l M_l$ . This allows us to define the local sparsity  $s_l = 1 - \frac{M_l}{N_l}$ , the sparsity of parameters of each layer, in addition to the global sparsity  $s$ . While  $s$  is kept constant throughout iterations in PECT, compositions of  $\phi_l$ ,  $\psi_l$ , and  $M_l$  (and thus  $s_l$ ) may change between iterations.

To initialize the parameter subset, we uniformly sample  $M_l^{(0)} = (1 - s)N_l$  parameter positions in each circuit layer, which defines  $\psi_l$  at iteration  $t = 0$  or  $\psi_l^{(0)}$ . When generating  $M_l^{(0)}$  for each layer, care must be taken in rounding  $M_l$ 's for all  $l$  such that the global sparsity is kept constant. The corresponding initial parameter values,  $\phi_l^{(0)}$  are either uniformly sampled, e.g. from the range  $[0, 4\pi]$  for each rotation angle, or use informed guesses such as MP2 amplitudes for the unitary coupled-cluster ansatz [44]. Once initialized, the parameter subset provides the description for an ansatz substructure, and the corresponding parameter values  $\phi$  ( $\phi_l$  for all  $l$ ) are tuned to maximize or minimize the objective function using some optimizer, e.g. L-BFGS-B [45] or SLSQP [46]. We refer to such optimization of an ansatz substructure as a “local optimization” (as shown in Figure 1) and the classical optimizer used as a “local optimizer.” One of the design choices that is left to the user is the choice of local optimizer and the corresponding hyperparameters. Choice of the local optimizer may depend on factors such as the maximum number of function evaluations or measurements the user would like to allocate to each local optimization step or the availability of high-precision gradients.

Once the optimization is complete, both  $\psi_l$ , which was defined before the local optimization step, and  $\phi_l$ , which was just optimized, are fed as inputs to the *parameter reallocation* algorithm to update the parameter subset. While we describe the parameter reallocation algorithm in greater detail in the following subsection, the goal of reallocating parameters is to update the parameter subset by removing and adding parameters the algorithm deems unimportant and important, respectively. The sequence of local optimization of a parameter subset followed by the parameter reallocation subroutine is repeated until some termination criteria are met. Our implementation utilizes a combination of termination criteria including convergence in the value of the objective function, limit on the overall maximum number of function calls (e.g.  $5 \times 10^5$  for optimizing LDCA circuits for estimating ground states of LiH), and detection of oscillating behavior in the final objective function values of the four latest PECT iterations. By the end of the PECT procedure, we obtain the optimized ansatz substructure that is a solution candidate for the variational quantum algorithm.

Overall, one can think of PECT as a procedure that hops over and optimizes different sparse parameterizations of an ansatz, which correspond to different ansatz substructures. In the following subsection, we describe the parameter reallocation algorithm in greater detail.

## A. Parameter reallocation for parameterized quantum circuits

The parameter reallocation algorithm was proposed by Ref. [40] to sparsely re-parameterize layers of deep convolutional neural networks every few hundred training epochs. Similarly, PECT employs the algorithm to re-parameterize or update the subset of PQC parameters. In the previous section, the global sparsity  $s$  was introduced as a hyperparameter for PECT. In addition to  $s$ , other hyperparameters of the parameter reallocation scheme include  $N_p$ ,  $H_{(0)}$ , and  $\delta$ . The hyperparameter  $N_p$  is the target number of parameters to prune at each iteration. One can think of  $N_p$  as a way of quantifying to what extent the user would like to “shuffle” the parameter subset per PECT iteration. One would assign higher values to  $N_p$  if the circuit at hand is highly redundant and thus the contribution of each parameter is not as significant. While  $N_p$  was fixed in the original formulation, one can in principle implement a schedule to vary  $N_p$  as the iterations progress. Next,  $H_{(0)}$  corresponds to the initial pruning threshold. In our numerical experiments, values of  $H_{(0)}$  were assigned based on the range over which random parameter values were sampled. That is, we chose a threshold value that was relatively small, e.g.  $H_{(0)} = 0.01$  for range  $[0, 4\pi]$  to prevent too many parameters being pruned in the first iteration. For cases in which we employed informative parameter guesses, e.g. using MP2 amplitudes to initialize UCCSD operators [44], we chose values of  $H_{(0)}$  that were slightly less than the average of the parameter guess magnitudes. Lastly,  $\delta$ , which defines the tolerance or neighborhood about  $N_p$  (see Algorithm 1), was fixed to be 0.1 for all of our numerical simulations after observing through preliminary calculations that  $\delta$  had no significant effect on the performance of PECT.

Inputs to the parameter reallocation algorithm are  $\{\psi_l\}$ ,  $\{\phi_l\}$ ,  $M^{(t)}$ ,  $H_{(t)}$ , which correspond to (sparse) representations of the current parameter vector, the size of the parameter subset at the current iterate  $t$ , and the current pruning threshold respectively. The re-definition or reallocation of the parameter subset comprises three main steps: (1) magnitude-based pruning, (2) adjustment of the pruning threshold, and (3) redistribution of free parameter slots. In step (1), for each layer of the PQC at iterate  $t$ , parameters (and their corresponding gates) are pruned or removed from the parameter subset based on the current value of the pruning threshold,  $H_{(t)}$ . That is, if the absolute value of each parameter in  $\phi_l$  is less than the value of  $H_{(t)}$ , the parameter is removed from the parameter subset (from both  $\psi_l$  and  $\phi_l$ ), and its corresponding gate operation is removed from the ansatz substructure. Assuming convergence of the local optimization before the pruning step, the gradient should be zero or near-zero, justifying the removal of parameterized gates with near-zero parameter magnitudes. At this step, the numbers of parameters that are pruned and survived in each circuit layer are stored

and denoted as  $K_l^{(t)}$  and  $R_l^{(t)} = M_l^{(t)} - K_l^{(t)}$  respectively. In the following step of the parameter reallocation subroutine, we adjust the pruning threshold for the next iteration,  $H_{(t+1)}$ , by comparing  $K = \sum_l K_l$ , the total number of parameters pruned, to the target number  $N_p$ :

---

**Algorithm 1:** Pruning threshold update

---

```

Input:  $K, H_{(t)}, N_p, \delta$ 
Output:  $H_{(t+1)}$ 
// Case 1: Too many parameters pruned
1 if  $K > (1 + \delta)N_p$  then
2    $H_{(t+1)} = \frac{H_{(t)}}{2}$ 
// Case 2: Too few parameters pruned
3 else if  $K < (1 - \delta)N_p$  then
4    $H_{(t+1)} = 2H_{(t)}$ 
5 else
6    $H_{(t+1)} = H_{(t)}$ 
```

---

If too many parameters are pruned compared to the target number  $N_p$ , the pruning threshold for the next iteration is lowered, and if too few parameters are pruned, the pruning threshold is increased. The user can customize the rate of increase or decrease for the pruning threshold. In our implementation, we multiply or divide by 2 in respective cases. In the final step of parameter reallocation subroutine, we grow or add  $K$  parameter slots in total (and their corresponding gate operations) to the PQC and update  $\psi_l$ ,  $\phi_l$ , and  $M_l$  for the next iteration. The growth rate for each circuit layer is determined by how many parameters in the original subset “survived” the pruning phase, which was quantified by  $R_l$ . That is, the heuristic growth rate for each layer  $l$  at iterate  $t$  is defined as:

$$G_l^{(t)} = \frac{R_l^{(t)}}{\sum_l R_l^{(t)}} \sum_l K_l^{(t)}. \quad (1)$$

We note that care must be taken into rounding the growth rates for all  $l$  such that the total number of grown parameters is equal to the total number of pruned parameters which was determined from the first step of the parameter reallocation scheme. This keeps the global sparsity  $s$  constant.<sup>1</sup> In practice, parameters are grown by updating  $\psi_l$ , in which  $G_l^{(t)}$  parameter positions or indices that have not been activated or selected in the current  $\psi_l$  are randomly sampled for circuit layer  $l$ . We then initialize the corresponding parameter values in  $\phi_l$  to 0. Lastly, we update the parameter subset size for each layer:  $M_l^{(t+1)} = M_l^{(t)} - K_l^{(t)} + G_l^{(t)}$ . As previously discussed, according to Equation 1, more parameter slots are to be allocated to PQC layers that have greater numbers of surviving parameters. The algorithm deems these

layers as being more important to the optimization than circuit layers that had greater number of gates with parameters near 0 and thus were pruned or removed. While further investigation is needed to determine if circuit layers with fewer pruned parameters are indeed more important, e.g. correspond to parameters with larger gradient magnitudes, this heuristic approach affords us a way of adding free parameter slots to the parameter subset without making calls to the quantum computer, which, especially in the NISQ era, are the main bottlenecks in the performance of variational quantum algorithms.

## B. Prior work

PECT combines and leverages two ideas that have been explored in past studies: dynamic construction of the parameterized ansatz and ansatz reduction by pruning. The idea of adaptively constructing an ansatz for variational quantum algorithms was proposed in ADAPT-VQE [19], in which the method was shown to be effective for preparing ground states of various molecular systems. While ADAPT-VQE iteratively adds or grows a parameterized ansatz from an *unordered* pool of quantum operations based on the corresponding gradients one operation at a time, the core objective in PECT is to find a compact representation of the solution state using a subset of gates from a pre-defined ansatz structure (i.e. an *ordered* pool of quantum operations) and then pruning and growing multiple gates in a single iteration to refine the ansatz substructure. Second, the idea of pruning an ansatz was explored in a previous work that developed the “Ansatz Architecture Search” (AAS) to solve QAOA problem instances using fewer two-qubit gates [47]. In each iteration, the AAS method reduces the current circuit architecture by generating all the unique architectures after removing one two-qubit operation. These pruned architectures are then ranked according to some defined scoring function, and circuit architectures with the best scores are chosen to be candidates for pruning for the following iteration. PECT also involves pruning or removal of gates, but in each iteration, multiple gates can be removed based on their optimized parameter values and the current pruning threshold. In addition to pruning, PECT also involves an ansatz growth phase to reallocate a number of parameters/gates to update the ansatz substructure.

## III. NUMERICAL EXPERIMENTS

In this work, we use the Variational Quantum Eigensolver (VQE) as the VQA-of-choice to demonstrate optimization of parameterized ansatze using PECT. While running a sequence of VQEs may seem like a costly task, each VQE instance likely has a shorter optimization runtime due to fewer circuit parameters to explore and update. In addition, activating a fraction of the parameters

---

<sup>1</sup> For the case in which the updated number of parameters at layer  $l$  for iteration  $t+1$  exceeds the total number of parameters in that layer, i.e.  $M_l^{(t+1)} > N_l$ , we redistribute the excess parameters randomly to different circuit layers, as was done in [40].

likely corresponds to executing circuits with fewer gate operations and lower circuit depth on the quantum computer. Although we did not apply this idea in our simulations, one might also consider establishing a maximum number of measurements to use in each local optimization, and applying techniques such as the one presented in Ref. [23] to optimize the use of measurements throughout the PECT execution. This way, one can guarantee the multiple VQE runs do not exceed a certain measurement budget. For our numerical experiments, we investigate optimizations of three different VQE ansatze: the standard UCCSD ansatz [32], its low-depth and generalized variant  $k$ -UpCCGSD ansatz [37], and the LDCA [36].

### A. Ansatze

The UCC method, a unitary variant of the traditional coupled-cluster method, has been widely studied in the context of simulating quantum chemistry using quantum computers [32, 39, 44]. However, the UCC wavefunction considering singles and doubles excitations (UCCSD) suffers from circuit depths that grow as  $O(N^5)$  with system size. More recently, this depth was reduced by approximately a linear factor using low-rank decomposition and the fermionic swap network while assuming linear connectivity [48]. The number of parameters for UCCSD grows as  $O(N_o^2 N_v^2)$ , where  $N_o$  and  $N_v$  are numbers of occupied and virtual spin orbitals respectively, or in the worst case, as  $O(N^4)$ . Moreover, UCCSD has been shown to be insufficient for describing particular molecular systems [49]. Since then, a recent work developed a shallower and more flexible variant of the UCC method, called  $k$ -UpCCGSD [37].

The  $k$ -UpCCGSD ansatz comprises  $k$  layers of a circuit that prepares a wavefunction considering generalized singles and paired doubles excitations. Compared to UCCSD, the  $k$ -UpCCGSD ansatz has a significantly shorter circuit depth i.e. linear in the system size or  $O(kN)$ , making  $k$ -UpCCGSD better suited for NISQ applications. While the ansatz was able to well describe the ground and the first excited states of various strongly correlated molecular systems in Ref. [37], initialization and optimization of the circuit parameters appeared to be challenging tasks with increasing  $k$ .

