Variational quantum algorithms (VQAs) are expected to establish valuable applications on devices in important fields such as quantum computers and quantum devices, VQAs adopt parameterized quantum circuit (PQC) (also known as ansatzes) [41] and utilize classical computers to optimize the parameters to minimize the cost functions that are designed for solving target problems.

Among the numerous VQAs, the variational quantum eigensolver (VQE) for ground state estimation is a central one of both practical and theoretical interests. For VQE, one common approach is to utilize fixed structure ansatzes, such as the hardware-efficient ansatz [42] and the unitary coupled cluster ansatz [8], [43], [44], which require a large depth to achieve high accuracy when the problem scale is large. On the other hand, there are adaptive structure ansatzes [45], [46], [47], [48], [49], which usually suffer from relatively high costs of both quantum and classical resources.

Despite the success of VQE, expressibility [50], [51], [52] and trainability [53], [54], [55] are still two critical challenges of designing ansatzes for it. The expressibility of a

**Mitigating Barren Plateaus of Variational Quantum Eigensolvers**

**XIA LIU**, **GENG LIU**, **HAO-KAI ZHANG**, **JIAXIN HUANG**, **AND XIN WANG** (Senior Member, IEEE)

1. Thrust of Artificial Intelligence, Information Hub, Hong Kong University of Science and Technology (Guangzhou), Guangzhou 510530, China
2. Key Laboratory of Cyberspace Security Defense, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China
3. Institute for Quantum Computing, Baidu Research, Beijing 100193, China
4. Institute for Advanced Study, Tsinghua University, Beijing 100084, China

Corresponding author: Xin Wang (e-mail: felixxinwang@hkust-gz.edu.cn).

This work was supported in part by the Hong Kong University of Science and Technology (Guangzhou) through Start-up Fund under Grant G010/1000151 in part by Guangdong Provincial Quantum Science Strategic Initiative under Grant GDQC2303007, in part by the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things under Grant 2023B1212010007, in part by the Quantum Science Center of Guangdong–Hong Kong–Macao Greater Bay Area, and in part by the Education Bureau of Guangzhou Municipality. (Xia Liu, Geng Liu, and Hao-Kai Zhang contributed equally to this work.)

**ABSTRACT** Variational quantum algorithms (VQAs) are expected to establish valuable applications on near-term quantum computers. However, recent works have pointed out that the performance of VQAs greatly relies on the expressibility of the ansatzes and is seriously limited by optimization issues, such as barren plateaus (i.e., vanishing gradients). This article proposes the state-efficient ansatz (SEA) for accurate ground state preparation with improved trainability. We show that the SEA can generate an arbitrary pure state with much fewer parameters than a universal ansatz, making it efficient for tasks like ground state estimation. Then, we prove that barren plateaus can be efficiently mitigated by the SEA and the trainability can be further improved most quadratically by flexibly adjusting the entangling capability of the SEA. Finally, we investigate a plethora of examples in ground state estimation where we obtain significant improvements in the magnitude of the cost gradient and the convergence speed.

**INDEX TERMS** Barren plateaus, near-term quantum computing, parameterized quantum circuits (PQCs), quantum algorithms, quantum simulation, variational quantum eigensolvers (VQEs).

**I. INTRODUCTION** Quantum computers are expected to achieve quantum advantages [1] in solving valuable problems [2], [3], [4], [5]. Before arriving at universal quantum computing, a key direction is to explore the power of noisy intermediate-scale quantum (NISQ) [6] devices in important fields such as quantum chemistry [7], [8] and quantum machine learning [9], [10], [11], [12], [13]. Recent results [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24] have shown that it is promising to realize quantum advantages with such devices in specific tasks. One common paradigm for designing quantum solutions using NISQ devices is variational quantum algorithms (VQAs) [25], [26], [27]. VQAs are promising to deliver applications in many important topics, including ground state preparation [7], [8], [28], [29], [30], [31], quantum data compression [32], [33], machine learning [9], [19], [24], [34], [35], [36], [37], [38], and combinatorial optimization [39], [40]. Combining the advantages of classical
PQC refers to how many quantum states it can represent. A circuit with high expressibility can represent a broad class of quantum states, and thus it has a large opportunity to provide a good approximation to the target state, which is a necessary condition for solving the problem successfully. The trainability of a parameterized circuit refers to the extent to which a common optimization algorithm can run successfully, e.g., the average gradient magnitude. A circuit with high trainability can lead to a higher probability of finding the optimum of the objective function. To improve the accuracy of the algorithms, deep ansatzes are usually preferred for their strong expressibility. This could be seen in the same spirit of deep neural networks in machine learning. However, recent works [52], [56], [57] imply that strong expressibility will lead to poor trainability due to the barren plateau phenomenon [53], [58], [59], i.e., exponentially vanishing gradients in training with respect to the number of qubits, which results in the optimization process becoming stuck in local minima or failing to converge to the correct ground state energy altogether. In other words, it is difficult to strike a balance between the expressibility and the trainability of ansatzes to efficiently complete the task. This technical bottleneck seriously restricts the scalability of VQE. Despite a plethora of recent attempts to address the barren plateaus through adjusting initialization [60], [61], [62], cost functions [63], [64], and architectures [65], a recent work [66] pointed out that the most existing methods for avoiding barren plateaus are essentially classically simulable. Thus, there is still a strong need for extending the scalability of VQE by improving trainability while preserving effectiveness and being nonclassically simulable.

To overcome these challenges, we propose the state efficient ansatz (SEA), as shown in Fig. 1. By removing the redundancy between the sets of universal unitary and universal pure quantum states, SEA can represent an arbitrary pure quantum state with fewer parameters, resulting in its efficiency in all the tasks that are essentially learning a quantum state, which we call the state-oriented tasks (such as VQE). Moreover, by adjusting the number of CNOT gates in the entangling layer (see Fig. 2), SEA has the ability to learn a target ground state with low bipartite entanglement, where we proved that barren plateaus can be effectively mitigated by the SEA. To be specific, the SEA can effectively enhance the gradient magnitude up to a square root of the scaling of the global 2-design case, which means that the entire randomly initialized circuit forms a 2-design ensemble. We further investigate a plethora of examples in ground
state estimation through numerical experiments and establish evident improvements in both the overall behavior and the magnitude of the cost gradient.

II. STATE EFFICIENT ANSATZ

The underlying concept of SEA is to prune the redundant expressibility of an ansatz in the state-oriented tasks by utilizing the Schmidt representation [29], [31], [67], [68] of pure states. This approach is also inspired by the tree tensor network [69], [70], [71], [72]. Specifically, the input system consists of two subsystems, A and B, and the ansatz is composed of three parts. The first part is the Schmidt coefficient layer $U_1$ acting on subsystem A. The second part is the entangling layer $V$ acting on the entire system $AB$ to create entanglement between two subsystems. The last part is the local basis changing layer (LBC layer), which consists of two local circuits, $U_2$ and $U_3$, applied to the two subsystems, respectively. The whole structure of SEA is depicted in the bottom left corner of Fig. 1 and can be expressed as

$$S(\theta) \equiv (U_2(\theta_2) \otimes U_3(\theta_3))V(U_1(\theta_1) \otimes I)$$

where $S$ is the unitary representation of SEA, $\theta = \{\theta_1, \theta_2, \theta_3\}$, and each $\theta_i (i = 1, 2, 3)$ is a parameter vector. The whole structure is based on Schmidt decomposition and we can tune parameters to get the desired orthonormal basis in each subsystem so that we can arrive at the target state at last. The entangling layer $V$ in SEA serves as a block that connects these two subsystems and generates entanglement between them. In order to showcase how SEA could generate arbitrary pure states with the help of Schmidt decomposition, we simply take the CNOT gates as the component of the entangling layer in our article, which can evolve $\sum_{k=0}^{2^N-1} \lambda_k |k\rangle_A \otimes |v_k\rangle_B$ into $\sum_{k=0}^{2^N-1} \lambda_k |k\rangle_A \otimes |v_k\rangle_B$, where $|\{k\}\rangle_{k=0}^{2^N-1}$ represents the computational basis. Although we used CNOTs as the entangling layer in this article for demonstration, it is not necessarily composed of CNOT gates. Other equivalent structures that can generate entanglement between two subsystems $A$ and $B$ are also available in the entangling layer.

We commence by elucidating how SEA reduces redundant expressibility from the standpoint of degrees of freedom (DOF). For state-oriented tasks, the DOF of the 2 $N$-qubit pure state set is $2^{2N+1} - 2$, whereas the DOF of the unitary group $U(2^N)$ equals $2^N$ [73]. Consequently, employing a universal 2 $N$-qubit PQC capable of representing a universal unitary set to generate a universal pure state set is overkill. In contrast, SEA boasts a quadratic advantage in the parametric DOF compared to universal PQCs. It is noteworthy that a universal $N$-qubit PQC $U_i$ possesses $O(4^N)$ DOF [74]. As a result, a 2 $N$-qubit SEA with universal $U_i (i = 1, 2, 3)$ also has $O(4^N)$ parametric DOF. By comparison, a 2 $N$-qubit universal PQC has $O(4^2N)$ parametric DOF. Thus, SEA necessitates quadratically fewer parameters than a general PQC of the equivalent dimension while preserving the capacity to generate an arbitrary pure state. It is also worth highlighting that SEA can attain the optimal DOF for generating pure states set, which is the $2 \times 4^N - 2$, by eliminating the redundant freedom in $U_1$, $U_2$, and $U_3$. In this sense, we refer to it as the SEA.

Having established that SEA is efficient in generating an arbitrary pure state, we now elaborate on the effectiveness of SEA. For simplicity, we assume both subsystems $A$ and $B$ have $N$ qubits each (actually $A$ and $B$ do not be equal), and let $|0\rangle \otimes 2^N$ (or other tensor product states) be an initial state fed into SEA. Suppose the target state is $|\phi\rangle$, it can be expressed as $|\phi\rangle = \sum_{k=0}^{2^N-1} \lambda_k |v_k\rangle_A \otimes |v_k\rangle_B$ according to the Schmidt decomposition. An example of SEA is depicted in Fig. 2. To be specific, $U_1$ serves as the Schmidt coefficient layer, with $U_1|0\rangle \otimes N = \sum_{k=0}^{2^N-1} \lambda_k |k\rangle_A$, where $|\{k\}\rangle_{k=0}^{2^N-1}$ represents the computational basis. In the entangling layer, we set a composition of $\text{CNOT}$ controlled and targeted on the qubit-pairs $(i, N+i)_{i=1}^{N-1}$. In the last layer, we apply $U_2$ and $U_3$ to subsystems $A$ and $B$, respectively, such that $U_2|k\rangle_A = |v_k\rangle_A$ and $U_3|k\rangle_B = |v_k\rangle_B$. Consequently, after SEA, $|0\rangle \otimes 2^N$ will evolve into $\sum_{k=0}^{2^N-1} \lambda_k |v_k\rangle_A \otimes |v_k\rangle_B$, which is the desired target state. It is worth noting that the evolution process exactly forms the Schmidt decomposition of pure states with tunable Schmidt coefficients and bases. Since any pure state has a form of Schmidt decomposition, SEA can evolve an initial state $|0\rangle \otimes 2^N$ into an arbitrary 2 $N$-qubit pure state if $U_1$ can generate an arbitrary $N$-qubit pure state (that is, $U_1$ possessing universal wavefunction expressibility) and $U_2$, $U_3$ are universal. The entire process of SEA acting on the initial state $|0\rangle \otimes 2^N$ is illustrated in Fig. 2. The analysis above implies that SEA, with a $U_1$ of the universal wavefunction expressibility and $U_2$, $U_3$ being universal, can accurately solve state-oriented tasks. It is important to note that an $N$-qubit SEA possesses universal wavefunction expressibility under the same conditions, allowing it to serve as the Schmidt coefficient layer of a 2 $N$-qubit SEA to generate an arbitrary 2 $N$-qubit pure state.

Owing to SEA’s capability to evolve $|0\rangle \otimes 2^N$ into an arbitrary 2 $N$-qubit pure state, we can infer that SEA is effective for the VQE, which aims to variationally find the ground state of a given Hamiltonian. To be specific, for any 2 $N$-qubit Hamiltonian $H$, the objective of the VQE is to find the ground state energy $E_0 = \min_{\psi} \langle \psi | H | \psi \rangle$, where the minimization is over the 2 $N$-qubit pure state set $|\{\psi\}\rangle$. Since a 2 $N$-qubit SEA could represent the zone of an arbitrary 2 $N$-qubit pure state, which means any 2 $N$-qubit pure state can be obtained by tuning the parameters of SEA. Thus, SEA will obtain $E_0$ of a Hamiltonian after the optimization. This exactly illustrates the effectiveness of SEA when employed in VQE. More details can be found in Section B of the Appendix. We note that the likelihood of SEA generating a ground state depends on the optimization task and the experimental settings, such as the optimizer used.

When dealing with a specific task, we can optimize SEA further if we have certain information about the entanglement of the target quantum state. The previously mentioned SEA
can generate a pure state with a full Schmidt rank. However, if the target quantum state is weakly entangled, we can modify the structure of the entangling layer to reduce the cost without sacrificing performance. The effectiveness of SEA in this scenario is described in the following proposition.

**Proposition 1:** If \( U_1 \) can generate any \( N \)-qubit pure state that is a superposition of at most \( K \) computational basis states, then for any \( |\phi\rangle \), there exists an \( N \)-qubit output state \( |\psi\rangle \) with \( F(|\phi\rangle, |\psi\rangle) \geq \min(\frac{1}{2}, 1) \), where \( F(|\phi\rangle, |\psi\rangle) \) is the fidelity between \( |\phi\rangle \) and \( |\psi\rangle \), and \( r \) is the Schmidt rank of \( |\phi\rangle \).

The main idea behind proving Proposition 1 involves using the assumption of \( U_1 \) to generate a quantum state with a Schmidt number of \( \min(K, r) \), and then leveraging the universality of \( U_2 \) and \( U_3 \), which means \( U_2 \) and \( U_3 \) can represent the space of \( n \)-qubit quantum states, to convert the basis of the two subsystems, respectively. Consequently, the fidelity between \( \phi \) and the output state can achieve at least \( \min(\frac{1}{2}, 1) \). The detailed proof can be found in Section C of the Appendix.

Note that the \( K \) is a theoretical concept that means the number of Schmidt numbers greater than some \( \epsilon \), where \( \epsilon \in [0, 1] \) is artificially fixed. After the Schmidt decomposition of a quantum state, \( K \) can be used to remove the less entangled components. In the experiment, we can tune the value \( K \), which corresponds to the entangling capability, by adjusting the number of \( \text{CNOT} \) gates in the entangling layer. This proposition guarantees that SEA with fewer \( \text{CNOT} \) gates maintains robust expressibility for learning weakly entangled quantum states.

Up until now, our analysis of SEA’s effectiveness has been based on the assumption of perfect training. However, in practice, trainability is also a crucial factor to be considered. Many common ansatzes suffer from the notorious barren plateau phenomenon [53]. It has been demonstrated that the cost gradient vanishes exponentially with the number of qubits for a randomly initialized PQC with adequate depth if the unitary ensemble generated by the PQC is sufficiently random to accord with the Haar distribution over the unitary group up to the second moment, i.e., forming a unitary 2-design. The subsequent proposition suggests that SEA does not form a unitary 2-design, even if it possessed sufficient expressibility for state-oriented tasks. In other words, SEA could mitigate barren plateaus without losing effectiveness on VQE. We summarize this conclusion as Proposition 2.

**Proposition 2:** SEA with \( U_i(i = 1, 2, 3) \) being local 2-design and \( \text{CNOT} \) as entangling layer does not form a 2-design on the global system.

The main idea of the proof is directly calculating the value of \( \langle 0^{\otimes 2N} | \int_U U^\dagger C U \rho U^\dagger DU dU | 0^{\otimes 2N} \rangle \) with \( C = D = \rho = \langle 0^{\otimes 2N} U^{\dagger} DU dU | 0^{\otimes 2N} \rangle \), where \( U \) denotes the unitary ensemble generated by SEA. Then, if a \( 2N \)-qubit SEA forms a 2-design, it should hold that \( \langle 0^{\otimes 2N} | \int_U U^\dagger C U \rho U^\dagger DU dU | 0^{\otimes 2N} \rangle \) is \( \frac{2}{d^2(d^2+1)} \) due to the closed form of integral formula shown in [75]. Based on the result in [75], we further propose its variant form for the problem with the bipartite systems in the proof of Proposition 2. Thus, a key proof ingredient that explores the central structure of SEA, we find that the actual value of the target term is \( \frac{2d+6}{d(d^2+1)} \). This leads to a direct contradiction and the fact that SEA with \( U_i(i = 1, 2, 3) \) being local 2-design structures and \( \text{CNOT} \) entangling layer does not form an exact 2-design from the global perspective. Even if we change the number of \( \text{CNOTs} \) in the entangling layer, we can still similarly prove that SEA will not form a 2-design. Detailed proofs of Proposition 2 can be found in Section D of the Appendix.

As we have known that a large class of random PQCs will end up forming a 2-design when the depth is large [53] and this is the usual case in VQAs, Proposition 2 implies that even though the local structures of SEA have strong expressibility by forming 2-designs, SEA will still not form a global 2-design. Thus, SEA could mitigate the known definite zone of barren plateaus without losing effectiveness on VQE. Moreover, SEA will maintain this property even if we change the number of \( \text{CNOTs} \) in the entangling layer. In fact, with less expressibility, SEA will be farther from being a 2-design.

However, not being a unitary 2-design does not guarantee good trainability. In the following, we conduct a more detailed analysis of the trainability of the SEA in terms of the gradient variance.

### III. MITIGATING BARREN PLATEAUS

In this section, we evaluate the trainability and expressibility of the SEA by calculating the variance of the cost gradient [53], [63] and the frame potential [56], respectively, given that the subblocks in the SEA are sampled from local unitary 2-design ensembles. We demonstrate that by tuning the number of \( \text{CNOT} \) gates in the entangling layer, SEA can continuously alter the entanglement of the prepared state, thereby striking a flexible balance between trainability and expressibility [52]. Especially, if the target state is weakly entangled, the SEA can attain a square root advantage over the common circuit ansatzes, such as the hardware-efficient ansatz.

Trainability is one of the most critical concerns in VQAs. If the adopted ansatz forms a unitary 2-design, the optimization process will be crippled due to the exponentially vanishing gradients, resulting in a severe issue with the trainability of VQAs. Specifically, we consider the VQE cost function \( C(\theta) \) with an initial state \( |0^{\otimes 2N}\rangle \), an objective operator \( H \), and a circuit ansatz \( U(\theta) \) on \( 2N \) qubits. We analyze the gradient component \( \partial_\mu C = \frac{\partial C}{\partial \mu} \) concerning the variational parameter \( \mu \). Here, we assume that all subblocks with variational parameters in SEA, i.e., \( U_1, U_2, \) and \( U_3 \), are sampled from local unitary 2-designs, allowing us to integrate them using the Weingarten calculus [76], [77]. If \( \partial_\mu \) is located at \( U_i \), we assume that the two parts in \( U_i \), split by the gate
corresponding to $\theta_A$ (as in Fig. 7), are both local 2-designs. The trainability of SEA under this assumption can be improved by reducing the number of cnot gates in the entangling layer. Hence, SEA can effectively mitigate barren plateaus for low-entangled quantum states while maintaining its expressibility. Proposition 3 below presents an explicit relationship between the trainability of SEA and the number of cnot gates in the entangling layer.

**Proposition 3 (Trainability of SEA):** For an SEA defined on $2N$ qubits with all subblocks being local 2-designs, the variance of the cost gradient scales of the number of qubits as

$$\text{Var}_{\text{SEA}}[\partial_{\mu}C] \in \mathcal{O}(2^{-(N+M)})$$

where $C$ represents the cost function, $\partial_{\mu}C = \frac{\partial C}{\partial \mu}$, Var$_{\text{SEA}}[*]$ is the variance of $*$, $M \in \{0, 1, \ldots, N\}$ denotes the number of cnot gates used in the entangling layer and the variance is taken over all SEA subblocks independently.