Lastly, LDCA is another ansatz with depth scaling linearly with system size introduced by Dallaire-Demers *et al.* [36]. The LDCA uses nearest-neighbor parameterized two-qubit gates that are native to devices with tunable couplers. These templates of nearest-neighbor two-qubit gates are repeated several times, producing a highly compact and nested circuit structure and thus accruing significant numbers of parameters.<sup>2</sup> For instance,

using the original formulation of LDCA, given a 12-qubit VQE problem using 5 superlayers or “cycles” of LDCA, this circuit already has 1662 parameters. By contrast, for an instance of a 12-qubit VQE problem solving for ground state energies of LiH, UCCSD uses 56 parameters, and 2-UpCCGSD uses 60 parameters. While LDCA demonstrated potential to well describe strongly correlated fermionic systems, its application to larger systems has so far been limited by the difficulty in optimization. In our implementation of LDCA, in order to reduce circuit depth, we apply the unitary  $U_{\text{VarMG}}$  defined in Ref. [36] to the Hartree-Fock state but do not apply  $U_{\text{Bog}}^\dagger$  afterwards. To generate wavefunctions that preserve particle number and reduce parameter count, we employed parameter sharing (also known as correlating parameters [16]), in which the  $XX$  and  $YY$  entanglers share the same parameter and the  $XY$  and  $YX$  entanglers share the same parameter with opposite signs, such that each two-qubit LDCA block has three free parameters, instead of five.

To summarize, in terms of circuit depth,  $k$ -UpCCGSD and LDCA have comparable depths (up to some  $k$ ) while UCCSD has a larger circuit depth for given system size. In parameter count, UCCSD has the largest asymptotic scaling though for small systems such as ones considered in this study, it may correspond to the fewest number of parameters out of the three ansatze. Even in cases where the number of parameters for UCCSD may be larger than those for  $k$ -UpCCGSD and LDCA, advantages of using UCCSD include well-motivated techniques for initializing the ansatz parameters and relative ease of optimization. In terms of defining what constitutes a layer for each ansatz to implement PECT, we define UCCSD as having a single layer,  $k$ -UpCCGSD having  $k$  layers (each layer being a repetition of the UpCCGSD circuit), and LDCA having  $L * \frac{N}{2} * (N - 1)$  layers, where  $L$  is the number of LDCA superlayers and  $N$  is the number of qubits. For LDCA, this means that we define each layer for PECT as a two-qubit LDCA block which has three free parameters, as shown in Figure 10 in Appendix A.

In the following subsections, we describe the computational details for VQE wavefunction simulations of LiH (STO-3G) and H<sub>2</sub>O (STO-3G) and show our results comparing VQE optimizations with and without PECT.

### B. Computational details

For testing PECT, we consider two molecular systems: (1) lithium hydride (4 electrons, 12 spin-orbitals/qubits) and (2) symmetric stretching of O-H bonds in a water molecule with a fixed H-O-H angle of 104.5 degrees (8 electrons, 12 spin-orbitals/qubits after freezing core orbitals), both in the STO-3G basis. We run VQE calculations to estimate the ground state energies of both systems using two optimization methods: the traditional method of optimizing the entire ansatz, which we refer to as “non-PECT” calculations, and optimization using

---

<sup>2</sup> We further explain and provide an example of the LDCA circuit structure in Appendix A.

System	Ansatz	Global sparsity, $s$	Parameter count (PECT/non-PECT)	Initial pruning threshold, $H_0$	Target number of parameters to prune, $N_p$
LiH	UCCSD	0.2	45/56	0.0001	8
	2-UpCCGSD	0.5	30/60	0.001	6
	LDCA (L-BFGS-B)	0.73 (rest)	291/1002 (rest)	0.1	15 (rest)
		0.6 (0.9, 1.1 Å)	499/1200 (0.9, 1.1 Å)	0.1	30 (0.9, 1.1 Å)
H <sub>2</sub> O	LDCA (SLSQP)	0.6	420/1002	0.1	30
	UCCSD	0.15 (rest), 0.1 (2.6 Å)	39/46 (rest), 49/54 (2.6 Å)	0.001	5
	4-UpCCGSD	0.3	84/120	0.01	10
	5-UpCCGSD	0.3	105/150	0.01	10
	LDCA	0.6 (rest)	420/1002 (rest)	0.05 (rest)	50 (rest)
		0.65 (2.2 Å) 0.5 (2.4 Å)	440/1200 (2.2 Å) 618/1200 (2.4 Å)	0.1 (2.2, 2.4 Å)	40 (2.2, 2.4 Å)

Table I: Hyperparameters and parameter counts for PECT calculations reported in this study.

PECT. To fairly compare VQE optimizations with and without PECT, we use the same classical optimizers, L-BFGS-B [45] and SLSQP [46], as well as initial parameters for a given ansatz. In addition, we use the same optimizer arguments and options for PECT and non-PECT calculations, e.g. step size used for numerical gradients in L-BFGS-B or the maximum number of function evaluations.

For each ansatz, we employed different parameter initialization strategies. For the UCCSD ansatz, we pre-computed MP2 amplitudes for parameter initialization [44]. With  $k$ -UpCCGSD ansatze for various  $k$  values, we used MP2 amplitudes as initial parameter values for the first layer ( $k = 1$ ) then randomly initialized parameters of subsequent layers by uniformly sampling from the range [-0.1, 0.1]. We employed this parameter initialization technique due to advantages in using informed guesses such as MP2 amplitudes to initialize UCC-like wavefunctions [44], and we assumed that parameter settings of the first circuit layer or  $k = 1$  would be more influential than those of subsequent layers. For LDCA, initial parameters were randomly generated by uniformly sampling from the range  $[0, 4\pi]$ .

We report the PECT hyperparameters for each system and ansatz type in Table I. The fractional tolerance for  $N_p$ , or  $\delta$ , is fixed to 0.1 for all simulations. We chose the global sparsities based on the number of parameters for each ansatz/problem and  $H_0$  based on the initial guesses. For example, for UCCSD, the pre-computed MP2 amplitudes for LiH were small in magnitude. To prevent too many parameters being pruned in the first few local optimizations, we set  $H_0$  for those simulations to be slightly less than the average of the magnitudes of the MP2 amplitudes. Lastly, we chose the values of  $N_p$  based on the number of parameters and our assumptions on the relative parameter redundancy of each ansatz.

### C. LiH

We show the results for estimating ground state energies of LiH over various bond lengths using the three types of ansatz in Figure 2. All of the calculations, both PECT and non-PECT optimizations, shown in this Figure employ the L-BFGS-B as the (local) optimizer. We observe that the UCCSD ansatz optimized without PECT achieves the highest energy accuracy but corresponds to the deepest circuits across all bond lengths. With PECT, it is possible to reduce the total depth by up to approximately 20% in some cases, although at the cost of losing some accuracy in the final energy. With 2-UpCCGSD, we achieve energy values with errors that fall below the chemical accuracy threshold (i.e.  $\leq 1$  kcal/mol or  $1.6 \times 10^{-3}$  Ha) while maintaining substantially lower circuit depths. With PECT, 2-UpCCGSD further reduces circuit depths but also at the cost of higher ground state energy estimates. Still, for both UCCSD and 2-UpCCGSD, errors using PECT are below the chemical accuracy threshold. Lastly, LDCA optimized with PECT is shown to have comparable depths to 2-UpCCGSD employing PECT. Naively, this ansatz with 5-6 superlayers has 1002-1200 parameters respectively. A direct optimization of this number of parameters using the L-BFGS-B optimizer did not achieve convergence despite multiple attempts with randomly initialized parameters. In contrast, using PECT, we were able to reduce the number of parameters to 291 at most bond lengths (and 499 at two bond lengths), as shown in Table I. Although for a few bond lengths we were not able to achieve chemical accuracy with PECT, we note that without PECT, optimization of the LDCA ansatz was not possible at any bond length. We also note that the number of function evaluations using this optimizer were often around the maximum number allocated, or  $5 \times 10^5$ .

To improve the performance of optimizing LDCA, we explored an alternative local optimizer and global sparsities. Using the SLSQP optimizer and lowering the global

sparsity from 0.73 to 0.6 for all bond lengths (i.e. increasing parameter count from around 291 to 420 parameters), we were able to improve the energies at most bond lengths with all the energy errors along the potential energy surface falling below chemical accuracy, as shown in Figure 3. In addition, using these PECT settings, we reduced the number of function calls on average by approximately half, compared to the LDCA calculations using L-BFGS-B with PECT. In Figure 3, we also tried optimizing LDCA with SLSQP but without PECT to demonstrate that optimization of LDCA at this system size and number of layers needs a strategy like PECT to achieve reasonable energies.

#### D. Symmetric stretching of O-H bonds in H<sub>2</sub>O

Symmetric dissociation of water is a common benchmark for the classical multi-reference methods [50–53] and has been recently studied in the context of  $k$ -UpCCGSD by Lee et al. [37] and by Sokolov et al. [54] as a test for the orbital optimized UCC methods. Similarly structured as the previous subsection, we show the VQE optimization results for the symmetric stretching of O-H bonds in a water molecule in Figure 4. For optimizing LDCA ansatze, SLSQP was employed as the local optimizer based on the relative performance of SLSQP compared to L-BFGS-B in VQE results of LiH (Figure 3). On the other hand, for UCCSD and  $k$ -UpCCGSD where  $k = 4, 5$ , we employed L-BFGS-B as the local optimizer after preliminary results showed lower final energies obtained using L-BFGS-B over SLSQP. Plots in row (a) in Figure 4 show that in the case of UCCSD, the ground state energy estimates become worse as the O-H bonds are stretched, which is consistent with results from a past study investigating the performance of UCCSD in the same system [49]. In addition, some of the energy error may also come from an unfavorable ordering of cluster operators when the UCCSD unitary is Trotterized in our implementation [55, 56].

Regarding the performance of PECT, it was interesting to note that, unlike in optimizations of LiH ansatze in which using PECT consistently produced higher ground state energy estimates, using PECT to optimize the UCCSD wavefunction yielded significantly lower energy than optimization without PECT when  $R_{\text{O-H}} = 2.4\text{\AA}$ . Additionally, at this bond length, PECT found a subset of gates with the same circuit depth as the un-pruned version of the ansatz.

Regarding the  $k$ -UpCCGSD ansatz, while  $k = 2$  was sufficient to reproduce the LiH ground state energy, we observed that up to  $k = 4$  and  $k = 5$  layers were needed in the case of water. As noted in the original paper [37], optimizing  $k$ -UpCCGSD ansatz for higher  $k$  proved to be challenging as the quality of the ansatz becomes highly sensitive to the initial values of the ansatz parameters. In Figure 4, plots in rows (b) and (c) show the energy errors, circuit depths, and number of function call for

the  $k = 4$  and  $k = 5$  instances of the ansatz, respectively. These results show that PECT improves the accuracy for some bond lengths but not others. In all cases, however, it achieves reductions in the circuit depths.

For those geometries in which the optimizer was unable to produce final energies with errors below the chemical accuracy threshold, this may be due to unfavorable initial points, versus the limited expressiveness of the ansatz. To support this claim, we ran multiple trials of a 4-UpCCGSD PECT-based optimization from ten different initial points, in which one of the trials achieved chemical accuracy, proving the sufficient expressiveness of the ansatz. We describe this numerical experiment in further detail in Section IV B. Final energies using instances of the  $k$ -UpCCGSD ansatz were generally lower than energies using the UCCSD ansatz in plots (a), especially in regions of the potential energy surface where UCCSD had high errors. However, at  $k = 4$  and  $k = 5$ , the resulting circuit depths of the  $k$ -UpCCGSD ansatz are comparable to or higher than the depths of the UCCSD ansatz. Comparing optimizations of  $k$ -UpCCGSD with and without PECT, we observe that PECT converges to comparable or lower energies in the region where O-H bond is stretched (e.g.  $R = 2.4\text{\AA}$  for  $k = 5$  in Figure 4(c)) using a fraction of the total circuit depth and comparable number of function calls. To improve the performance of optimizing  $k$ -UpCCGSD using PECT, further investigation is needed to find better strategies for initializing parameters of the ansatz. In this study, we observe that initializing the parameters of the first layer ( $k = 1$ ) with corresponding MP2 amplitudes performed better than when the entire parameter vector was randomly initialized. Nevertheless, even with this initialization scheme, we do not observe reliable convergence to low energies or energies within chemical accuracy.

Lastly, we considered the LDCA ansatz to estimate the ground state energies of water. From the LDCA optimization results for LiH in Figure 3 and initial numerical experiments, we observed high energy errors when optimizing LDCA without PECT.<sup>3</sup> This implies that to estimate the ground state potential energy curve of water, which is more difficult to describe than that of LiH, it is infeasible to optimize the corresponding LDCA circuit without PECT. Thus, we limited our VQE optimizations of LDCA for water to ones employing PECT, as shown in Figure 4(c). For all the bond lengths considered, we were able to generate VQE-optimized circuits that corresponded to energy errors falling below the chemical accuracy threshold. With the exception of ground states at two O-H bond lengths, all PECT-based optimizations were executed on 5 superlayers of LDCA, which originally has depth 1382 and two-qubit gate count of 3300. Using PECT, the effective circuit depth and gate count are re-

---

<sup>3</sup> The average VQE energy error across bond lengths of LiH using the LDCA wavefunction without PECT was approximately 0.04 Hartrees.

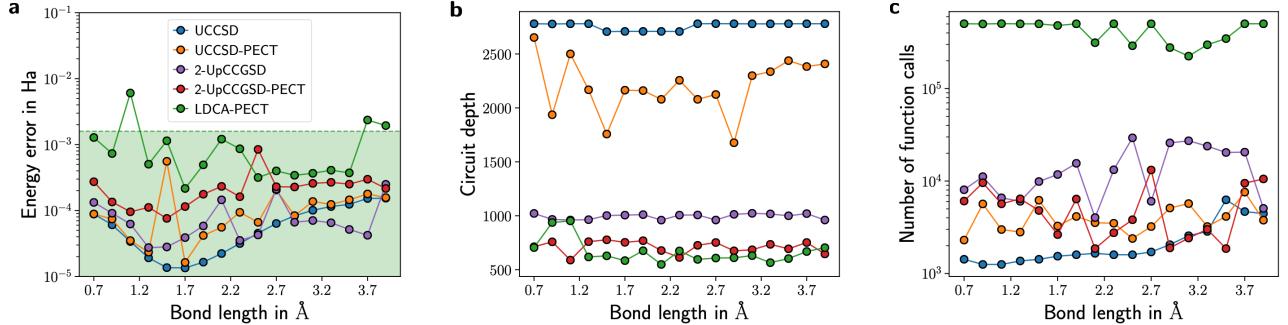


Figure 2. VQE performance metrics computed for estimating ground state energies of LiH at various bond lengths. (a) Energy errors are shown for different ansatze that were optimized without or with the PECT strategy. The optimizer used is L-BFGS-B. The green shaded region indicates energies below chemical accuracy. (b) Depths of corresponding circuits. For circuits optimized using PECT, we report the depth averaged over PECT iterations. (c) Number of total function calls for each ansatz.

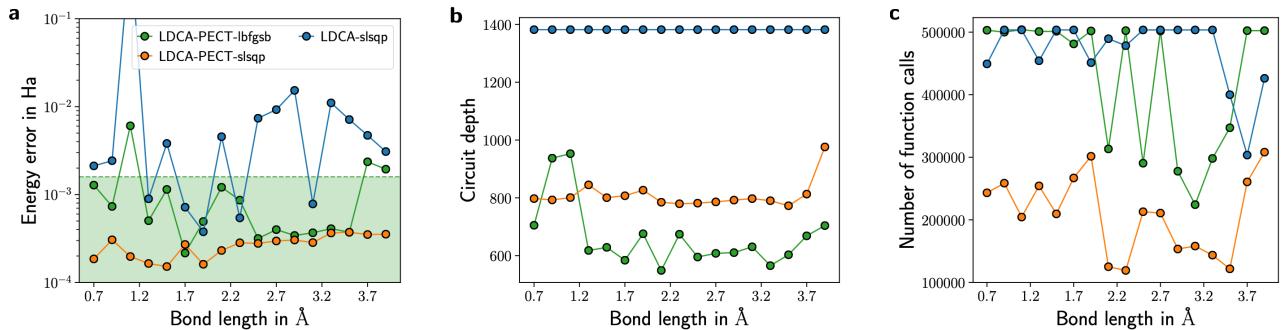


Figure 3. Comparing VQE performance metrics computed for estimating ground state energies of LiH by optimizing LDCA with PECT but employing two different local optimizers: L-BFGS-B (green) and SLSQP (orange). We additionally plotted the VQE performance metrics for LDCA optimized without PECT (blue) using the SLSQP optimizer to highlight the advantages of using PECT for optimizing LDCA circuits. (a) Energy errors for different PECT optimizers and global sparsities ( $s = 0.73$  and  $0.6$  for L-BFGS-B and  $s = 0.6$  for SLSQP optimizers). The green shaded region indicates energy errors below chemical accuracy. (b) Average depths of corresponding circuits. (c) Number of function calls for each PECT calculation.

duced, as discussed in more detail in Section IV C. For the two O-H bond lengths, indicated on Figure 3(d) in red boxes, 6 superlayers of LDCA were needed to prepare the ground state, which corresponded to simulating circuits with depths of around 1658 and two-qubit gate counts of around 3960. Compared to the other two types of ansatz, LDCA circuits had at least half the circuit depths though they required significantly more function evaluations to optimize. This is due to the greater number of parameters LDCA has compared to the other ansatze for the instances we studied. Computing gradients, analytical or numerical, requires two function evaluations per parameter. Based on our observations and results, it appears that while LDCA circuits take longer to optimize, they are less sensitive to parameter initialization compared to  $k$ -UpCCGSD when optimized with PECT. For each bond length, fewer than 5 random parameter initializations for LDCA were used to generate the plots in Figure 4(d)

which report the best results.<sup>4</sup> We explore this further in Subsections IV A and IV B.

#### IV. DISCUSSION

From the previous section, we gained insight on each ansatz' ability to reproduce the potential energy surfaces of LiH and H<sub>2</sub>O using PECT. Here, we provide further analyses on PECT, including its advantages in robustness of parameter initialization for LDCA as well as general reductions in circuit resources (i.e. depth and two-qubit

<sup>4</sup> We set the PECT hyperparameters for optimizing LDCA for water based on corresponding values used for optimizing LDCA for LiH. Likewise, when optimizing e.g. two circuits of similar widths and depths using PECT, it may be possible to infer the hyperparameters for one run based on the other.

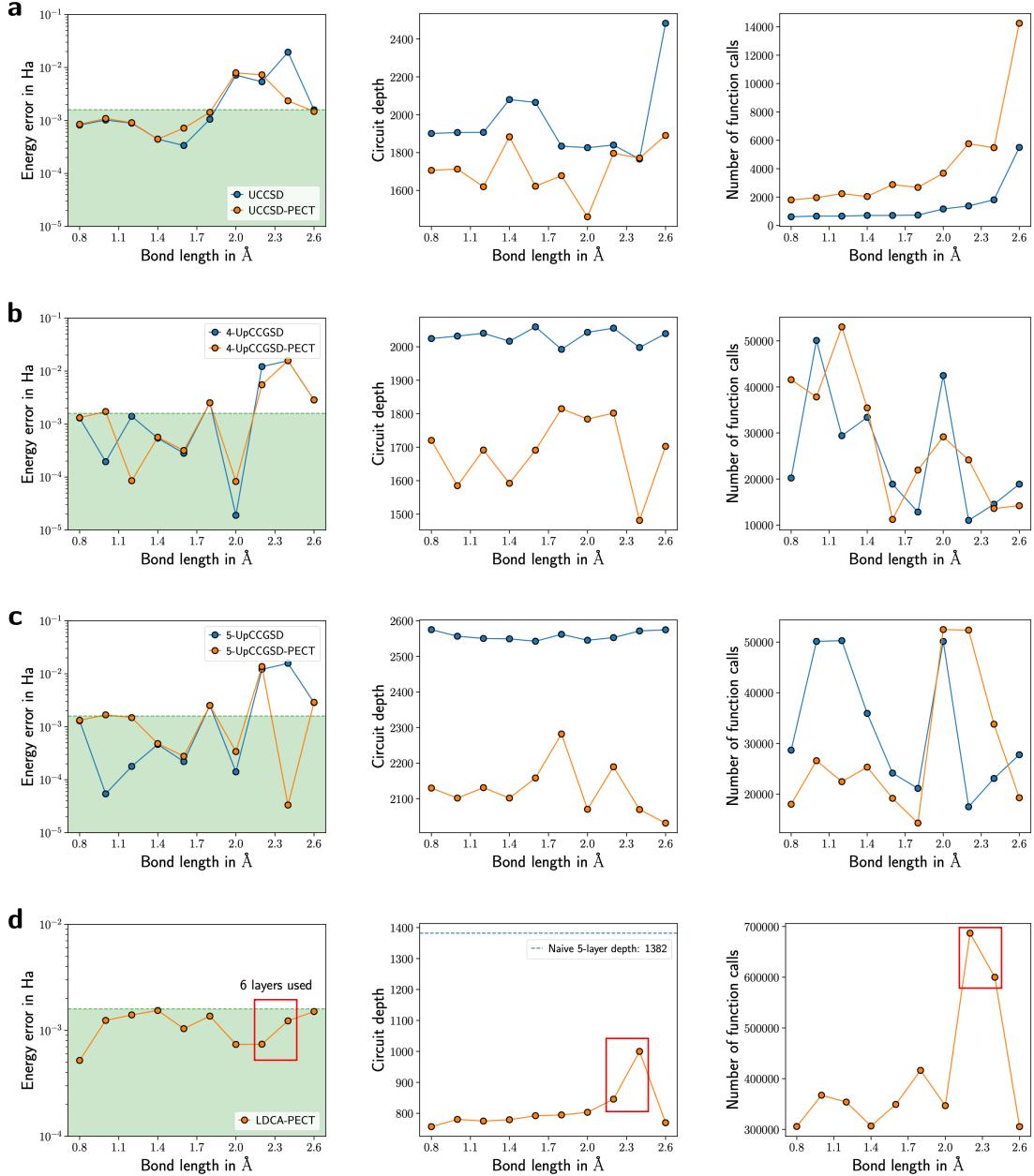


Figure 4. VQE performance metrics computed for estimating ground state energies of  $\text{H}_2\text{O}$  using various ansatze: (a) UCCSD, (b) 4-UpCCGSD, (c) 5-UpCCGSD, and (d) LDCA. For two of the bond lengths as shown in the red boxes in (d), 6 superlayers of LDCA were used with PECT as using 5 superlayers led to poor final energies. For each ansatz, the same initial parameters are used for PECT and non-PECT methods. We show the resulting potential energy surface for each ansatz in Appendix D.

gate count) and in optimization runtimes. Our PECT calculations also reveal the relative ‘‘prunability’’ of each of the three ansatze and provided numerical evidence of how parameters of early layers of multi-layered PQCs converge to their final values before parameters of later layers do. We then propose a way for extending PECT to circuits with more parameters by combining the method with the layerwise circuit training strategy [28, 41, 57]. We test this combined strategy by optimizing VQE circuits to estimate the ground state densities of the one-

dimensional Fermi-Hubbard model across various chain lengths.