**Proof:** Suppose $U = (U_2 \otimes U_3)\text{CNOT}(U_1 \otimes I_B)$ is an SEA on a equally bipartite system $AB$ of dimension $d = d_Ad_B$ and $d_A = d_B = 2^N$. Note that the cnot in $U$ represents a collection of $M$ cnot gates between subsystem $A$ and $B$. We focus on a single parameter $\theta$ within $U_1$ such that $U_1 = U_5(e^{-i\Omega})U_4$. The case where $\theta$ locates at $U_2$ or $U_3$ can be analyzed similarly. Denote $W = (U_2 \otimes U_3)\text{CNOT}(U_5 \otimes I_B)$, $V = U_4 \otimes I_B$ and then we have $U = W(e^{-i\Omega}) \otimes I_B)V$. The cost gradient becomes

$$\frac{\partial C}{\partial \theta} = i\text{Tr}\left(V\rho V^\dagger \left[\Omega \otimes I_B, \hat{W}^\dagger H \hat{W}\right]\right)$$

where we denote $\hat{W} = W(e^{-i\Omega}) \otimes I_B$, $\hat{S}_5 = U_5(e^{-i\Omega})$ and $\rho = |0\rangle\langle 0| \otimes 2^N$. Suppose $H$ is traceless without loss of generality and all $U_i$ are sampled from local 2-designs. Since the tensor product of two 1-designs is still a 1-design, the expectation of the cost gradient is zero. Then, consider the variance

$$\text{Var}_{\text{SEA}}[\partial_{\mu}C] = -E\left[\text{Tr}\left(V\rho V^\dagger \left[\Omega \otimes I_B, \hat{W}^\dagger H \hat{W}\right]\right)\right]^2$$

where the integration is taken over $U_2$, $U_3$, $U_4$, and $U_5$ independently. We exploit the RTNI package [78] to calculate this integral. It turns out that the exact expression of the variance is dominated by

$$\text{Var}_{\text{SEA}}[\partial_{\mu}C] \rightarrow \frac{\text{Tr}((\text{Tr}_A H)^2(\text{Tr}_B H)^2))}{d_A(d_B^2-1)^2(d_B^2-1)}$$

According to the properties of the copy and xor tensors [79], a single 2-qubit CNOT gate in the SEA contributes one diagrammatic loop in (5) while a 2-qubit identity contributes two loops, where each loop contributes a factor 2. Combining the contributions from Tr$((\text{Tr}_A H)^2) \sim d_A^2 d_B$ and Tr$((\text{Tr}_B H)^2) \sim d_A^2 d_B$ and Tr$(\Omega^2) \sim d_A$, the dominant term shall scale with the number of qubits $2N$ as $O(2^{-(N+M)})$. Note that, here we suppose the Hamiltonian $H$ contains terms acting on $A$ or $B$ trivially, which is reasonable for physical and chemical models with local interactions.

The variance of the cost gradient scales of a global 2-design ansatz (such as hardware-efficient ansatz) is $O(2^{-2N})$, where $2N$ is the number of qubits. However, Proposition 3 states that the variance of the cost gradient scales of the SEA is $O(2^{-N+M})$, where $M$ is the number of cnot gates. This means that the degree of entanglement of the quantum states also affects the gradient scales of the SEA. The result notably implies that, by tuning the number of cnot gates in the entangling layer, the SEA can effectively enhance the gradient magnitude up to a square root of the scaling of the global 2-design case $\text{Var}_{\text{Haar}}[\partial_{\mu}C] \in O(2^{-2N})$.

**FIGURE 3.** Second frame potentials of 8-qubit SEA with different Schmidt coefficient layers and entangling layers. Yellow triangles, black dots, and red diamonds represent the frame potentials of the 8-qubit SEA when changing the number of qubits associated with the Schmidt coefficient layer, entangling layer, and both two layers, respectively. The blue dotted line represents the frame potential of an exactly 2-design circuit. Each value is estimated using 5000 samples of unitaries.

**FIGURE 4.** Example of a 4-qubit ALT with two layers. Each layer is composed of the rotation operator gate $R_y$ and control-Z gate.
FIGURE 5. Numerical simulations of VQE on $N_2$ (14-qubit) and Heisenberg model (14-qubit). (a) Numerical simulations of VQE on $N_2$. The molecular geometry, specified in Cartesian format with the unit in Angstrom, is provided as detailed coordinates: "N [0, 0, 0]" and "N [0, 0, 1.09]." We utilized the "sto-3 g" basis set, and the y-axis unit is Hartree. The blue dotted line is the theoretical ground energy of $N_2$, and the lines from top to bottom represent the experimental results of ALT, the random circuit, SEA of Schmidt coefficient layer as subSEA, SEA with 2 CNOTs, and SEA with 3 CNOTs, respectively. $j(j=2, 3)$ CNOTs means that we set a composition of $j$ CNOTs controlled and targeted on the qubit-pairs $(i, N+i)$ for $j-1$ times. (b) Numerical simulations of VQE on the Heisenberg model. The Hamiltonian is chosen to be the 14-qubit 1-D spin-1/2 antiferromagnetic Heisenberg model with periodical boundary condition, i.e., $H = \sum_{i=1}^{14}(X_iX_{i+1} + Y_iY_{i+1} + Z_iZ_{i+1})$. The blue dotted line is the theoretical ground energy of the Hamiltonian $H$, and the other lines from top to bottom represent the results of the ALT, random circuit, SEA of Schmidt coefficient layer as subSEA, SEA with 3 CNOTs and SEA with 2 CNOTs, respectively.

FIGURE 6. (a) Comparison of the scaling of the average variance of gradients between different ansatzes on the Heisenberg model. This shows the semi-log plot of the average value of the variance of each parameter in the circuit. The depth of each part of SEA is 30 and we ensure different ansatzes have similar numbers of parameters by setting different depths. The solid part of the fitted lines represents the range we experiment with, while the dashed part represents the expected performance on a larger range. (b) Comparison of the scaling of variance between the general structure of SEA and general random circuit. The sea-Schmidt and the sea$_{\lfloor N/2\rfloor}$-Schmidt show the semi-log plot of the variance of the gradient with respect to the parameter in the Schmidt coefficient layer. The sea$_1$-LBC and the sea$_{\lfloor N/2\rfloor}$-LBC show the semi-log plot of the variance of the gradient with respect to the parameter in the LBC layer. The markers on the fitted lines represent the results of our experiments. The dotted part of the lines represents the expected performance on a larger range.

FIGURE 7. Structures of SEA used for studying the scaling of variance of general circuits. (a) $R_x$ gate located in the Schmidt coefficient layer. (b) $R_x$ gate located in the LBC layer. Parameter $\theta \in [0, 2\pi]$ is uniformly chosen and $U_1, U_11, U_{12}, U_2, U_{21}, U_{22}$, and $U_3$ are Haar randomly chosen from the unitary group in each round of sampling. CNOT-s is composed of a number of CNOT-s.
as numerically verified in Fig. 6. This is achieved by sacrificing the ability to express the highly entangled states since the number of \(\text{CNOT} \) gates determines the upper bound to the Schmidt rank of the prepared state. However, if the target state is known to be low-entangled, e.g., the Schmidt rank does not scale or scales slowly with the number of qubits, the SEA could effectively mitigate barren plateaus and at the same time keep the capability of expressing the target state. For the VQE, which focuses on the task of ground state preparation, this assumption of low-entanglement is reasonable according to the entanglement area law for the ground states of gapped systems and the logarithmic growth for the ground states of gapless systems, at least in the 1-D case [80]. Although SEA currently focuses on entanglement over a particular bipartition, it is expected to serve as a building block of the ansatz design and corresponding analysis based on more complicated entanglement structures.

The tradeoff between the trainability and the expressibility [52] motivates us to further quantitatively analyze the expressibility of SEA using the \(t\)-degree frame potential [56], which is defined by

\[
\mathcal{F}^{(t)} = \mathbb{E} \left[ | \langle 0 | U^t V | 0 \rangle |^{2t} \right]
\]

where \(| 0 \rangle\) denotes the zero state of the whole system and the expectation with respect to \(U, V\) is taken over two copies of the ansatz ensemble independently. The \(t\)-degree frame potential measures the expressibility of the ansatz ensemble up to the \(t\)th moment. It is known that the Haar random ensemble achieves the minimum frame potential \(\mathcal{F}^{(t)}_{\text{Haar}}\) for each \(t\), and this value can be achieved if and only if the ansatz ensemble is a \(t\)-design ensemble [81]. In addition, a smaller frame potential implies a stronger expressibility. Under the assumption of local 2-designs, we have the following proposition for the expressibility of the SEA.

**Proposition 4 (Expressibility of SEA):** For an SEA defined on \(2N\) qubits with all subblocks being local 2-designs, the first and second frame potential satisfies

\[
\mathcal{F}_{\text{SEA}}^{(1)} = 2^{-2N}, \quad \mathcal{F}_{\text{SEA}}^{(2)} \in \mathcal{O}(2^{-4N}).
\]

**Proof:** Since the SEA with local 2-designs forms an exact 1-design, the first frame potential of the SEA with local 2-designs is optimal, i.e., equal to that of a global 1-design \(2^{-2N}\). The calculation of the second frame potential is more complicated. Similar to the proof of Proposition 3, we again exploit the RTNI package [78] to calculate this integral. It turns out that the exact expression of \(\mathcal{F}_{\text{SEA}}^{(2)}\) with full \(\text{CNOTs}\) is also dominant by

\[
\mathcal{F}_{\text{SEA}}^{(2)} \xrightarrow{d \to \infty} \frac{2}{(d_A^2 - 1)^3(d_B^2 - 1)^3} \left(\begin{array}{c|c|c|c}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right)^4
\]

\[
= \frac{2d_A^4}{(d_A^2 - 1)^3(d_B^2 - 1)^3},
\]

of scaling \(\mathcal{O}(2^{-4N})\). This is the same as that of the global 2-design case, though the exact value of \(\mathcal{F}_{\text{SEA}}^{(2)}\) is larger.

This can be easily understood if we replace \(\text{CNOT}\) gates with identities as the tensor product ansatzes (TEN) in [57], which is easier to compute. Specifically, suppose the number of qubits \(2N\) is divisible by \(k\) so that the \(2N\) qubits could be divided equally into \(k\) subsystems with \(2N/k\) qubits. The tensor product ansatz is defined by \(U = \bigotimes_{i=1}^{k} U_i\) with each unitary \(U_i\) acting on each subsystem. If the ensembles of \(U_i\) are all unitary 2-designs, then the following equality hold:

\[
\mathcal{F}_{\text{TEN}}^{(2)} = 2^{k-1}, \quad \mathcal{F}_{\text{Haar}}^{(2)} = 2^{2N/k} + 1/k \mathcal{F}_{\text{SEA}}^{(2)},
\]

For large \(2N\) and fixed \(k\), we have \(\mathcal{F}_{\text{TEN}}^{(2)} \approx 2^{k-1} \mathcal{F}_{\text{Haar}}^{(2)}\), which means that there is no difference in the sense of scaling between the finite tensor product ensemble and the Haar ensemble. They both behave as \(\mathcal{O}(2^{-4N})\). This partially explains why the second frame potential of the SEA can have the same scaling as the extreme one.