#### A. Robustness to random parameter initialization in LDCA

In this section, we investigate the robustness to parameter initialization by applying PECT to optimize LDCA circuits. While depth-efficient, LDCA accrues a large

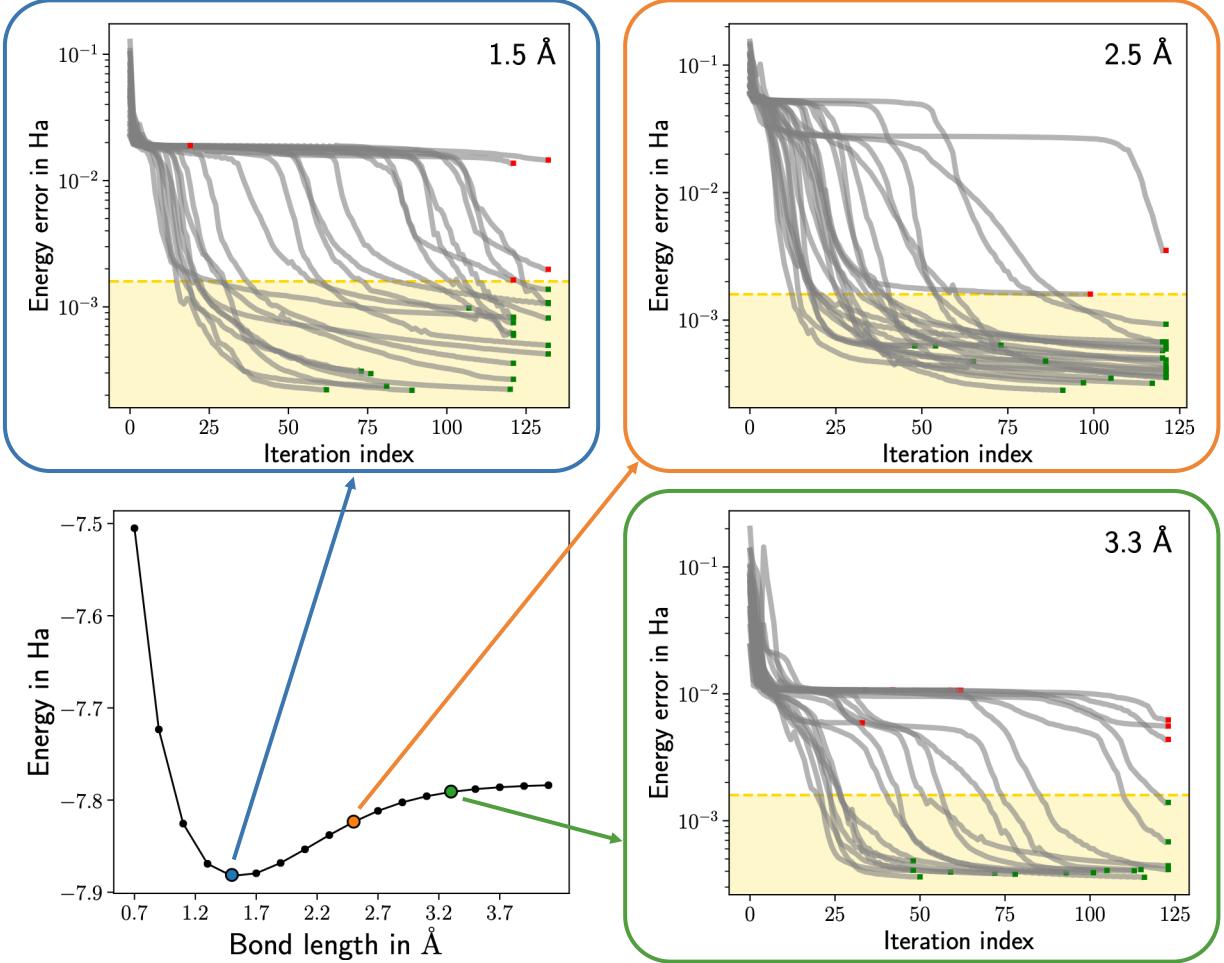


Figure 5. Energy convergence for finding the ground state of LiH using the LDCA circuit. Twenty-five random parameter initializations were executed for each of three bond lengths to ensure PECT can provide high-quality solutions with high probability. The green square at the end of each optimization curve indicates a “successful” run, in which the final energy error fell below chemical accuracy. Yellow shaded region indicates energy errors below chemical accuracy threshold. The x-axis refers to the iteration index of PECT. For these simulations, we fixed the maximum number of function evaluations to  $5 \times 10^5$ .

number of parameters. Using PECT, however, we numerically show that optimization runs are robust to random initialization. In Fig. 5, we consider three different bond lengths of LiH:  $R = 1.5, 2.5$ , and  $3.3\text{\AA}$ . For each bond length, we execute 25 randomly initialized optimizations of the 12-qubit, 5-superlayered LDCA which naively has 1002 parameters. Using PECT with the L-BFGS-B optimizer, we fixed the global sparsity  $s$  to 0.73 such that we were optimizing around 291 parameters during each local optimization. At bond length of  $1.5\text{\AA}$ , 20 out of 25 runs achieved final energies with errors that fell below the chemical accuracy threshold. For  $2.5\text{\AA}$ , 23 out of 25 runs succeeded, and for  $3.3\text{\AA}$ , 16 out of 25 runs. The smaller success ratio for  $R = 3.3\text{\AA}$  is expected due to the increasing difficulty of describing the ground state as the Li-H bond is stretched. That is, as the interatomic distance increases, the ground state wave function acquires a multi-

reference character. This may have implications for the quality of the ansatz and likely changes the “roughness” of the objective function landscape, e.g. addition of local minima. For a parameter-heavy ansatz such as LDCA, we noted relatively high success probabilities at the three bond lengths. This not only highlights the advantage of employing PECT for optimization but also sheds light on the amount of redundancy in the parameterization of LDCA, which we discuss in the following subsection.

## B. Prunability of ansatz

From our optimizations of UCCSD,  $k$ -UpCCGSD, and LDCA using PECT, we were able to identify which ansatz is more “prunable,” that is, well-suited for optimization strategies such as PECT that utilizes prun-

ing or removal of gates. Specifically, PECT seemed the most effective strategy for LDCA out of the three types of ansatze investigated. For instance, in the case of estimating the ground state energy of the water molecule, for all the bond lengths considered, PECT was able to estimate energies with errors below the chemical accuracy threshold with fewer than 5 trials of random initial parameters, with several runs converging on the first trial. For  $k$ -UpCCGSD with PECT, the optimization remained sensitive to initial parameter values as well as the sets of gates selected during the PECT procedure. To test the sensitivity, in the case of 4-UpCCGSD for  $R_{O-H} = 2.2\text{\AA}$ , where in Figure 4(b) we observed a high energy error, ten independent PECT optimizations were run with random parameter initializations. These optimizations were originally designed to test whether the high energy error observed was due to the deficiency of the ansatz or the sensitivity to parameter initialization. Out of the ten trials, only one run achieved a final energy with error below chemical accuracy, while the average energy error across the ten trials was approximately 0.01 Hartrees with standard deviation of 0.005 Hartrees. With most trials producing energy errors that are close to the mean, this confirmed the sensitivity to initial parameters. One open question to investigate would be whether using PECT makes  $k$ -UpCCGSD *less* sensitive to parameter initialization than optimizing the ansatz without PECT.

From our results, we determined the UCCSD ansatz is the least prunable ansatz of the three. For the sizes of problems investigated in this work, UCCSD, though high in depth, is the most compact in terms of its parameterization. That is, the numbers of parameters are already low enough such that optimizing subsets of parameters is unnecessary. In addition, the bigger challenge in using UCCSD is not in its optimization but in its expressibility or capability of describing ground states of (strongly correlated) molecular systems. As noted earlier, effectiveness of UCCSD also highly depends on the ordering of Trotterized cluster operators [55, 56]. Therefore, because PECT assumes a fixed ordering of operations (i.e. operates based on a pre-defined ansatz structure), a method such as ADAPT-VQE, which chooses operations from an unordered pool, may be better suited for optimizing an ansatz comprising cluster operators.

Comparing the performance of PECT for LDCA and  $k$ -UpCCGSD circuits in terms of energy error, the difference is likely due to the higher degree of redundancy in the LDCA circuit, compared to the  $k$ -UpCCGSD circuit. That is, even if we remove some parameter  $\theta$  in the LDCA circuit optimization, there are likely several other parameters that can replace  $\theta$ , i.e. that can be tuned with the same effect as tuning  $\theta$ . In addition, this may help explain the robustness in parameter initialization for LDCA observed in Figure 5. In conclusion, our analysis of the three ansatze indicates that PECT is *not* an optimization strategy that will work well for any parameterized quantum circuit. For optimal performance, the circuit should be parameter-heavy and contain some amount of

redundancy in its parameterization. Fortunately, several PQC designs, especially ones that are efficient to particular hardware architectures and connectivities [41, 58], satisfy these conditions.

### C. Ansatz cost reduction

During the local optimization phase of PECT, a subset of parameters of the PQC is optimized while values of other parameters are set to 0 corresponding to identity operations. This likely results in an ansatz substructure that has reduced circuit depth and lower (two-qubit) gate count compared to the original ansatz with full parameter count. For instance, for optimization of the UCCSD wavefunction estimating LiH ground states, using 45 out of 56 parameters for PECT allowed for a 20% reduction in circuit depth and 20% reduction in two-qubit gate count, averaged over optimizations at different bond lengths. For optimization of 2-UpCCGSD, using 30 out of 60 parameters produced a reduced ansatz with 28% reduction in depth and 42% fewer two-qubit gates. Lastly, with LDCA, using 420 out of 1002 parameters led to a 42% reduction in circuit depth and 58% reduction in two-qubit gate count.

For optimizations of ground states of the water molecule, it is difficult to make equally meaningful statements on the reductions in circuit resources using UCCSD and  $k$ -UpCCGSD as we have done for LiH due to energy errors. With the UCCSD ansatz, there were regions in the potential energy surface in which both SLSQP and SLSQP-with-PECT were not able to converge to chemically accurate energies. With  $k$ -UpCCGSD where  $k = 4$  and 5, PECT improved accuracies with respect to regular optimization only for certain geometries. Nevertheless, we computed the percent reductions in circuit depths and two-qubit gate counts for UCCSD, 4-UpCCGSD, and 5-UpCCGSD averaged over all considered bond lengths. Optimizing the UCCSD ansatz with PECT led to an average of 12% reduction in depth and 12% reduction in two-qubit gate count. With 4-UpCCGSD, using PECT corresponded to an average of 17% reduction in depth and 31% reduction in two-qubit gate count. Using PECT with the 5-UpCCGSD ansatz, we observed an average of 17% reduction in depth and 32% reduction in two-qubit gate count. To provide a fair analysis of the reductions in circuit resources, in Appendix B we report reductions in circuit depth and gate count for bond lengths at which both non-PECT and PECT based optimizations achieved final energies that are chemically accurate. With LDCA, because we did not optimize the ansatz without PECT, we report the depths of the full 5- and 6-superlayered circuit and show that using PECT, the circuit depths were significantly reduced. For bond lengths at which 5 superlayers of LDCA were used, there was an average of 42% reduction in depth compared to the depth of the full 5-superlayered LDCA, and for 6 superlayers of LDCA, there was an average of

39% reduction in depth. The reader may wonder why for certain bond lengths, e.g.  $R_{O-H} = 2.2$  and  $2.4\text{\AA}$ , we used 6 superlayers instead of 5 superlayers with PECT if the resulting circuit depths at those bond lengths are still lower than those of a fully-dense 5-superlayered LDCA circuit. We point out that the PECT optimization acting on 5 superlayers of LDCA did not achieve chemical accuracy for these geometries, while the PECT optimization on 6 superlayers did. The final circuit of the latter optimization contains gates from the sixth LDCA superlayer.

We showed in various instances that PECT is able to produce more depth-efficient circuits compared to the original ansatz. In theory, such depth reductions make circuits generated using PECT more resilient to noise in the quantum device. We leave analysis of PECT under noise to future work.

#### D. Optimization cost estimation

In this section, we discuss estimating the cost associated with optimizing a parameterized quantum circuit in terms of the number of objective function evaluations with the goal of *comparing* optimization runtimes of different ansatze as opposed to providing accurate estimates of the actual runtimes. The total time associated with the VQE optimization step can roughly be estimated as:

$$t_{\text{opt, est}} = N * t_{\text{circuit}} * f_{\text{eval}}, \quad (2)$$

where  $N$  is the number of circuit repetitions or shots per function evaluation,  $t_{\text{circuit}}$  is the time required to execute the circuit, and  $f_{\text{eval}}$  is the number of function evaluations. While small in magnitude, circuit execution time is a significant component in estimating the overall optimization runtime due to its relation to the circuit depth. Deeper circuits, especially in the NISQ era, will execute with lower circuit fidelities, corresponding to lower-quality estimates of the objective function values. We expect optimizations using noisier estimates to be more difficult and expensive, e.g. requiring more function evaluations. To reflect this, we can further approximate the total optimization time by replacing  $t_{\text{circuit}}$  with the circuit depth, which indirectly accounts for the expected increase in cost associated with optimizing noisier circuits:

$$t_{\text{proxy}} \approx N * D * f_{\text{eval}}, \quad (3)$$

where  $D$  is the circuit depth. The PECT scheme provides a way to potentially reduce this time by reducing the circuit depth required at each function evaluation though this may come at the cost of a greater number of function evaluations to search for an effective parameter subset. For PECT, the time proxy can be defined as:

$$t_{\text{proxy, PECT}} \approx \sum_i N_i * D_i * f_{\text{eval, i}}, \quad (4)$$

where  $i$  runs over the number of local optimizations. Assuming the required precision for the cost function is the same for all local optimizations, we simplify  $N$  or  $N_i$ , the number of circuit repetitions or measurements, as being constant for each VQE sub-module and set it to 1. In practice, the number of circuit measurements for a VQE experiment is expected to be very high (and thus time-intensive) for variational quantum algorithms and will likely be a large multiplicative factor. For an in-depth runtime analysis especially in a cloud-based quantum computation setting, it may be necessary to use a more sophisticated runtime model that considers sampling, circuit batching, and network latency [25].

We compute the following runtime proxies (Equations 3 and 4) for non-PECT and PECT optimizations respectively for the three ansatze solving for the ground states of LiH. These runtime estimates are shown in Figure 6. This plot indicates that using PECT can help speed up optimizations of 2-UpCCGSD and LDCA for most or all bond lengths, but for an ansatz like UCCSD, it is better, in terms of runtime, to execute non-PECT optimization using the entire circuit. However, as noted earlier, in the NISQ era, reducing circuit depth may be a priority over reducing the overall optimization runtime due to noise in the device, which increases with the number of gate operations and circuit depth. Because PECT reduces the effective circuit depth and gate counts during the optimization, it increases the fidelity of the ansatz circuit and therefore could improve the optimization results and their accuracy.

For the water molecule, the UCCSD and  $k$ -UpCCGSD ansatze did not achieve chemically accurate energies for several bond lengths. We report the estimated runtimes of UCCSD over different bond lengths in the top plot of Figure 7. For all the bond lengths at which both non-PECT and PECT calculations reached chemically accurate energies, indicated using bars with hash patterns, we observe that using PECT increased the optimization times. This is expected due to the compact parameterization of UCCSD for the systems considered in this work. Like UCCSD, both 4-UpCCGSD and 5-UpCCGSD were not able to produce chemically accurate energies for all considered bond lengths, with or without PECT. However, while we believe the reason for the lack of convergence to accurate energies in UCCSD, especially as the O-H bond is stretched, is due to the inherent limitations in the expressiveness of the ansatz, 4-UpCCGSD and 5-UpCCGSD were unable to reproduce accurate energies likely due to the sensitivity in parameter initialization. With more trials testing different random parameter initializations, as we tried in Section IV B,  $k$ -UpCCGSD can converge to energies with errors below chemical accuracy, as was shown in Ref. [37]. The middle and bottom plots of Figure 7 show runtime estimates for 4-UpCCGSD and 5-UpCCGSD, respectively. These results suggest that the benefit of PECT in reducing runtime is more apparent for larger  $k$  in the  $k$ -UpCCGSD ansatz. Lastly, while we did not run optimizations for LDCA without PECT

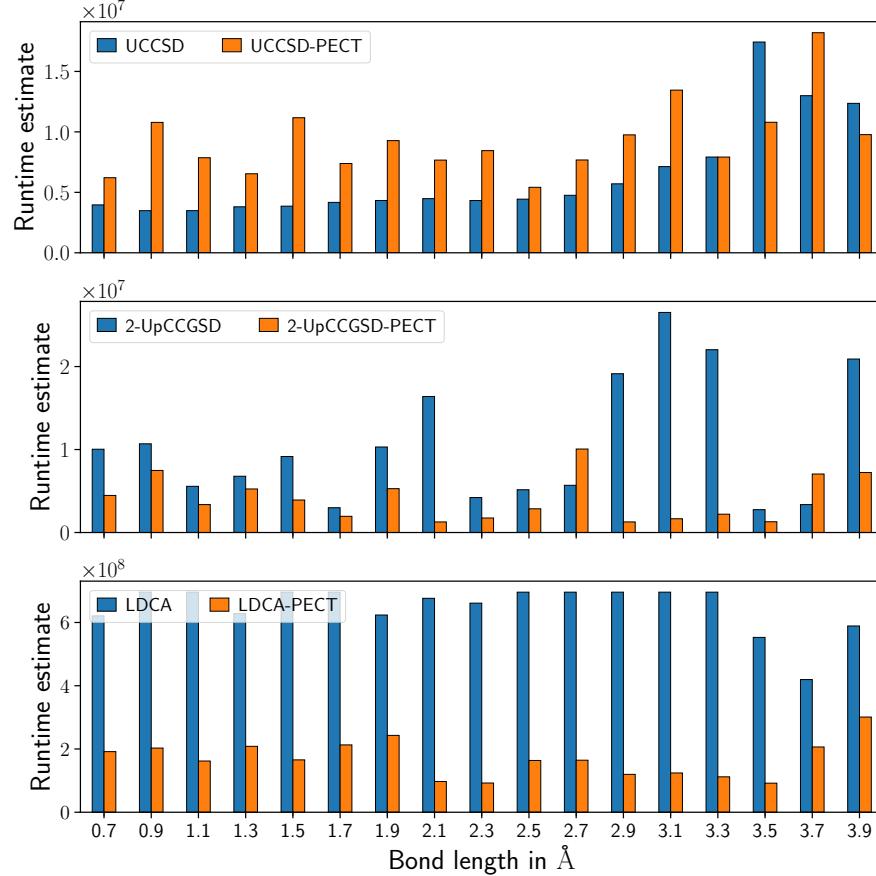


Figure 6. Runtime estimates (depth  $\times$  number of function calls) of LiH VQE calculations for the three ansatze, with and without PECT, shown in orange and blue colored bars respectively. Definitions for runtime estimates are shown in Equations 3 and 4 for non-PECT and PECT optimizations respectively.

based on results for LiH, we expect significant reductions in the runtime, in addition to producing accurate energies, as was observed for LiH in the bottom plot of Figure 6.

### E. Scaling up: layerwise PECT

As we approach VQE problems using ansatze with larger parameter counts, PECT will also suffer from inability to sufficiently explore and optimize in the parameter space. In such cases, it may be useful to combine the “layerwise” training [28, 57] strategy with PECT. To employ a simplified version of the layerwise training scheme, we first divide up the parameterized ansatz into  $k$  layers and employ PECT to optimize each layer. Once the PECT procedure is complete for the  $j$ -th layer ( $j \in [k]$ ), the final parameter values of the layer are “frozen,” and the  $j + 1$ -st layer is optimized with PECT. While fixing parameters of the previous layers may lead to local minima, we observe that relaxing these parameters in the optimization of subsequent layers can also lead to preparing a state corresponding to a higher energy. It

may then take many function calls to go back to the previous state, which can increase the overall cost of optimization. We initialize the  $j + 1$ -st layer with parameter values sampled uniformly from the range  $[0, 0.02]$  such that it constructs a near-identity operation to avoid potentially expending many function evaluations to return to the energy achieved with  $j$  layers.

We demonstrate the utility of layerwise PECT by running VQE optimizations for estimating the energy densities of the one-dimensional Fermi-Hubbard systems studied in Ref. [41]. In this work, the one-dimensional Fermi-Hubbard model (FHM) at half-filling was proposed as a well-defined benchmark system to test the capabilities of NISQ devices [41]. That is, VQEs on various chain lengths can be executed on a processor to determine the largest chain length for which the quantum computer can reliably approximate the ground state energy. For the variational ansatz, the authors suggested a parameterized quantum circuit assuming a two-dimensional grid layout of qubits using qubit connectivity and gate operations that are native to the Sycamore processor, which has recently demonstrated the ability to outperform classical computers at certain tasks [59]. While we refer the read-

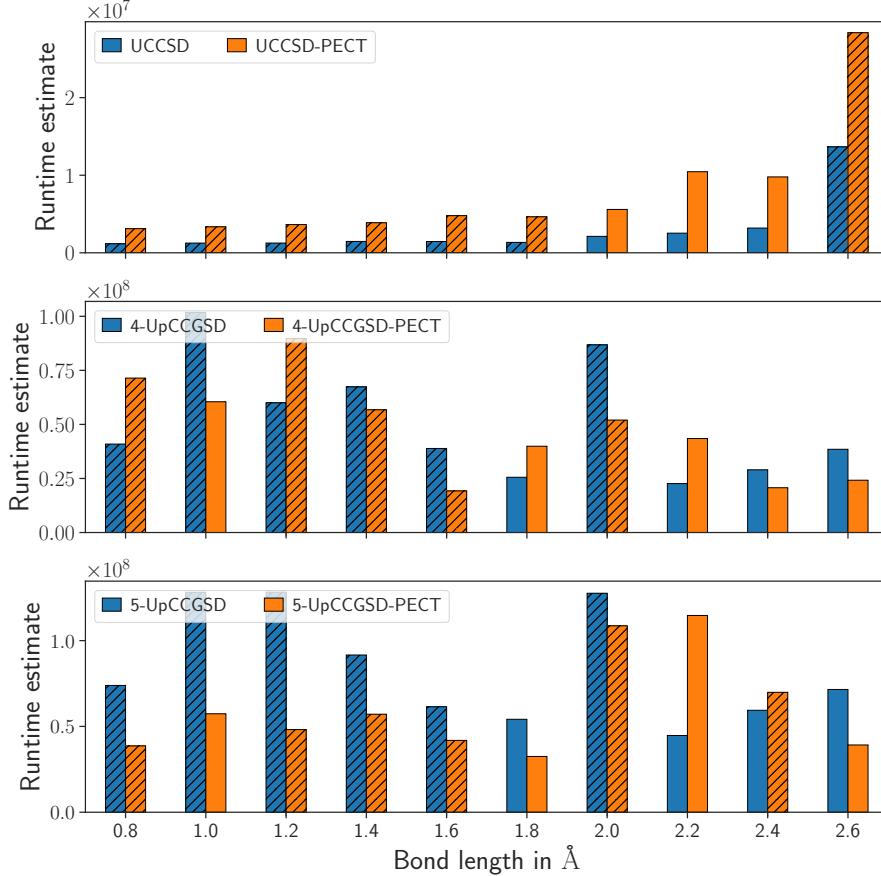


Figure 7. Runtime estimates (depth  $\times$  number of function calls) of H<sub>2</sub>O VQE calculations for UCCSD, 4-UpCCGSD, and 5-UpCCGSD. For each type of ansatz, we report runtimes with and without PECT, shown using orange and blue colored bars respectively. Definitions for runtime estimates are shown in Equations 3 and 4 for non-PECT and PECT optimizations respectively. Unlike LiH calculations, optimizations at certain bond lengths did not converge at chemically accurate energies. We distinguish optimizations that have reached chemically accurate final energies with hash patterns on bars.

ers to Ref. [41] for further details on the system as well as the ansatz, in this work, we present VQE optimization results that show substantial improvement in the circuit depths while producing comparable energy density errors.

In Figure 8, we show the energy density error as a function of the number of circuit layers, where each circuit layer is defined in [41]. The original results from Ref. [41], shown in blue markers, employed layerwise training (without PECT) with the SLSQP optimizer. We additionally note that their layerwise training method did not freeze parameters of previous layers. Each “layer” in layerwise training comprised four sub-layers of the proposed ansatz. Using the layerwise PECT approach, where each “layer” comprises two sub-layers of the ansatz, the optimization can converge to comparable energy densities for different chain lengths using significantly fewer circuit layers, or equivalently, lower circuit depths. Despite the reduction in circuit depth, as noted in previous results, PECT required greater numbers of function evaluations before convergence. These numerical experiments exem-

plify the potential trade-off that exists between reducing the circuit depth and increasing the cost of evaluating cost function and gradient values (here represented by the number of function calls) when using PECT as the optimization strategy. That is, if reducing circuit depth is an important objective in a VQE experiment and the large numbers of function calls is affordable, then PECT may be a promising optimization technique. However, if reducing the number of function calls is more important, then PECT may not be as effective as an optimization strategy. In the near term, due to the rapid accrual of noise in NISQ devices, we expect circuit depth reduction to be more critical in the application of variational algorithms. Furthermore, while potentially expensive in the number of function calls, the overall measurement cost of PECT could be mitigated using measurement-aware optimization strategies, such as the one proposed in Ref. [23].

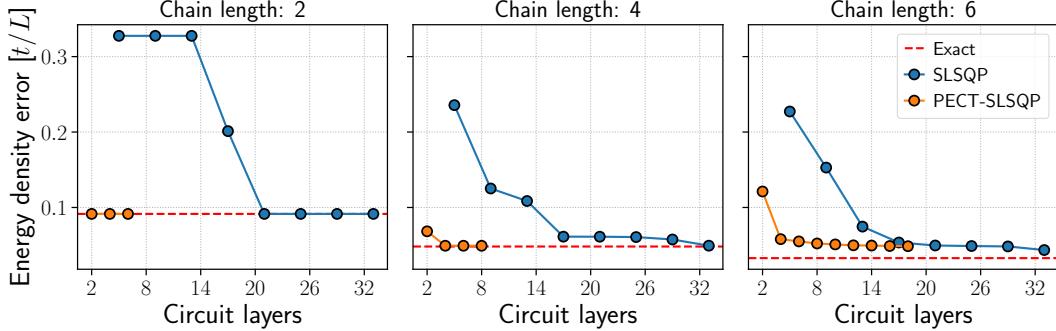


Figure 8. PECT results for optimizing VQE calculations for the one-dimensional Fermi-Hubbard model. Energy density errors (in units of kinetic energy,  $t$  over  $L$  sites) with respect to the infinite chain energy density are plotted against number of circuit layers of the ansatz proposed in Ref. [41]. Blue markers indicate addition of new circuit layers that are layerwise-optimized using the SLSQP optimizer (data points are obtained from Figure 9 of Ref. [41]). Orange points indicate addition of new circuit layers that are layerwise-optimized using PECT employing SLSQP as the local optimizer. The red dashed line indicates the exact energy density at a given chain length.