To sum up, the first frame potential of the SEA is exactly identical to that of a global state 1-design \(\mathcal{F}_{\text{Haar}}^{(1)} = 2^{-2N}\) and the second frame potential is in the same scaling as that of a global 2-design \(\mathcal{F}_{\text{Haar}}^{(2)} = 2^{1-2N}(2^{-2N} + 1)\), in spite that the exact value of \(\mathcal{F}_{\text{SEA}}^{(2)}\) is larger than \(\mathcal{F}_{\text{Haar}}^{(2)}\). So we can regard the SEA as a better ansatz in the sense that it has better trainability while the expressibility is not sacrificed too much. From the simulation results in Fig. 3, we can see that the second frame potential of the SEA is pretty close to the optimal value.

It is worth noting that a recent work [66] argues that strategies aimed at avoiding barren plateaus invariably result in a polynomially sized subspace constraining the evolved observable, rendering them classically simulable. We acknowledge that our SEA aligns with the fundamental assumption of this work, employing a loss function formulated as the expectation of an observable for a state evolved under a PQC. However, it is essential to note that SEA distinguishes itself from the primary focus of this work, which centers around strategies for avoiding barren plateaus. SEA presents a methodology designed to mitigate their impact with acceptable sacrifice of expressibility. Consequently, we remark that although the SEA cannot completely avoid barren plateaus since the gradient still vanishes exponentially if the subblocks form local 2-designs, the SEA can greatly mitigate barren plateaus in the sense of slower exponential decay by adjusting the number of \(\text{CNOTs}\) in the...
entangling layer reasonably. Due to this reason, SEA does not fall within the category of classically simulable algorithms in the same manner as the strategies considered in [66], thus having the potential to achieve quantum advantages in practical use. In the following, we will confirm our results with numerical simulations.

IV. NUMERICAL SIMULATION OF EXPERIMENTS
To verify the advantages of SEA on both efficiency and trainability, we investigate the performance as well as the magnitude of cost gradient of SEA and other ansatzes in estimating the ground state energy of chemistry and physics models by carrying out numerical simulations of experiments. All simulations are performed using the Paddle Quantum [82] toolkit on the Paddle–Paddle Deep Learning Platform [83].

We note that all our experiments below are considered under the environment without noise. Previous work [84] showed that noise could also induce barren plateaus. However, unlike “standard” barren plateaus where gradients vanish exponentially with the number of qubits due to measure concentration in a large Hilbert space, the noise-induced barren plateau means that the gradient vanishes exponentially at every point on the cost function landscape with the number of layers in the quantum circuit. Therefore, while both phenomena present challenges to trainability, the underlying causes are different and independent. As a consequence, we only consider the noise-free circumstance in this article.

A. APPLYING TO CHEMISTRY AND PHYSICS MODELS
We adopt the alternating layered ansatz (ALT) [63] (depicted in Fig. 4) as the subblocks of SEA. We emphasize that the purpose of our experiments is merely to demonstrate the better trainability of SEA over general ansatzes, such as the ALT given an equal number of parameters, instead of proving that it outperforms any other ansatzes in any tasks, especially those specifically designed ansatzes to solve quantum chemistry and physical models, such as the UCCSD ansatz.

Here we use 14-qubit dinitrogen (N2) molecule and the 14-qubit Heisenberg model as examples for numerical simulations of VQE. The structure of SEA is shown in Fig. 2, where U1, U2, and U3 are all 7-qubit ALTs with a depth of 30. For different Hamiltonians, we could adjust the K in Proposition 1 to obtain SEAs with different entangling capabilities. Considering that the N2 molecule and the Heisenberg model we choose are both weakly entangled, we use two kinds of SEA named SEA2 and SEA3 to learn these models’ ground state energies. To be specific, SEA2 is the 14-qubit SEA of Schmidt coefficient layer as a 2-qubit ALT and entangling layer as 2 CNOTS, which means a composition of 2 CNOTS controlled and targeted on the qubit-pairs \{ (0, 7), (1, 8) \}. SEA3 is the 14-qubit SEA of Schmidt coefficient layer as a 3-qubit ALT and entangling layer as 3 CNOTS, which implement on the qubit-pairs \{(i, 7 + i)\}_{i=0}^{7}.

The specific structure of SEA2 and SEA3 can be found in Section E of the Appendix. As mentioned before, we can also set the Schmidt coefficient layer as a subSEA. To verify this point, we use an SEA with the Schmidt coefficient layer being the 7-qubit subSEA, which is called SEA_Sch, in our simulations.

To study the advantages of SEA, we compare a 14-qubit SEA with a 14-qubit ALT and a 14-qubit random circuit [53] in the task of VQE with a similar number of parameters by setting different depth for different ansatzes. The calculation methods of the number of parameters are given in the Table I and the specific depths and number of parameters in different ansatzes used in our numerical experiments are shown in Table II. The reason for choosing an ALT and a random circuit is that they both have a strong expressibility and they can be regarded as a 2-design when their depths are large [53], [63]. After setting up the ansatzes, we employ a stochastic gradient descent optimizer to iteratively update parameters for 400 iterations.

The simulation results of N2 molecule and the Heisenberg model are shown in Fig. 5(a) and (b), respectively.

Both figures illustrate the efficiency of SEA on VQE. We can observe that with a similar number of parameters, there is a visible gap in the accuracy of estimating the ground energy by different ansatzes. Both SEA 2 and SEA3 converge rapidly and get an approximate value of ground energy while the results of the ALT and random circuit are hard to be optimized in 400 iterations. Furthermore, it is also valid to replace the three parts of SEA with other structures. The SEA_Sch in Fig. 5 shows that when we choose the Schmidt coefficient layer as subSEA, this kind of SEA also has a good performance on VQE like SEA 2 or SEA3. In addition to SEA_Sch, the Section E of the Appendix presents the result of SEA whose three parts are constructed by other circuits. We can also see that this kind of SEA has better performance than ALT and random circuit, which further verifies the effectiveness of SEA in practical applications. Since we only use 2 or 3 CNOTS and adjust the structure of the Schmidt coefficient layer to obtain good results, the experiment results also support Proposition 1.

Due to the limitations of the experimental equipment, there is still a gap between our simulation result and the accurate ground state energy of these models. However, it is worth noting that our choice of these two models as case studies aims to highlight SEA’s training performance and its potential for larger systems, rather than pursuing high accuracy for specific systems like nitrogen molecules. As demonstrated in the numerical experiment of VQE, SEA exhibits superior trainability and faster convergence compared to other PQCs. Furthermore, by enhancing the structure of the SEA, such as deepening the layer of Ui and constructing the subblocks of SEA with special designed structures, we can improve the accuracy of simulation results. In addition, utilizing advanced
TABLE I. Comparison of the Number of Parameters in Different Ansatzes

| Ansatz   | Parameter Number |
|----------|------------------|
| SEA      | $3N + 6(N-1)D_{SEA}$ |
| ALT      | $2N + 2(2N - 1)D_{ALT}$ |
| Random   | $2N D_{R}$ |

TABLE II. Comparison of Depth and the Number of Parameters in Different Ansatzes in Numerical Experiments Shown in Fig. 5

| Ansatz | Depth | Parameter Number |
|--------|-------|------------------|
| SEA    | 30    | 1101             |
| ALT    | 42    | 1106             |
| Random | 79    | 1106             |

optimizers and employing clever parameters tuning methods are also expected to further enhance accuracy.

B. GRADIENT EXPERIMENTS

For both experiments on gradients with different numbers of qubits, we consider the 1-D spin-1/2 antiferromagnetic Heisenberg model on $2N$ qubits where the Hamiltonian reads $H = \sum_i (X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1})$ with periodic boundary condition. We use different ansätze and calculate the gradient of the cost function $C(\theta) = \langle 0 | \otimes^N U(\theta)^\dagger H U(\theta) | 0 \rangle | \otimes^N$ with respect to each parameter, where $U(\theta)$ represents the unitary of the ansatz.

We first conduct experiments with the ansätze used in the previous numerical simulations of VQE. The SEAs we chose are SEA$_1$ and SEA$_{N/2}$, which still adopt ALT as the Schmidt coefficient layer and the LBC layer.

For each $N$, ALT and random circuit with similar numbers of parameters as SEAs are selected for comparison. The average variances of parameters of different ansätze are compared in Fig. 6(a).

We also extract the variance of the largest partial derivative in each sample, which is shown in the Section E of the Appendix. It is clearly demonstrated that all these variances decay exponentially as the number of qubits increases. However, the slopes of the red and the blue lines are nearly half of the slopes of the other lines, which means the SEA has a much lower rate of decay of variance.

To better illustrate how our structure increases the scaling of variance, we further inspect the gradient with respect to the parameter in different parts of SEA. As shown in Fig. 7, we look into the gradients with respect to the parameters of the $R_y$ gates in the Schmidt coefficient layer and the LBC layer. As a comparison, we also consider a single parameter $R_y$ gate in the middle of a circuit that forms a 2-design with sufficient depth. The two parts split by the gate are both local 2-designs and are represented by Haar random unitaries in our experiments. Fig. 6(b) summarizes the results of the variance of parameters located in the Schmidt coefficient layer and the LBC layer, respectively.

From Fig. 6(b), we see that no matter where this parameter is located in SEA, the variance of its gradient has a decaying speed lower than the decaying speed of variance of the parameter in a sufficiently random circuit that forms a 2-design. This fact indicates that the variance of SEA has significant improvement compared to using an ansatz that forms a 2-design, indicating evident mitigation of barren plateaus. It is also worth noting that the fewer CNOTs in the entangling layer, the larger variance it will have, which is in line with Proposition 3. Specifically, in our experiments, if the scale of the problem is larger than 12 qubits, the variance of the gradients in a 2-design circuit will be less than $10^{-3}$. However, this bound of scale can be increased to 18 qubits using SEA with $\binom{N}{2}$ CNOTs according to our experiment results. With SEA of weaker expressibility, this bound can be further increased.

In other words, we can notably increase the scale of problems with the same limited scaling of gradients by using SEA instead of a 2-design circuit. Therefore, SEA is more likely to scale to larger systems on VQE tasks because of its improvement in trainability by mitigating barren plateaus.