#### F. Parameter dynamics of multi-layered PQCs

While we expect earlier PQC layers to influence the optimization of later layers, we provide numerical evidence of this phenomenon by investigating how parameters of each layer of a PQC are optimized to their final values. In Figure 9, we show the averaged convergence of parameters in each layer for two optimization instances, one using PECT and the other without.<sup>5</sup> We plot the absolute difference of parameter values at  $t$ -th optimization iteration from their corresponding final values, which are then averaged over parameters in circuit layer  $l$ . That is, the averaged absolute difference of parameter values for the  $l$ -th circuit layer at the  $t$ -th iteration is defined as:

$$|\Delta\theta|_l^{(t)} = \frac{1}{N_l} \sum_{i=0}^{N_l} |\theta_i^{(t)} - \theta_i^{(t_{\text{final}})}| \quad (5)$$

where index  $i$  enumerates parameters of layer  $l$ , and  $N_l$  is the total number of parameters in layer  $l$ . Thus,  $\theta_i^{(t)}$  is the  $i$ -th parameter value in layer  $l$  at iteration  $t$ , and  $\theta_i^{(t_{\text{final}})}$  is the final/optimized  $i$ -th parameter value in layer  $l$ . This quantity represents the average movement of parameter values per circuit layer. From Figure 9, we observe for both PECT and non-PECT optimizations that parameter values of earlier layers start closer and stray less from their final values, on average, compared to the movements of parameters of later layers. This appears to suggest the importance of strategically initializing the first few layers such that parameters of later layers, which change values by greater magnitudes throughout the optimization, can more efficiently refine or improve upon circuit parameter settings of previous layers. We clarify

that the optimizations we consider in Figure 9 do not employ layerwise training. We performed this analysis for other PECT and non-PECT optimization runs, with several results shown in Appendix C, and observed similar behavior. The relative lack of change in earlier layers appears intuitive as we expect parameter settings of earlier layers to influence how parameters of later layers are tuned. This observation also appears to justify layerwise training strategies for parameterized quantum circuits as well as suggest more carefully setting or initializing values of parameters of earlier layers.

#### V. FUTURE DIRECTIONS

PECT is a high-level optimization strategy for determining optimal parameter values for parameterized quantum circuits. Because PECT employs many subroutines, it remains suitable to leverage existing developments such as using the individual Coupled Adaptive Number of Shots (iCANs) optimizer [23] or the quantum natural gradient descent method [21] for local optimization. Several advancements made to the optimization of PQCs are bottle-necked by the number of parameters in the circuit, e.g. quantum natural gradient's dependency on the computation and inversion of the Fubini-Study metric tensor which scales with the number of parameters. Because PECT actively reduces the number of parameters in the PQC that are being optimized in each iteration, it also reduces the costs associated with these subroutines, thereby making the overall process more efficient. Another improvement may be to employ the more sophisticated layerwise training schemes (e.g. in [28]) for layerwise PECT.

As with many heuristic methods, PECT's performance is sensitive to the choice of hyperparameters, especially global sparsity of the PQC  $s$  and the target number of parameters to prune,  $N_p$ . For instance, from our simu-

<sup>5</sup> These optimization results are from estimating the ground state energy of LiH at  $R = 2.5\text{\AA}$  using the LDCA ansatz.

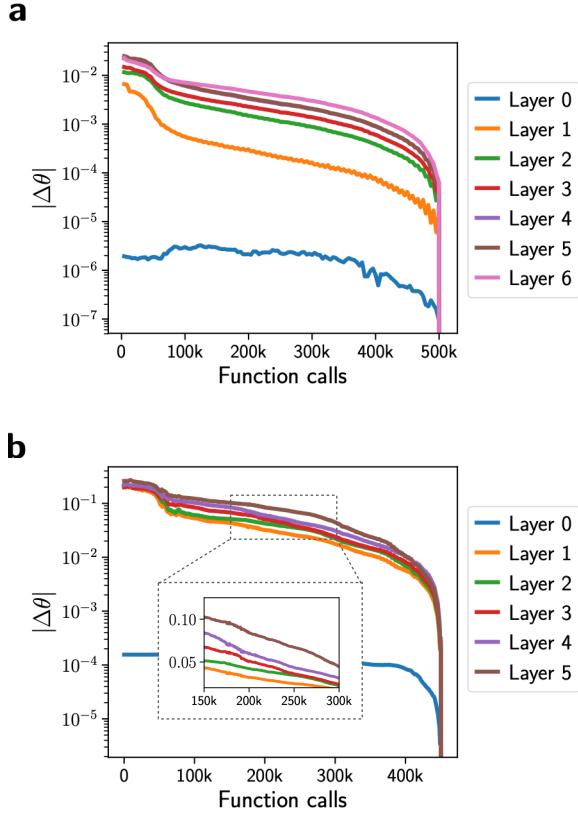


Figure 9. Plots tracking averaged parameter convergence to final values for each circuit layer. Plot in (a) shows the averaged parameter convergence per circuit layer (superlayer) for a PECT-based optimization of the LDCA modeling the LiH ground state. The averaged parameter convergence per layer is defined as in Equation 5. Plot in (b) shows the averaged parameter convergence per circuit layer for a non-PECT optimization of the LDCA modeling the LiH ground state. We generally observe that parameters of earlier layers start closer and change less with respect to their final values than parameters of later layers do.

lations, we observed that if the value of  $s$  was assigned too high, i.e. the parameter subset at each iteration was too small, this often resulted in the PECT simulation terminating due to oscillating energies. In such a case, by simply decreasing  $s$ , the performance often improved. On the other hand, if  $s$  was too low, PECT simulations would converge but expend more function calls due to the larger parameter space in each local optimization. Thus, a closer investigation into properly tuning these hyperparameters may be necessary to more effectively use PECT for variational algorithms.

Our current implementation of PECT, using local optimizers such as L-BFGS-B and SLSQP, does not make effective use of circuit batching, which may result in a high overall experimental runtime due to network latency [25]. One potential way to make PECT a better candidate for optimizing PQCs via the cloud is to employ local opti-

mizers that support evaluations of circuit batches, e.g. particle swarm optimization [60].

Lastly, significant effort has been dedicated to pruning or training sparse networks in classical machine learning since the 1990’s [61]. Just as PECT adapted the dynamic sparse reparameterization method, we expect several neural network pruning methods to be applicable and beneficial for optimizing and generating more compact quantum circuits for variational quantum algorithms.

## VI. CONCLUSION

Through random circuit sampling, quantum computers have recently been demonstrated to accomplish particular computational tasks exponentially faster than any classical computer can [59]. A natural next step is designing and demonstrating practical applications of quantum computers; variational quantum algorithms are believed to be promising candidates for such experiments. At this scale, optimization of parameterized quantum circuits will likely be a significant bottleneck among others, prompting a need for optimization strategies tailored to large circuits with many parameters. Fortunately, there have been significant efforts in designing powerful parameterized ansatze for modeling ground states of strongly correlated systems. Here, we contribute to the set of ansatz optimization techniques by introducing the “Parameter Efficient Circuit Training” (PECT) method. By optimizing only a subset of gates of a parameterized ansatz and iteratively removing and adding gates to the subset to refine the circuit composition, PECT can construct an optimized circuit that additionally saves on resources such as circuit depth, two-qubit gate count and, in some cases, the optimization runtime.

Through noiseless simulations of VQE calculations on small molecular systems, we show that PECT is a promising approach for optimizing parameterized quantum circuits that have so far been limited by their difficulty in parameter exploration and optimization. For instance, considering existing optimization strategies, PECT appears to be necessary for reliably converging calculations of molecular systems using LDCA and similar ansatze with high parameter counts and parameter redundancy. As noted earlier, in addition to facilitating convergence to lower energies, using PECT may significantly reduce circuit depths, two-qubit gate counts, and optimization runtimes. Some of these features are particularly critical for realizing useful application of variational algorithms in NISQ devices, where noise imposes stringent limits to the size of the circuits we can execute. PECT is compatible with other optimization strategies such as layer-wise optimization, and could be used in conjunction with measurement-aware optimizers in order to minimize the number of circuit repetitions used during the local optimizations. We expect that, used jointly with several of these approaches, PECT would become an effective opti-

mization strategy for refining and training parameterized quantum circuits across multiple quantum computing applications.

## ACKNOWLEDGEMENTS

S.S. is supported by the DOE Computational Science Graduate Fellowship under grant number DE-FG02-

97ER25308. All the authors would like to thank William Simon, Manuel Rudolph, Pierre-Luc Dallaire-Demers, and Peter Johnson for their valuable feedback. The optimization strategy and numerical simulations reported in this work were developed and executed through workflows managed by a pre-release version of the Orquestra platform [62], using methods from OpenFermion [63] and Psi4 [64] libraries to construct the problem Hamiltonians, Qiskit [65] and Cirq libraries to estimate circuit resource requirements, and using the Intel Quantum Simulator [66] to simulate quantum circuits.

- 
- [1] J. Preskill, *Quantum* **2**, 79 (2018).
  - [2] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, *Nat. Commun* **5**, 4213 (2014).
  - [3] E. Farhi, J. Goldstone, and S. Gutmann, “A Quantum Approximate Optimization Algorithm,” (2014), arXiv:1411.4028.
  - [4] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, *Nature* **567**, 209 (2019).
  - [5] E. Farhi and H. Neven, “Classification with Quantum Neural Networks on Near Term Processors,” (2018), arXiv:1802.06002.
  - [6] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, *Phys. Rev. A* **101**, 032308 (2020).
  - [7] J. Romero, J. P. Olson, and A. Aspuru-Guzik, *Quantum Sci. Technol* **2** (2017), 10.1088/2058-9565/aa8072.
  - [8] P.-L. Dallaire-Demers and N. Killoran, *Phys. Rev. A* **98**, 012324 (2018).
  - [9] S. Lloyd and C. Weedbrook, *Phys. Rev. Lett* **121**, 040502 (2018).
  - [10] J. Zeng, Y. Wu, J. G. Liu, L. Wang, and J. Hu, *Phys. Rev. A* **99** (2019), 10.1103/PhysRevA.99.052306.
  - [11] H. Situ, Z. He, Y. Wang, L. Li, and S. Zheng, *Inf. Sci.* **538**, 193 (2020).
  - [12] D. Zhu, N. M. Linke, M. Benedetti, K. A. Landsman, N. H. Nguyen, C. H. Alderete, A. Perdomo-Ortiz, N. Korda, A. Garfoot, C. Brecque, L. Egan, O. Perdomo, and C. Monroe, *Sci. Adv.* **5**, eaaw9918 (2019).
  - [13] J. Romero and A. Aspuru-Guzik, “Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions,” (2019), arXiv:1901.00848.
  - [14] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, *Nat. Commun* **9**, 4812 (2018).
  - [15] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, “Cost-Function-Dependent Barren Plateaus in Shallow Quantum Neural Networks,” (2020), arXiv:2001.00550.
  - [16] T. Volkoff and P. J. Coles, “Large gradients via correlation in random parameterized quantum circuits,” (2020), arXiv:2005.12200.
  - [17] M. Ostaszewski, E. Grant, and M. Benedetti, “Quantum circuit structure learning,” (2019), arXiv:1905.09692.
  - [18] M. Wilson, S. Stromwold, F. Wudarski, S. Hadfield, N. M. Tubman, and E. G. Rieffel, “Optimizing quantum heuristics with meta-learning,” (2019), arXiv:1908.03185.
  - [19] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, *Nat. Commun* **10**, 3007 (2019).
  - [20] H. L. Tang, E. Barnes, H. R. Grimsley, N. J. Mayhall, and S. E. Economou, “qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansatze on a quantum processor,” (2019), arXiv:1911.10205.
  - [21] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, *Quantum* **4**, 269 (2020).
  - [22] A. G. Rattew, S. Hu, M. Pistoia, R. Chen, and S. Wood, “A Domain-agnostic, Noise-resistant, Hardware-efficient Evolutionary Variational Quantum Eigensolver,” (2019), arXiv:1910.09694.
  - [23] J. M. Kübler, A. Arrasmith, L. Cincio, and P. J. Coles, *Quantum* **4**, 263 (2020), 1909.09083.
  - [24] D. Chivilikhin, A. Samarin, V. Ulyantsev, I. Iorsh, A. R. Oganov, and O. Kyriienko, “MoG-VQE: Multiobjective genetic variational quantum eigensolver,” (2020), arXiv:2007.04424.
  - [25] K. J. Sung, M. P. Harrigan, N. C. Rubin, Z. Jiang, R. Babbush, and J. R. McClean, “An Exploration of Practical Optimizers for Variational Quantum Algorithms on Superconducting Qubit Processors,” (2020), arXiv:2005.11011.
  - [26] D. Wecker, M. B. Hastings, and M. Troyer, *Phys. Rev. A* **92**, 042303 (2015).
  - [27] A. Garcia-Saez and J. I. Latorre, “Addressing hard classical problems with Adiabatically Assisted Variational Quantum Eigensolvers,” (2018), arXiv:1806.02287.
  - [28] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, “Layerwise learning for quantum neural networks,” (2020), arXiv:2006.14904.
  - [29] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, *Quantum* **3**, 214 (2019).
  - [30] G. Verdon, M. Broughton, J. R. McClean, K. J. Sung, R. Babbush, Z. Jiang, H. Neven, and M. Mohseni, “Learning to learn with quantum neural networks via classical neural networks,” (2019), arXiv:1907.05415.
  - [31] W. Lavrijzen, A. Tudor, J. Müller, C. Iancu, and W. de Jong, “Classical Optimizers for Noisy Intermediate-Scale Quantum Devices,” (2020), arXiv:2004.03004.
  - [32] Y. Shen, X. Zhang, S. Zhang, J. N. Zhang, M. H. Yung, and K. Kim, *Phys. Rev. A* **95** (2017), 10.1103/PhysRevA.95.020501.
  - [33] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, *Phys. Rev. X* **8**, 011044 (2018).

- [34] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K. L. Chan, and R. Babbush, *Phys. Rev. Lett.* **120** (2018), 10.1103/PhysRevLett.120.110501.
- [35] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, *Phys. Rev. A* **98**, 022322 (2018).
- [36] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, *Quantum Sci. Technol.* **4**, 045005 (2019).
- [37] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, *J. Chem. Theory Comput.* **15**, 311 (2019).
- [38] A. Choquette, A. Di Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, “Quantum-optimal-control-inspired ansatz for variational quantum algorithms,” (2020), arXiv:2008.01098.
- [39] M.-H. Yung, J. Casanova, A. Mezzacapo, J. McClean, L. Lamata, A. Aspuru-Guzik, and E. Solano, *Sci. Rep.* **4**, 3589 (2015).
- [40] H. Mostafa and X. Wang, “Parameter Efficient Training of Deep Convolutional Neural Networks by Dynamic Sparse Reparameterization,” (2019), arXiv:1902.05967.
- [41] P.-L. Dallaire-Demers, M. Stęchły, J. F. Gonthier, N. T. Bashige, J. Romero, and Y. Cao, “An application benchmark for fermionic quantum simulations,” (2020), arXiv:2003.01862.
- [42] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, *Adv. Quantum Technol.* **2**, 1900070 (2019).
- [43] S. E. Rasmussen, N. J. S. Loft, T. Bækkegaard, M. Kues, and N. T. Zinner, “Single-qubit rotations in parameterized quantum circuits,” (2020), arXiv:2005.13548.
- [44] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, *Quantum Sci. Technol.* **4**, 014008 (2018).
- [45] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, *SIAM J. Sci. Comput.* **16**, 1190 (1995).
- [46] D. Kraft, DFVLR-FB **88**, 33 (1988).
- [47] L. Li, M. Fan, M. Coram, P. Riley, and S. Leichenauer, *Phys. Rev. Res.* **2**, 023074 (2020).
- [48] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. K.-L. Chan, “Low rank representations for quantum simulation of electronic structure,” (2018), arXiv:1808.02625.
- [49] B. Cooper and P. J. Knowles, *J. Chem. Phys.* **133** (2010), 10.1063/1.3520564.
- [50] J. Ma, S. Li, and W. Li, *J. Comput. Chem.* **27**, 39 (2006).
- [51] X. Li and J. Paldus, *J. Chem. Phys.* **108**, 637 (1998).
- [52] J. Olsen, P. Jørgensen, H. Koch, A. Balkova, and R. J. Bartlett, *J. Chem. Phys.* **104**, 8007 (1996).
- [53] F. B. Brown, I. Shavitt, and R. Shepard, *Chem. Phys. Lett.* **105**, 363 (1984).
- [54] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, *J. Chem. Phys.* **152** (2020), 10.1063/1.5141835.
- [55] H. R. Grimsley, D. Claudino, S. E. Economou, E. Barnes, and N. J. Mayhall, *J. Chem. Theory Comput.* **16**, 1 (2020).
- [56] A. F. Izmaylov, M. Díaz-Tinoco, and R. A. Lang, *Phys. Chem. Chem. Phys.* **22**, 12980 (2020), 2003.07351.
- [57] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, *Adv. Neural Inf. Process. Syst.*, 153 (2007).
- [58] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Nature* **549**, 242 (2017).
- [59] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brando, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michelsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, *Nature* **574**, 505 (2019).
- [60] J. Kennedy and R. Eberhart, in *Proceedings of ICNN’95 - International Conference on Neural Networks*, Vol. 4 (IEEE, 1995) pp. 1942–1948.
- [61] C. Le, *Adv. Neural Inf. Process. Syst.* **2**, 598 (1990).
- [62] “Orquestra,” <https://www.orquestra.io/> (2018).
- [63] 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, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O’Brien, B. O’Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. P. D. Sawaya, B. Senjean, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, *Quantum Sci. Technol.* **5**, 034014 (2020).
- [64] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, *J. Chem. Theory Comput.* **13**, 3185 (2017).
- [65] H. Abraham, I. Y. Akhalwaya, G. Aleksandrowicz, T. Alexander, G. Alexandrowics, E. Arbel, A. Asfaw, C. Azaustre, AzizNgoueya, P. Barkoutsos, G. Barron, L. Bello, Y. Ben-Haim, D. Bevenius, L. S. Bishop, S. Bosch, S. Bravyi, D. Bucher, F. Cabrera, P. Calpin, L. Capelluto, J. Carballo, G. Carrascal, A. Chen, C.-F. Chen, R. Chen, J. M. Chow, C. Claus, C. Clauss, A. J. Cross, A. W. Cross, S. Cross, J. Cruz-Benito, C. Culver, A. D. Córcoles-Gonzales, S. Dague, T. E. Dandachi, M. Dartailh, DavideFr, A. R. Davila, D. Ding, J. Doi, E. Drechsler, Drew, E. Dumitrescu, K. Dumon, I. Duran, K. EL-Safty, E. Eastman, P. Eendebak, D. Egger, M. Everitt, P. M. Fernández, A. H. Ferrera, A. Frisch, A. Fuhrer, M. GEORGE, J. Gacon, Gadi, B. G. Gago, J. M. Gambetta, A. Gammanpila, L. Garcia, S. Garion, J. Gomez-Mosquera, S. de la Puente González, I. Gould, D. Greenberg, D. Grinko, W. Guan, J. A. Gannels, I. Haide, I. Hamamura, V. Havlicek, J. Hellmers, Ł. Herok, S. Hillmich, H. Horii, C. Howington, S. Hu, W. Hu, H. Imai, T. Imamichi, K. Ishizaki, R. Iten, T. Itoko, A. Javadi-Abhari, Jessica, K. Johns, T. Kachmann, N. Kanazawa, Kang-Bae, A. Karazeev, P. Kasse-

- baum, S. King, Knabberjoe, A. Kovyrshin, V. Krishnan, K. Krsulich, G. Kus, R. LaRose, R. Lambert, J. Latone, S. Lawrence, D. Liu, P. Liu, Y. Maeng, A. Malyshev, J. Marecek, M. Marques, D. Mathews, A. Matsuo, D. T. McClure, C. McGarry, D. McKay, S. Meesala, M. Mevissen, A. Mezzacapo, R. Midha, Z. Minev, N. Moll, M. D. Mooring, R. Morales, N. Moran, P. Murali, J. Müggenburg, D. Nadlinger, G. Nannicini, P. Nation, Y. Naveh, P. Neuweiler, P. Niroula, H. Norlen, L. J. O’Riordan, O. Ogunbayo, P. Ollitrault, S. Oud, D. Padilha, H. Paik, S. Perriello, A. Phan, M. Pistoia, A. Pozas-iKerstjens, V. Prutyanov, D. Puzzuoli, J. Pérez, Quintiii, R. Raymond, R. M.-C. Redondo, M. Reuter, J. Rice, D. M. Rodríguez, M. Rossmannek, M. Ryu, T. SAPV, SamFerracin, M. Sandberg, N. Sathaye, B. Schmitt, C. Schnabel, Z. Schoenfeld, T. L. Scholten, E. Schoute, I. F. Sertage, K. Setia, N. Shammah, Y. Shi, A. Silva, A. Simonetto, N. Singstock, Y. Siraichi, I. Situdikov, S. Sivarajah, M. B. Sletfjerding, J. A. Smolin, M. Soeken, I. O. Sokolov, D. Steenken, M. Stypulkoski, H. Takahashi, I. Tavernelli, C. Taylor, P. Taylour, S. Thomas, M. Tillet, M. Tod, E. de la Torre, K. Trabing, M. Treinish, TrishaPe, W. Turner, Y. Vaknin, C. R. Valcarce, F. Varchon, A. C. Vazquez, D. Vogt-Lee, C. Vuillot, J. Weaver, R. Wieczorek, J. A. Wildstrom, R. Wille, E. Winston, J. J. Woehr, S. Woerner, R. Woo, C. J. Wood, R. Wood, S. Wood, J. Wootton, D. Yerlin, R. Young, J. Yu, C. Zachow, L. Zdanski, C. Zoufal, Zoufalc, azulehner, bcamorison, brandhsn, chlorophyll zz, dime10, drholmie, elfrocampeador, faisaldebouni, fanizzamarco, gruu, kanejess, klinvill, kurarr, lerongil, ma5x, merav aharoni, ordmoj, sethmerkel, strickroman, sumitpuri, tigerjack, toural, vvilpas, willhbang, yang.luh, and yotamvkninibm, “Qiskit: An open-source framework for quantum computing,” (2019).
- [66] G. G. Guerreschi, J. Hogaboam, F. Baruffa, and N. P. D. Sawaya, *Quantum Sci. Technol.* **5**, 034007 (2020).

## APPENDIX

### Appendix A: LDCA Details

In this work, we consider three types of ansatze: (a) the unitary coupled-cluster ansatz including singles and doubles excitations (UCCSD), (b) a variant of UCC ansatz using chained generalized singles and paired doubles excitations called  $k$ -UpCCGSD [37], and (c) a simplified version of the low-depth circuit ansatz (LDCA) [36]. While gate compositions for (a) and (b) can be found in Refs. [44] and [37] respectively, we provide the gate composition for the particle number conserving construction of LDCA in Figure 10. LDCA has a compact nested structure with several levels of layers. Its outermost layer, or what we call “superlayer,” comprises  $\lceil \frac{n}{2} \rceil$  “sublayers,” where  $n$  is the number of qubits, and each “sublayer” comprises two layers of nearest neighbor two-qubit blocks. The first layer of two-qubit blocks starts with the first qubit, and the second layer starts with the second qubit. Each two-qubit block is composed of five two-qubit gate operations with three free parameters, as shown in Figure 10(b). While the original formulation of LDCA allowed five free parameters, one for each two-qubit gate, and included particle number constraints, our preliminary calculations (unreported) showed that using the particle number conserving version of LDCA led to significantly better convergence behaviors. For each two-qubit gate operation labeled  $P(\theta)$ , the gate is defined as  $\exp(-i\frac{\theta}{2}P)$ .

### Appendix B: Ansatz cost reductions for $\text{H}_2\text{O}$

In Section IV C, we provide average percent reductions in circuit resources, namely circuit depth and two-qubit gate counts, from using PECT as an optimization strategy. For simulations estimating ground state energies of water, these quantities may not be as meaningful because optimizations of several bond lengths did not converge to chemically accurate energies. To provide a more meaningful report of reductions in circuit resources, we provide tables with reduction numbers at bond lengths at which both non-PECT and PECT calculations produced final energies with errors below the chemical accuracy threshold. We show resulting percent reductions in depths and two-qubit gate counts arriving from using PECT for UCCSD in Table II, 4-UpCCGSD in Table III, and 5-UpCCGSD in Table IV.