V. CONCLUDING REMARKS

The barren plateau phenomenon, which is associated with the strong expressibility of an ansatz, severely restricts the trainability of the VQE, rendering it impractical for many quantum applications. This limitation has been recognized as one of the main obstacles to achieving quantum advantages in VQAs. Consequently, addressing barren plateaus is crucial for the practical use of VQAs on larger quantum computers. While there have been numerous attempts to tackle BP from various algorithmic perspectives, it is still an unsolved issue for general problems, and even a tiny step toward solving it is significant as algorithms scale up. In our work, the main contribution is proposing an ansatz structure, SEA, to mitigate barren plateaus in optimizing parameterized circuits for ground state preparation while preserving useful expressibility. In other words, SEA broadens a new perspective to address BP by balancing the tradeoff between expressibility and trainability, which is essential for the development of NISQ and the advancement of quantum computing. Specifically, with the idea of Schmidt decomposition, we invent the SEA by removing the redundant expressibility in universal PQCs. We in particular have provided an explicit construction of SEA that possesses sufficient expressibility to generate arbitrary pure quantum states. Moreover, SEA does not form a unitary 2-design, and by tuning the number of CNOT gates, it can efficiently mitigate barren plateaus. From the perspective of frame potential, we demonstrate that SEA maintains high trainability without sacrificing too much expressibility, allowing for a tradeoff between trainability and expressibility. Numerical simulations show that SEA can be applied to approximate the ground state of the N$_2$ molecule and the Heisenberg model, with numerical results confirming the effectiveness of SEA. We further compare the scaling of variance in partial derivatives between different ansätze and...
establish that SEA can obtain a quadratically lower rate of decay of variance, implying that SEA can considerably mitigate the barren plateaus of VQE. To the best of the authors’ knowledge, SEA is the very first ansatz that can continuously deepen the circuit without forming a 2-design and can still effectively complete tasks. Our work also opens a new avenue toward solving trainability obstacles of VQAs via designing tailored ansatzes considering the feature of entanglement and quantum information theory.

We note that SEA is also effective in principle for other tasks with the sole purpose of generating a pure state, such as combinatorial optimization [85] and entanglement detection [86]. It is natural to explore how to extend our structure by considering multipartite systems and other possible designs. We also remark that there are other proposals for addressing barren plateaus [49], [87], [88], [89], and using meta-learning to improve efficiency [90], [91], [92], [93]. Combining SEA with other approaches, such as adaptive methods, may merit future studies. Moreover, with recent works exploring the symmetry of VQAs [94], [95], [96], it is also worth investigating how to enhance the performance of SEA from this perspective.

APPENDIX

A. PRELIMINARIES

Here, we present some supplemental lemmas and define some notations to be used throughout the proof. We write a target quantum state, which is a pure state to be approximated by a PQC, as \( |\psi(\theta)\rangle \) and let the actual circuit output be \( |\psi(\lambda)\rangle \). We may simply write \( |\psi\rangle \) if the circuit parameters \( \theta \) are unimportant in the context. Moreover, let \( F(|\phi\rangle, |\psi\rangle) = \langle|\psi|\phi\rangle^2 \) be fidelity. Unless otherwise specified, \( \{|k\rangle\}_{k=0}^{2^N-1} \) are the N-qubit computational bases. We denote \( \text{cnot} \equiv \prod_{i=0}^{N-1} \text{cnot}(i, N+i) \), which is a composition of \( N \) cnots controlled on the pair \((i, N+i)\) for all \( i = 0, \ldots, N-1 \). Let \( \tilde{i} \) be a vector for the binary expansion of \( i \), and therefore \( \tilde{i}_j \), where \( 0 \leq j < \lceil \log_2 i \rceil \), is the \( j \)th leading digit of the expansion.

We first recall the definition of a \( t \)-design. Assume \( U = \{U_k\}_{k=1}^{K} \) is a finite set of unitary operators on \( \mathbb{C}^d \), and \( P_{t,i}(U) \) is a polynomial of degree at most \( t \) in the matrix elements of \( U \) and at most \( t \) in those of \( U^\dagger \). Then, we define that \( U \) is a \( t \)-design if for every polynomial \( P_{t,i}(U) \) we have

\[
\frac{1}{K} \sum_{k=1}^{K} P_{t,i}(U_k) = \int P_{t,i}(U)d\eta(U) \tag{12}
\]

where the integrals with respect to the Haar measure over the unitary group. Particularly, when \( t = 2 \), the definition is equivalent to the following definition [97].

**Definition 1:** \( \{U_k\}_{k=1}^{K} \) forms a unitary 2-design if and only if for any linear operators \( C, D, \rho \in L(\mathbb{C}^d) \), we have

\[
\frac{1}{K} \sum_{k=1}^{K} U_k^\dagger C U_k \rho U_k^\dagger D U_k = \int_{U(d)} U^\dagger C U \rho U^\dagger D d\eta(U). \tag{13}
\]

Based on this definition, we can know whether a unitary group is a 2-design by comparing the both sides of (13). Fortunately, according to the Schur’s lemma [98], the RHS of (14) has a closed form, which is shown in Lemma S1. The proof of Lemma S1 can be seen in [75].

**Lemma S1:** For any linear operators \( C, D, \rho \in L(\mathbb{C}^d) \), we have

\[
\int_{U(d)} U^\dagger C U \rho U^\dagger D d\eta(U) = \frac{\text{Tr}(CD)[\text{Tr}(\rho)]}{d} \frac{1}{d} + \left( \frac{\text{Tr}(C)[\text{Tr}(D)] - \text{Tr}(CD)}{d(d^2-1)} \right) \times \left( \rho - \frac{\text{Tr}[\rho]}{d} I \right). \tag{14}
\]

Furthermore, we present the following lemma so that we can solve the problem with bipartite system.

**Lemma S2:** For any bipartite state \( \rho_{AB} (d_A = d_B = d) \), and arbitrary linear operators \( C, D \in L(\mathbb{C}^d) \), we have

\[
\int_{U(d) \otimes U(d)} d\eta(U) (U^\dagger C U \rho U^\dagger D) = t_0 \rho + t_1 \frac{\rho_A \otimes I_B}{d} + t_2 \frac{I_A \otimes \rho_B}{d} + t_3 I_{AB} \text{Tr}(\rho_{AB}) \tag{15}
\]

where \( t_0 = \text{Tr}[\rho_{AB}], t_1 = \text{Tr}[(\rho_A \otimes I_B)_{AB}], t_2 = \text{Tr}[(I_A \otimes \rho_B)_{AB}], \) and \( t_3 = \text{Tr}(I_{AB}) \) can be computed from the following linear system of equations:

\[
\begin{align*}
\text{Tr}(CD) &= t_0 d^2 + t_1 d^2 + t_2 d^2 + t_3 d^4 \tag{16} \\
\text{Tr}(C_A D_A) &= t_0 d^3 + t_1 d^2 + t_2 d^2 + t_3 d^3 \tag{17} \\
\text{Tr}(C_B D_B) &= t_0 d^3 + t_1 d^2 + t_2 d^2 + t_3 d^3 \tag{18} \\
\text{Tr}(C) \text{Tr}(D) &= t_0 d^2 + t_1 d^2 + t_2 d^2 + t_3 d^2 \tag{19}
\end{align*}
\]

that is

\[
\begin{bmatrix}
1 \\
t_0 \\
t_1 \\
t_2 \\
t_3
\end{bmatrix} = \left( \frac{1}{(d^2-1)^2} \right) \begin{bmatrix}
\text{Tr}(CD) \\
\text{Tr}(C_A D_A) \\
\text{Tr}(C_B D_B) \\
\text{Tr}(C) \text{Tr}(D)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\text{Tr}(CD) - \text{Tr}(C_A D_A) + \text{Tr}(C_B D_B)
\text{Tr}(CD) + \text{Tr}(C_A D_A) - \text{Tr}(C_B D_B)
\text{Tr}(CD) + \text{Tr}(C_A D_A) + \text{Tr}(C_B D_B)
\text{Tr}(CD) - \text{Tr}(C_A D_A) - \text{Tr}(C_B D_B)
\end{bmatrix}
\tag{20}
\]

Proof: Similar to the proof of Lemma S1, Schur’s lemma implies that the LHS of (15) has an explicitly expression, which is denoted as follows:

\[
\int_{U(d) \otimes U(d)} d\eta(U) (U^\dagger C U \rho U^\dagger D) = t_0 \rho + t_1 \frac{\rho_A \otimes I_B}{d} + t_2 \frac{I_A \otimes \rho_B}{d} + t_3 I_{AB} \text{Tr}(\rho_{AB}). \tag{22}
\]
For simplicity, we define the following two functions:

\[ f^{(1)}(\rho) = t_0 \rho + t_1 \rho_A \otimes \frac{I_B}{d} + t_2 I_A \otimes \frac{\rho_B}{d} + t_3 I_{AB} \text{Tr}(\rho_{AB}) \]  

(23)

\[ f^{(2)}(\rho) = \int_{U(d) \otimes U(d)} d\eta(U) U^\dagger C U \rho U^\dagger DU \]  

(24)

where \( f^{(1)} = f^{(1)} \).

Then, we only need to consider the coefficients \( t_0, t_1, t_2, t_3 \).

To calculate the coefficients, we defined the following four operators:

\[ T_1(f) = \sum_{i,j} (ij|i)(ij) \]  

(25)

\[ T_2(f) = \sum_{i,j,k} (ik|f(|i))(j) \otimes I_B |j)k) \]  

(26)

\[ T_3(f) = \sum_{i,j,k} (ij|f(I_A \otimes |k)(j)|i)l \]  

(27)

\[ T_4(f) = \sum_{i,j,k,l} (ij|f(|k)(j)|l)k \]  

(28)

For (16), we have

\[ T_1(f^{(1)}) = t_0 \sum_{i,j} (ij|i)(ij) + t_1 \sum_{i,j} (ij|i)(j)A \otimes I_B |j) \]  

(29)

\[ = t_0 d^2 + t_1 d^2 + t_2 d^2 + t_3 d^3 \]  

(30)

\[ T_1(f^{(2)}) = \sum_{i,j} \int_{U(d) \otimes U(d)} d\eta(U) U^\dagger C U \eta U^\dagger DU |ij \]  

(31)

\[ = \text{Tr}(CD). \]  

(32)

Since \( T_1(f^{(1)}) = T_1(f^{(2)}) \), then \( \text{Tr}(CD) = t_0 d^2 + t_1 d^2 + t_2 d^2 + t_3 d^3 \).  