### Appendix C: Parameter dynamics of $k$ -UpCCGSD and LDCA optimizations

In Section IV F, we discussed how parameters of a multi-layered PQC evolve throughout an optimization. For example, in the case of LDCA describing ground states of LiH, parameters of earlier layers started and generally remained closer to their final values compared to parameters of later layers. In Figure 11, we provide further numerical evidence for the  $k$ -UpCCGSD and LDCA ansatze describing the ground state of LiH at  $R = 2.5\text{\AA}$  and ground state of  $\text{H}_2\text{O}$  at  $R_{\text{O-H}} = 2.4\text{\AA}$ . In each plot, similar behavior in the average parameter movement is observed. For certain

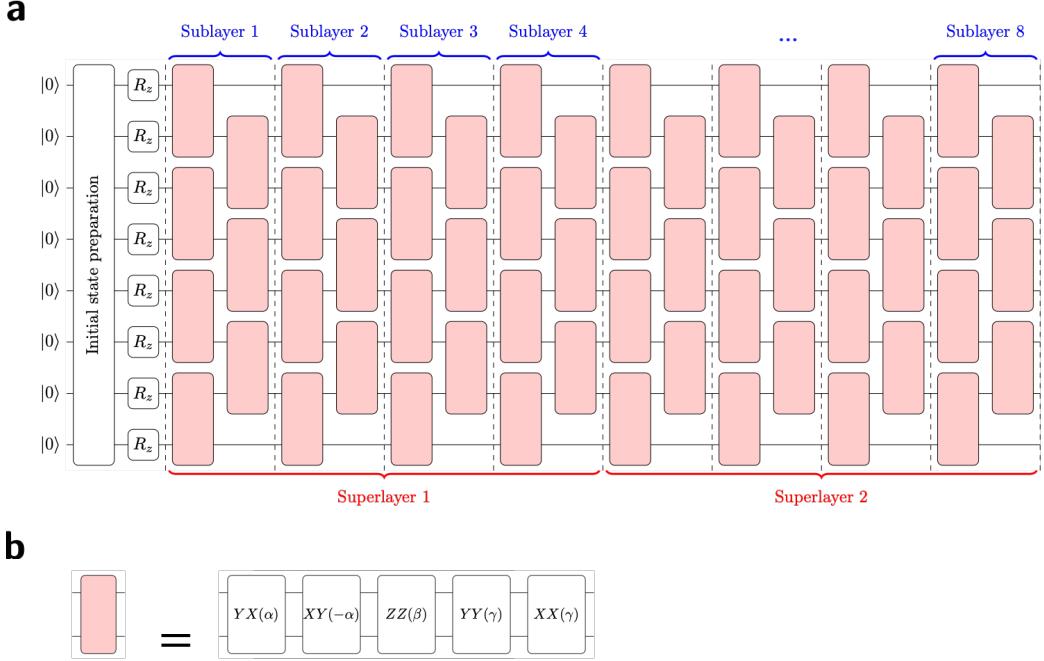


Figure 10. Gate composition of a simplified version of the low-depth circuit ansatz (LDCA) from Ref. [36]. Circuit in (a) is an instance of LDCA for 8 qubits. Each entangling operation in LDCA is composed of five two-qubit gate operations with three free parameters, as shown in (b). The  $YX$  and  $XY$  operations sharing a parameter (with opposite signs) and  $XX$  and  $YY$  operations sharing another parameter leads to an ansatz that preserves particle number, assuming the initial state is e.g. the Hartree-Fock state.

Bond length [Å]	% reduced depth	% reduced two-qubit gate count
0.8	10%	10%
1.0	10%	10%
1.2	15%	15%
1.4	9%	9%
1.6	21%	20%
1.8	9%	8%
2.6	24%	24%

Table II: Percent reductions in the circuit depths and two-qubit gate counts from utilizing PECT to optimize VQE calculations with the UCCSD ansatz at various O-H bond lengths of a water molecule. At these bond lengths, both non-PECT and PECT calculations achieved energy errors below the chemical accuracy threshold.

optimization instances such as optimizing LDCA for  $\text{H}_2\text{O}$  using PECT in Figure 11(c), we observed parameters of layer 6 changing less than parameters of layers 3, 4, and 5. This appears to suggest that after some number of circuit layers, the order in the degree of movement becomes less structured. Nevertheless, plots in Figure 11 support the idea that parameters of especially the first few layers should be carefully initialized to facilitate the tuning of later circuit layers.

#### Appendix D: Estimating the potential energy surface for $\text{H}_2\text{O}$

We plot the VQE-optimized energies for water in Figure 12. Since potential energy surfaces are often used as an input for small scale quantum dynamics simulations, the uniformity of the energy errors across different bond lengths (i.e. smoothness) is an important comparison metric for electronic structure methods. We observe that using UCCSD

Bond length [Å]	% reduced depth	% reduced two-qubit gate count
0.8	15%	32%
1.2	17%	29%
1.4	21%	33%
1.6	18%	37%
2.0	13%	25%

Table III: Percent reductions in the circuit depths and two-qubit gate counts from utilizing PECT to optimize VQE calculations with the 4-UpCCGSD ansatz at various O-H bond lengths of a water molecule. At these bond lengths, both non-PECT and PECT calculations achieved energy errors below the chemical accuracy threshold.

Bond length [Å]	% reduced depth	% reduced two-qubit gate count
0.8	17%	30%
1.2	16%	36%
1.4	18%	30%
1.6	15%	27%
2.0	19%	35%

Table IV: Percent reductions in the circuit depths and two-qubit gate counts from utilizing PECT to optimize VQE calculations with the 5-UpCCGSD ansatz at various O-H bond lengths of a water molecule. At these bond lengths, both non-PECT and PECT calculations achieved energy errors below the chemical accuracy threshold.

with PECT, the resulting potential energy surface appears smoother than the one generated using UCCSD without PECT, which has a bump in the surface due to the high energy error at  $R_{\text{O-H}} = 2.4\text{\AA}$ . With  $k$ -UpCCGSD ansatze, the potential energy surfaces are less smooth due to inconsistent improvements in energies at particular bond lengths. For smoother surfaces, we may need alternative methods for initializing  $k$ -UpCCGSD parameters. With LDCA, we note that its PECT calculation yielded a smooth and accurate potential energy surface compared to resulting surfaces of other methods.

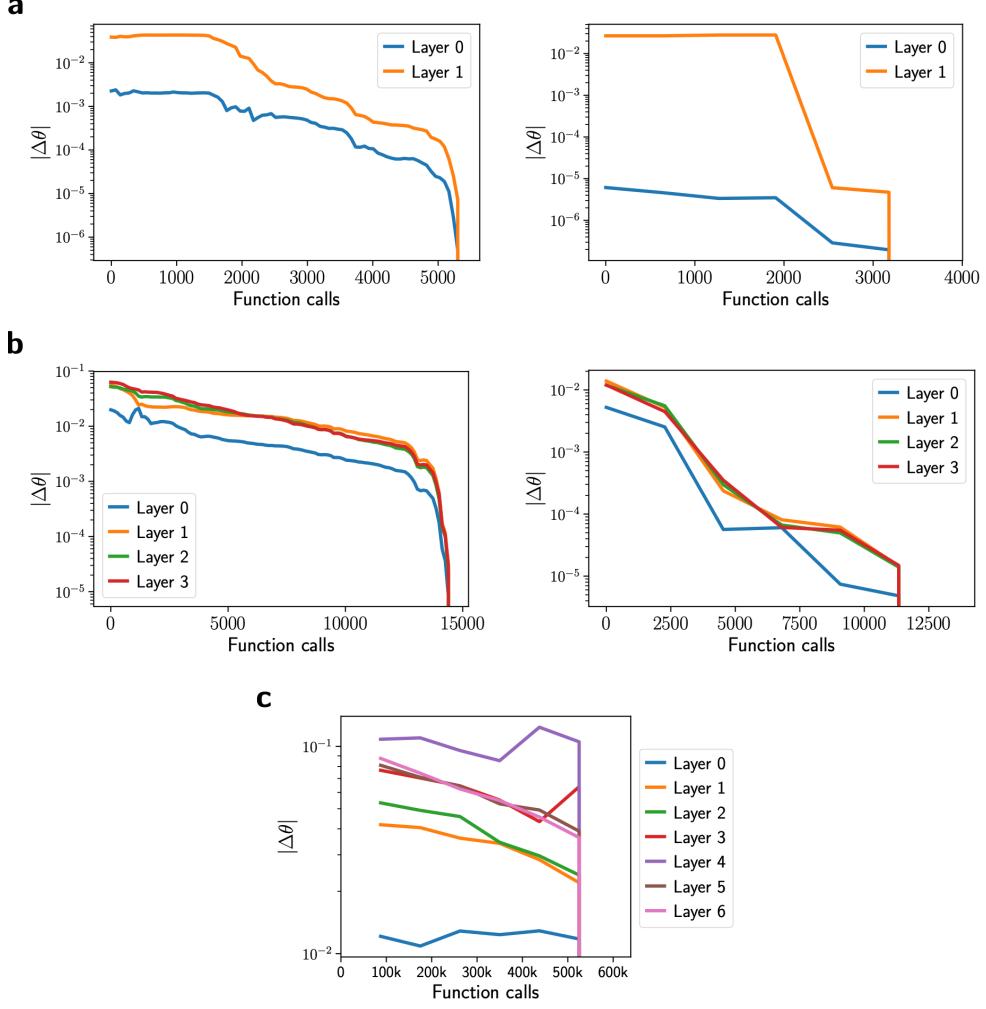


Figure 11. Absolute difference of parameter values with respect to their final values averaged over each circuit layer at each optimization iteration for  $k$ -UpCCGSD and LDCA circuits. We compute the averaged parameter distance for: (a) 2-UpCCGSD circuits describing ground state of LiH without PECT (left) and with PECT (right), (b) 4-UpCCGSD circuits describing ground state of H<sub>2</sub>O without PECT (left) and with PECT (right), and (c) LDCA circuit for describing ground state of H<sub>2</sub>O with PECT.

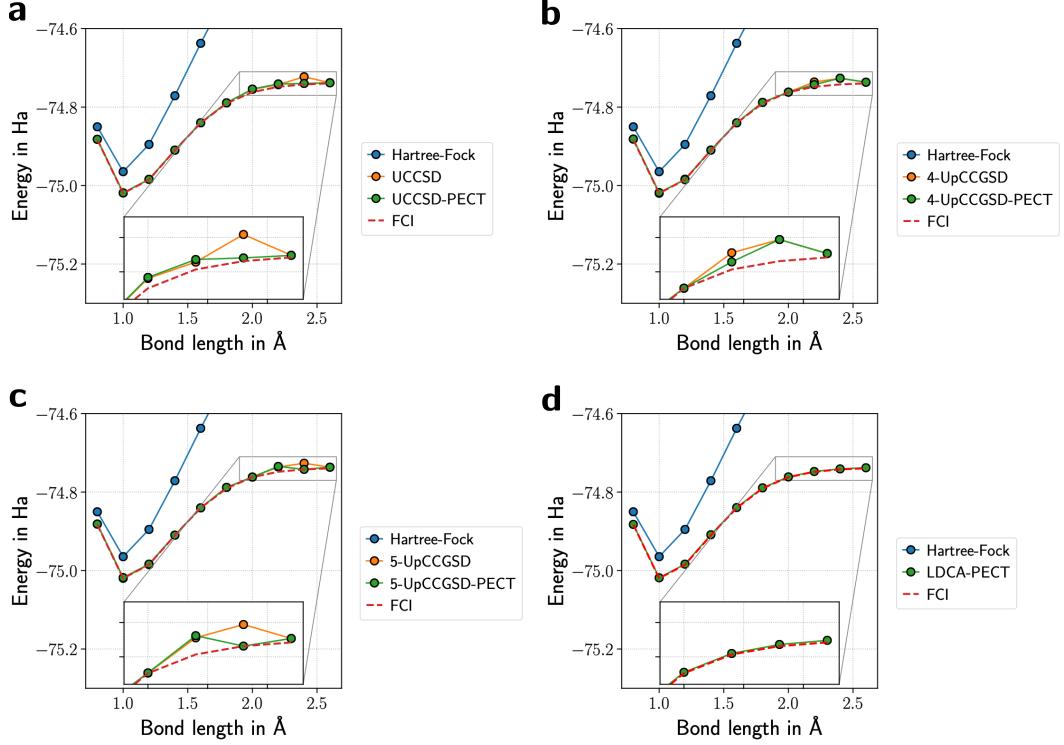


Figure 12. Potential energy surfaces of  $\text{H}_2\text{O}$  constructed using optimized (a) UCCSD, (b) 4-UpCCGSD, (c) 5-UpCCGSD, and (d) LDCA circuits. We show the energies from optimizations without PECT in orange markers and ones from optimizations with PECT in green markers. Inset in each plot zooms in on the same region from  $R_{\text{O-H}} = 1.9$  to  $2.65\text{\AA}$  and energies from -74.77 to -74.71 Hartrees.