For (18), we have

\[ T_2(f^{(1)}) = t_0 \sum_{i,j,k} (ik|f(|i)(j) \otimes I_B |j)k) + t_1 \sum_{i,j,k} (ik|f(|i)(j) \otimes I_B |j)k) \]  

(33)

\[ = t_0 d^3 + t_1 d^3 + t_2 d + t_3 d^3 \]  

(34)

\[ T_2(f^{(2)}) = \sum_{i,j,k} (ik|f(|i)(j) \otimes I_B |j)k \int_{U(d) \otimes U(d)} d\eta(U) U^\dagger C U \eta U^\dagger DU |ij \]  

(35)

\[ = \sum_{i,j,k} \int_{U(d) \otimes U(d)} d\eta(U) \cdot (ik|U^\dagger C (|i)(j)A \otimes I_B U^\dagger DU |j)k) \]  

(36)

\[ = \text{Tr}(C_B D_B). \]  

(37)

Since \( T_2(f^{(1)}) = T_2(f^{(2)}) \), then \( \text{Tr}(C_B D_B) = t_0 d^3 + t_1 d^3 + t_2 d + t_3 d^3 \).  

Similarly, we have

\[ T_3(f^{(1)}) = t_0 d^3 + t_1 d^3 + t_2 d^2 + t_3 d^3 \]  

(38)

\[ T_3(f^{(2)}) = \text{Tr}(C_A D_A) \]  

(39)

\[ T_4(f^{(1)}) = t_0 d^4 + t_1 d^2 + t_2 d^2 + t_3 d^2 \]  

(40)

\[ T_4(f^{(2)}) = \text{Tr}(C) \text{Tr}(D). \]  

(41)

Thus

\[ \text{Tr}(C_A D_A) = t_0 d^3 + t_1 d + t_2 d^2 + t_3 d^3 \]  

(42)

\[ \text{Tr}(C) \text{Tr}(D) = t_0 d^4 + t_1 d^2 + t_2 d^2 + t_3 d^2. \]  

(43)

Up to now, we have proved (16) to (19).

Let

\[ R = \begin{pmatrix}
  d^2 & d^2 & d^2 & d^3 \\
  d^2 & d^3 & d^3 & d^3 \\
  d^2 & d^3 & d^3 & d^3 \\
  d^3 & d^3 & d^3 & d^3
\end{pmatrix}. \]  

(47)

Then

\[ R^{-1} = \frac{1}{(d^2 - 1)^2} \begin{pmatrix}
  \frac{1}{d^2 - 1} & -\frac{1}{d^2} & \frac{1}{d^2} & 1 \\
  -\frac{1}{d^2} & \frac{1}{d^2} & -\frac{1}{d^2} & -1 \\
  -\frac{1}{d^2} & \frac{1}{d^2} & -\frac{1}{d^2} & 1 \\
  1 & -\frac{1}{d^2} & -\frac{1}{d^2} & \frac{1}{d^2}
\end{pmatrix}. \]  

(48)

Hence, we have

\[ \begin{bmatrix}
  t_0 \\
  t_1 \\
  t_2 \\
  t_3
\end{bmatrix} = R^{-1} \begin{bmatrix}
  \text{Tr}(CD) \\
  \text{Tr}(C_A D_A) \\
  \text{Tr}(C_B D_B) \\
  \text{Tr}(C) \text{Tr}(D)
\end{bmatrix} \]

\[ = \frac{1}{(d^2 - 1)^2} \left[
\begin{array}{c}
  \text{Tr}(CD) - \frac{\text{Tr}(C_B D_B)}{d^2} - \frac{\text{Tr}(D)}{d^2} + \text{Tr}(C) \text{Tr}(D) \\
  -\text{Tr}(CD) + \frac{\text{Tr}(C_B D_B)}{d^2} + d \text{Tr}(C_B D_B) - \text{Tr}(C) \text{Tr}(D) \\
  -\text{Tr}(CD) + d \text{Tr}(C_A D_A) + \frac{\text{Tr}(C_B D_B)}{d^2} - \text{Tr}(C) \text{Tr}(D) \\
  \text{Tr}(CD) - \frac{\text{Tr}(C_A D_A)}{d^2} - \frac{\text{Tr}(D)}{d^2} + \text{Tr}(C) \text{Tr}(D)
\end{array}\right]. \]
**B. EFFECTIVENESS OF SEA ON VQE**

In this section, we present a proof of the effectiveness of SEA on VQE. We first give the Proposition S1 to illustrate that SEA can generate an arbitrary pure state. Then, in Proposition S2 we demonstrate that the SEA in Proposition S1 can find the ground state energy of a Hamiltonian, which implies the effectiveness of SEA on VQE.

**Proposition S1:** If $U_1$ can generate an arbitrary $N$-qubit pure state, then given any $2N$-qubit pure state $|\phi\rangle$, a $2N$-qubit SEA can generate $|\phi\rangle$ with a certain set of parameters $\hat{\theta} = \{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\}$, such that

$$S(\hat{\theta})|0\rangle^{\otimes 2N} = |\phi\rangle. \quad (49)$$

**Proof:** Beginning with Schmidt decomposition, we can write the target state $|\phi\rangle$ as

$$|\phi\rangle = (\hat{U}_2 \otimes \hat{U}_3) \sum_{k=0}^{2^N-1} \lambda_k |k\rangle_A |k\rangle_B \quad (50)$$

where $A$ and $B$ are two $N$-qubit subsystems, $\hat{U}_2$ and $\hat{U}_3$ are two unitary operators acting on these two subsystems, $\{|k\rangle\}_{k=0}^{2^N-1}$ are the $N$-qubit computational bases, and $(\lambda_k)_{k=0}^{2^N-1}$ are Schmidt coefficients.

Because $U_1(\theta_1)$ can generate an arbitrary $N$-qubit pure state, and $U_2(\theta_2), U_3(\theta_3)$ are universal, we can choose a certain set of parameters $\hat{\theta} = \{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\}$ such that

$$U_1(\hat{\theta}_1)|0\rangle^{\otimes N} = \sum_{k=0}^{2^N-1} \lambda_k |k\rangle \quad (51)$$

$$U_2(\hat{\theta}_2) = \hat{U}_2 \quad (52)$$

$$U_3(\hat{\theta}_3) = \hat{U}_3. \quad (53)$$

Combining with $V$, which is a composition of $N$-cNOTs controlled and targeted on the qubit-pairs $(i, N+1)_{i=0}^{N-1}$, there has

$$V(U_1(\hat{\theta}_1) \otimes I)|0\rangle^{\otimes N} |0\rangle^{\otimes N} = \sum_{k=0}^{2^N-1} \lambda_k V|k\rangle_A |0\rangle_B \quad (54)$$

$$= \sum_{k=0}^{2^N-1} \lambda_k |k\rangle_A |k\rangle_B. \quad (55)$$

Therefore, we have

$$S(\hat{\theta})|0\rangle^{\otimes 2N} = (U_2(\hat{\theta}_2) \otimes U_3(\hat{\theta}_3))V(U_1(\hat{\theta}_1) \otimes I)|0\rangle^{\otimes N} |0\rangle^{\otimes N} \quad (56)$$

$$= (\hat{U}_2 \otimes \hat{U}_3) \sum_{k=0}^{2^N-1} \lambda_k |k\rangle_A |k\rangle_B \quad (57)$$

$$= |\phi\rangle. \quad (58)$$

**Proposition S2:** For any $2N$-qubit Hamiltonian $H$, it holds that

$$\min_{\hat{\theta}} \langle 0 | S^{\otimes 2N} S^{\dagger} HS | 0 \rangle^{\otimes 2N} = E_0 \quad (59)$$

where $E_0$ is the ground state energy of $H$ and the optimization is over all unitaries reachable by SEA with $U_1$ having universal wavefunction expressibility.

**Proof:** A given $2N$-qubit Hamiltonian $H$ can be written as follows according to spectral decomposition:

$$H = \sum_{i=0}^{m} E_i P_i \quad (60)$$

where $\{E_i\}_{i=0}^{m}$ are eigenvalues of $H$ such that $E_0 < E_1 < \cdots < E_m$ and $P_i$ is a projector onto the eigenspace $V_i$ corresponding to the eigenvalue $E_i$. Then, given an arbitrary pure state $|\psi\rangle \in V_0$, we have

$$H|\psi\rangle = E_0|\psi\rangle. \quad (61)$$

Note that when the optimization is over all unitaries reachable by SEA with $U_1$ having universal wavefunction expressibility, there exists an $S$ such that $|\psi\rangle = S|0\rangle^{\otimes 2N}$. Therefore

$$\langle 0 | S^{\otimes 2N} S^{\dagger} HS | 0 \rangle^{\otimes 2N} = \langle \psi | H | \psi \rangle = E_0 \quad (62)$$

$$= E_0|\psi\rangle \quad (63)$$

$$= E_0. \quad (64)$$

Since it is trivial that $\min_{S} \langle 0 | S^{\otimes 2N} S^{\dagger} HS | 0 \rangle^{\otimes 2N} \geq E_0$, we have

$$\min_{\hat{\theta}} \langle 0 | S^{\otimes 2N} S^{\dagger} HS | 0 \rangle^{\otimes 2N} = E_0. \quad (65)$$

**C. PROOF OF PROPOSITION 1**

In this section, we give a detailed proof of Proposition 1. We first give the following lemma that is used in the proof.

**Lemma S3:** For any descending sequence $\{x_i\}_{i=1}^{p} \geq x_{i+1}$, and any $n \geq N \geq M \geq 1$ ($n, N, M \in \mathbb{Z}$), the following is true:

$$\frac{1}{M} \sum_{i=1}^{M} x_i \geq \frac{1}{N} \sum_{i=1}^{N} x_i. \quad (66)$$

**Proof:** Because $N \geq M \geq 1$ and $x_i \geq x_{i+1}$, we have the following formula:

$$\frac{1}{M} \sum_{i=1}^{M} x_i - \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{1}{M} \sum_{i=1}^{M} x_i - \left( \frac{1}{N} \sum_{i=1}^{M} x_i + \frac{1}{N} \sum_{i=M+1}^{N} x_i \right) \quad (67)$$

$$= \left( \frac{1}{M} - \frac{1}{N} \right) \sum_{i=1}^{M} x_i - \frac{N}{N} \sum_{i=M+1}^{N} x_i \quad (68)$$
\[
\frac{1}{N} \left( \frac{N - M}{M} \sum_{i=1}^{M} x_i - \frac{1}{N} \sum_{i=M+1}^{N} x_i \right)
\]

(69)

\[
\leq \frac{N - M}{N} \left( \frac{M}{1} \sum_{i=1}^{N} x_i - \frac{1}{N - M} \sum_{i=M+1}^{N} x_i \right)
\]

(70)

\[
\geq \frac{N - M}{N} \left( \frac{1}{M} \sum_{i=1}^{M} x_M - \frac{1}{N - M} \sum_{i=M+1}^{N} x_{M+1} \right)
\]

(because, \(x_i \geq x_{i+1}\))

(71)

\[
= \frac{N - M}{N} (x_M - x_{M+1})
\]

(72)

\[
\geq 0,
\]

(73)

Thus, we obtain \(\frac{1}{M} \sum_{i=1}^{M} x_i \geq \frac{1}{N} \sum_{i=1}^{N} x_i \geq 0\), that is, \(\frac{1}{M} \sum_{i=1}^{M} x_i \geq \frac{1}{N} \sum_{i=1}^{N} x_i \geq 0\).

**Proposition 1:** If \(U_1\) can generate any \(N\)-qubit pure state that is a superposition of at most \(K\) computational basis states, then for any \(|\phi\rangle\), there exists an SEA output state \(|\psi\rangle\) with \(F(|\phi\rangle, |\psi\rangle) \geq \min(\frac{1}{N}, 1)\), where \(F(|\phi\rangle, |\psi\rangle)\) is the fidelity between \(|\phi\rangle\) and \(|\psi\rangle\), and \(r\) is the Schmidt rank of \(|\phi\rangle\).

**Proof:** For any 2\(N\)-qubit target state \(|\phi\rangle = \sum_{k=0}^{r-1} \lambda_k |v_k\rangle_A|v_k\rangle_B\), we explicitly construct an SEA, such that \(|\psi\rangle = S(\hat{\theta})|0\rangle^{\otimes 2N}\) and \(F(|\phi\rangle, |\psi\rangle) \geq \min(\frac{1}{N}, 1)\), where \(\hat{\theta} = \{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\}\) is a certain set of parameters.

Suppose \(\{\lambda_k\}_{k=0}^{r-1}\) is the Schmidt coefficients of \(|\phi\rangle\) sorted in descending order. Similar to the proof of Proposition S1, we can choose a certain set of parameters \(\hat{\theta}\) such that

\[
U_1(\hat{\theta}_1)|0\rangle^A = \frac{1}{\sqrt{M}} \sum_{k=0}^{\min(K,r)-1} \lambda_k |k\rangle
\]

(74)

\[
U_2(\hat{\theta}_2)|k\rangle_A = |v_k\rangle_A
\]

(75)

\[
U_3(\hat{\theta}_3)|k\rangle_B = |v_k\rangle_B
\]

(76)

where \(M = \sum_{k=0}^{\min(K,r)-1} \lambda_k^2\). Then

\[
|\psi\rangle = S(\hat{\theta})|0\rangle^{\otimes 2N} = \frac{1}{\sqrt{M}} \sum_{k=0}^{\min(K,r)-1} \lambda_k |v_k\rangle_A|v_k\rangle_B.
\]

(77)

Hence

\[
F(|\phi\rangle, |\psi\rangle) = |\langle \phi | \psi \rangle|^2
\]

(78)

\[
= \left( \frac{1}{\sqrt{M}} \sum_{k=0}^{\min(K,r)-1} \lambda_k^2 \right)^2
\]

(79)

\[
= \frac{1}{M} \left( \sum_{k=0}^{\min(K,r)-1} \lambda_k^2 \right)^2
\]

(80)

\[
\leq \sum_{k=0}^{\min(K,r)-1} \lambda_k^2
\]

(81)

\[
\geq \frac{\min(K,r)}{r} \sum_{k=0}^{\min(K,r)-1} \lambda_k^2
\]

(82)

\[
= \min\left\{ \frac{K}{r}, 1 \right\}
\]

(83)

\[
= \frac{\min(K,r)}{r}
\]

(84)

with equality holds if and only if \(\lambda_i = \lambda_j, \forall i, j = 0, \ldots, r - 1\).

**D. PROOF OF PROPOSITION 2**

Here, we prove the Proposition 2 about SEA does not form a 2-design. We first present Lemma S4, which is used in the proof.

**Lemma S4:** \(C_2N\Omega = \sum_{i=0}^{2^n-1} |i\rangle\langle i|A \otimes V_{B_i}\), where \(A, B\) are subsystems and \(V_{B_i} = \bigotimes_{j \neq i}^{n} X_j, i \in \{0, 1\}\), that is the operator on the subsystem \(B\) that represents the binary bit of \(i\) is 1 acting on pauli \(X\), and 0 acting on \(I\). Then, \(V_{B_i}\) has the following properties:

1) \([V_{B_i}, V_{B_j}] = 0\);
2) \(V_{B_i}^2 = V_{B_i}\);
3) \(|0\rangle^{\otimes n}V_{B_i}|0\rangle^{\otimes n} = \delta_{ij}\)

where \(\delta_{ij}\) is Kronecker delta.

**Lemma S4** can be easily proved using properties of Pauli operators.

**Proposition 2:** SEA with \(U_i(i = 1, 2, 3)\) being local 2-design and \(C_2N\Omega\) as entangling layer does not form a 2-design on the global system.

**Proof:** Without losing generality, we only consider the SEA with 2\(n\) qubits in this proof. Lemma S1 says that if an ansatz forms a 2-design, the (14) should hold for any linear operators \(C, D, \rho \in L(\mathbb{C}^{d^2})\). Assuming \(d = 2^n\), \(C_0 = D_0 = \rho = |0\rangle\langle 0|^{\otimes 2n}\), then if SEA forms a 2-design, we have

\[
\int_{U(\theta)} |\langle 0|^{\otimes 2n}U^\dagger C_0 U \rho U^\dagger D_0 U|0\rangle^{\otimes 2n} d\eta(U)
\]

(85)

\[
= \frac{2}{d^2(d^2+1)}
\]

(86)

where \(U(\theta)\) denotes the subset of unitary group generated by SEA, and \(d(\eta(U)\) denotes the Haar measure.

In the following steps, we will explicitly calculate the (85) and prove that it cannot be \(\frac{2}{d^2(d^2+1)}\).
For the SEA with $U_i(i = 1, 2, 3)$ being local 2-design and $CNOT = V = \sum_{i=0}^{2^n-1} |i⟩⟨i| \otimes V_{B_i}$, the unitary of this ansatz is

$$U = \sum_{i=0}^{2^n-1} (U_2 \otimes U_3) \cdot |i⟩⟨i| \otimes V_{B_i} \cdot (U_1 \otimes I_B). \quad (87)$$

Therefore

$$U|0⟩^{\otimes 2n} = \sum_{i=0}^{2^n-1} (U_2 \otimes U_3) \cdot |i⟩⟨i| \otimes V_{B_i} \cdot (U_1 \otimes I_B)|0⟩^{\otimes 2n} \quad (88)$$

$$⟨0|^{\otimes 2n} U^† = \sum_{i=0}^{2^n-1} ⟨0|^{\otimes 2n} (U_1^† \otimes I_B) \cdot |i⟩⟨i| \otimes V_{B_i} \cdot (U_2^† \otimes U_3^†). \quad (89)$$

Then

$$⟨0|^{\otimes 2n} U^† C_0 U \rho U^† D_0 U|0⟩^{\otimes 2n} = \sum_{i=0}^{2^n-1} ⟨0|^{\otimes 2n} U^† (0|0⟩^{\otimes 2n} U((0|0⟩^{\otimes 2n} U^† (0|0⟩^{\otimes 2n} U|0⟩^{\otimes 2n} \quad (90)$$

$$⟨0|^{\otimes 2n} U^† \cdot (U_1^† \otimes I_B) \cdot V^† \cdot (U_2^† \otimes U_3^†) (0|0⟩^{\otimes 2n} (U_2 \otimes U_3) \cdot V \cdot (U_1 \otimes I_B) \cdot V^† \cdot (U_2^† \otimes U_3^†) \quad (91)$$

$$= \sum_{i=0}^{2^n-1} I_A \otimes \langle k|B \cdot V \cdot (U_1 \otimes I_B) (0|0⟩^{\otimes 2n} (U_1^† \otimes I_B) \cdot V^† \cdot (U_2^† \otimes U_3^†) \quad (92)$$

where

$$V = \sum_{i=0}^{2^n-1} |i⟩⟨i| A \otimes V_{B_i} \quad (93)$$

$$\rho_{1} = V \cdot (U_1 \otimes I_B) (0|0⟩^{\otimes 2n} (U_1^† \otimes I_B) \cdot V^†. \quad (94)$$

Therefore, we have

$$\rho_{1}^A = Tr_B[\rho_{1}] \quad (95)$$

$$= \sum_{k=0}^{d-1} I_A \otimes \langle k|B \cdot V \cdot (U_1 \otimes I_B) (0|0⟩^{\otimes 2n} (U_1^† \otimes I_B). \quad (96)$$

$$= \sum_{k=0}^{d-1} I_A \otimes (k|B \sum_{i=0}^{2^n-1} |i⟩⟨i|A\otimes V_{B_i}) (U_1 \otimes I_B) (0|0⟩^{\otimes 2n} \quad (97)$$

Then, combining Lemma S1 and Lemma S2, we have

$$\int_{U(θ)} |0⟩^{\otimes 2n} U^† C_0 U \rho U^† D_0 U|0⟩^{\otimes 2n} dθ(U) \quad (99)$$

$$= \int_{U(θ)} |0⟩^{\otimes 2n} (U_1^† \otimes I_B) \cdot V^† \cdot (U_2^† \otimes U_3^†) \cdot C_0 \cdot (U_2 \otimes U_3) \cdot \rho_{1} \cdot (U_2^† \otimes U_3^†) \cdot D_0 \cdot (U_2 \otimes U_3) \cdot V \cdot (U_1 \otimes I_B) (0|0⟩^{\otimes 2n} \quad (100)$$
FIGURE 8. Specific structure of 14-qubit SEA₂ and SEA₃. 14-qubit SEA₃ is the SEA of Schmidt coefficient layer as a 2-qubit ALT and entangling layer as 2 CNOTs, which is a composition of 2 CNOTs controlled and targeted on the qubit-pairs (i, j). SEA₂ is the SEA of Schmidt coefficient layer as a 3-qubit ALT and entangling layer as 3 CNOTs, which implement on the qubit-pairs (i, j, k). The LBC layers of SEA₂ and SEA₃ are 7-qubit ALT.

FIGURE 9. Numerical experiment of VQE on LiH (12-qubit). The blue dotted line is the theoretical ground energy of LiH, and the lines from top to bottom represent the experimental results of ALT, the random circuit, SEA with three random circuits as $U_j(i=1,2,3)$ and six CNOTs as entangling layer (i.e., SEA_RC), SEA with Schmidt coefficient layers as subSEA constructed by random circuits (i.e., SEA_Sch_RC), SEA with $R_\gamma(i)$ as $U_i$, and two ALTS as $U_3, U_5$ and 6 CNOTs as entangling layer (i.e., SEA_RC), SEA with Schmidt coefficient layer as subSEA constructed by ALTS (i.e., SEA_Sch_ALT), SEA with three ALTS as $U_i(i=1,2,3)$ and three CNOTs as entangling layer (i.e., SEA (ALT)), SEA with three ALTS as $U_i(i=1,2,3)$ and 6 CNOTs as entangling layer (i.e., SEA (ALT)), respectively. $J=3,6$ CNOTs means that we set a composition of $J$ CNOTs controlled and targeted on the qubit-pairs $(i, N+1)_{i\neq J}$.

FIGURE 10. Comparison of the scaling of variance between different ansatzes on the Heisenberg model. It shows the semi-log plot of the variance of the largest partial derivative among parameters in each round of sampling. We ensure different ansatzes have similar number of parameters by setting different depth. The solid part of the fitted lines represents the range we experimented with, while the dotted part represents the expected performance on a larger range.

\[
M_1 = \frac{t_1}{d} \int_{\mathcal{U}_{(d)}} (|0\rangle \otimes 2n(U_1^\dagger \otimes I_B) \cdot V^\dagger \cdot \rho_1^A \otimes I_B \\
\cdot V \cdot (U_1 \otimes I_B)(0) \otimes 2n d\eta(U))
\]

\[
M_2 = \frac{t_2}{d} \int_{\mathcal{U}_{(d)}} (|0\rangle \otimes 2n(U_1^\dagger \otimes I_B) \cdot V^\dagger \cdot I_A \otimes \rho_1^B \\
\cdot V \cdot (U_1 \otimes I_B)(0) \otimes 2n d\eta(U))
\]

\[
M_3 = t_3 \int_{\mathcal{U}_{(d)}} (|0\rangle \otimes 2n(U_1^\dagger \otimes I_B) \cdot V^\dagger \cdot I_{AB} \text{Tr}(\rho_1) \\
\cdot V \cdot (U_1 \otimes I_B)(0) \otimes 2n d\eta(U))
\]

Now the problem is how do we calculate $M_1$ and $M_2$. Combine with $V = \sum_{i=0}^{d-1} |i\rangle |i\rangle_A \otimes V_{B_i}$, there has

\[
M_1 = \frac{t_1}{d} \int_{\mathcal{U}_{(d)}} (\sum_{i=0}^{d-1} |i\rangle \otimes 2n(U_1^\dagger \otimes I_B)(0)) \otimes 2n d\eta(U)
\]

\[
M_2 = \frac{t_2}{d} \int_{\mathcal{U}_{(d)}} (\sum_{i=0}^{d-1} |i\rangle \otimes 2n(U_1^\dagger \otimes I_B)(0)) \otimes 2n d\eta(U)
\]
Therefore

\[ (110) = M_0 + M_1 + M_2 + M_3 \]

\[ = \frac{2d+6}{d^2(d+1)^3} \]

\[ \neq \frac{2}{d^2(d^2+1)} (n \geq 1, d = 2^n > 1) \]

\[ = (86). \]

Thus, SEA with \( U_i(i = 1, 2, 3) \) being local 2-design ansatizes and \( \text{CNOT} \) is not a unitary 2-design.

\( \square \)

E. SUPPLEMENTARY DESCRIPTION OF EXPERIMENTS

In this section, we present results from supplementary numerical experiments. Fig. 8 shows the specific structure of 14-qubit SEA_2 and SEA_3. Fig. 9 displays the results of VQE using the SEA of Schmidt coefficient layer as \( R_s(\theta)^{\otimes 6} \), entangling layer as six \text{CNOTs} and the LBC layers as two 6-qubit ALTs. Fig. 10 shows the variance of the largest partial derivative in each sample.

ACKNOWLEDGMENT

The authors thank Runyao Duan, Yin Mo, Chengkai Zhu, and Yuele Wang for their helpful discussions. The authors also thank the anonymous reviewers for their helpful suggestions, which helped us improve the manuscript. Part of this work was done when Xia Liu, Geng Liu, Hao-Kai Zhang, Jiaxin Huang, and Xin Wang were at Baidu Research.

REFERENCES

[1] J. Preskill, “Quantum computing and the entanglement frontier,” 2012, arXiv:1203.5813, doi: 10.48550/arXiv.1203.5813.

[2] S. Lloyd, “Universal quantum simulators,” Science, vol. 273, no. 5278, pp. 1073–1078, 1996, doi:10.1126/science.273.5278.1073.

[3] A. W. Harrow, A. Hassidim, and S. Lloyd, “Quantum algorithm for linear systems of equations,” Phys. Rev. Lett., vol. 103, no. 15, 2009, Art. no. 150502, doi:10.1103/PhysRevLett.103.150502.

[4] A. M. Childs and W. van Dam, “Quantum algorithms for algebraic problems,” Rev. Mod. Phys., vol. 82, no. 1, pp. 1–52, Jan. 2010, doi:10.1103/RevModPhys.82.1.

[5] A. Montanaro, “Quantum algorithms: An overview,” npj Quantum Inf., vol. 2, no. 1, Nov. 2016, Art. no. 1023, doi:10.1038/appi.jqiq.2015.23.

[6] J. Preskill, “Quantum computing in the NISQ era and beyond,” Quantum, vol. 2, 2018, Art. no. 79, doi:10.22331/q-2018-08-06-79.

[7] Y. Cao PD et al., “Quantum chemistry in the age of quantum computing,” Chem. Rev., vol. 119, no. 19, pp. 10856–10915, 2019, doi:10.1021/acs.chemrev.8b00830.

[8] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, “Quantum computational chemistry,” Rev. Mod. Phys., vol. 92, no. 1, 2020, Art. no. 015003, doi:10.1103/RevModPhys.92.015003.

[9] M. Schuld and F. Petruccione, Machine Learning With Quantum Computers, Cham, Switzerland: Springer, 2021, doi:10.1007/978-3-030-83098-4.

[10] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” Nature, vol. 549, no. 7671, pp. 195–202, Sep. 2017, doi:10.1038/nature23474.

[11] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, “Provably efficient machine learning for quantum many-body problems,” Science, vol. 377, no. 6613, 2021, Art. no. eabk3333, doi:10.1126/science.abk3333.

[12] S. Jerbi, L. J. Fiderer, H. P. Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, “Quantum machine learning beyond kernel methods,” Nature Commun., vol. 14, no. 1, 2021, Art. no. 517, doi:10.1038/s41467-023-36159-y.

[13] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, “Quantum circuit learning,” Phys. Rev. A, vol. 98, no. 3, Sep. 2018, Art. no. 032309, doi:10.1103/PhysRevA.98.032309.

[14] F. AruteGSL et al., “Quantum supremacy using a programmable superconducting processor,” Nature, vol. 574, no. 7779, pp. 505–510, 2019, doi:10.1038/s41586-019-1666-5.

[15] Y. Wu et al., “Strong quantum computational advantage using a superconducting quantum processor,” Phys. Rev. Lett., vol. 127, no. 18, Oct. 2021, Art. no. 180501, doi:10.1103/PhysRevLett.127.180501.

[16] H.-S. Zhong et al., “Quantum computational advantage using photons,” Science, vol. 370, no. 6523, pp. 1460–1463, Dec. 2020, doi:10.1126/science.abe8770.

[17] Y. Lü, Q. Gao, J. Lü, Y. Pan, and D. Dong, “Recent advances of quantum neural networks on the near term quantum processor,” 2020, arXiv:2206.03066, doi: 10.48550/arXiv.2206.03066.

[18] X. Xu, J. Sun, S. Endo, Y. Li, Simon C. Benjamin, and X. Yuan, “Variational algorithms for linear algebra,” Sci. Bull., vol. 66, no. 21, pp. 2181–2188, 2021, doi:10.1016/j.scib.2021.06.023.

[19] J. Liu et al., “Hybrid quantum-classical convolutional neural networks,” Sci. China Phys., Mech. Astron., vol. 64, no. 9, 2021, Art. no. 290311, doi:10.1007/s11433-021-1734-3.

[20] H.-S. Zhong et al., “Phase-programmable Gaussian boson sampling using simulated squeezed light,” Phys. Rev. Lett., vol. 127, no. 18, Oct. 2021, Art. no. 180502, doi:10.1103/PhysRevLett.127.180502.

[21] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, “The power of quantum neural networks,” Nature Comput. Sci., vol. 1, no. 6, pp. 403–409, Jun. 2021, doi:10.1038/s43588-021-00084-1.

[22] W. Li and D.-L. Deng, “Recent advances for quantum classifiers,” Sci. China Phys., Mech. Astron., vol. 65, no. 2, 2022, Art. no. 220301, doi:10.1007/s11433-021-1903-6.

[23] L. S. Madsen et al., “Quantum computational advantage with a programmable photonic processor,” Nature, vol. 606, no. 7912, pp. 75–81, 2022, doi:10.1038/s41586-022-04725-x.
H. L. Tang et al., “Qubit-adapt-VQE: An adaptive algorithm for constructing high-efficient ansatze on a quantum processor,” PRX Quantum, vol. 2, no. 2, 2021, Art. no. 020310, doi: 10.1038/S41570-020-02031-9.

I. G. Ryabinkin, R. A. Lang, S. N. Genin, and A. F. Izmaylov, “Iterative qubit coupled cluster approach with efficient screening of generators,” J. Chem. Theory Comput., vol. 16, no. 2, pp. 1055–1063, 2020, doi: 10.1021/acs.jctc.9b00184.

K. Bharti and T. Haug, “Iterative quantum-assisted eigensolver,” Phys. Rev. A, vol. 104, no. 4, Nov. 2021, Art. no. L050401, doi: 10.1103/PhysRevA.104.L050401.

H. R. Griemsley, G. S. Barron, E. Barnes, S. E. Economou, and N. J. Mayhall, “Adaptive, problem-tailored variational quantum eigensolver mitigates rough parameter landscapes and barren plateaus,” npj Quantum Inf., vol. 9, no. 1, p. 19, 2023.

T. Haug, K. Bharti, and M. S. Kim, “Capacity and quantum geometry of parametrized quantum circuits,” PRX Quantum, vol. 2, no. 4, 2021, Art. no. 045030, doi: 10.1038/s41570-021-00574-8.

Y. Du, Z. Tu, X. Yuan, and D. Tao, “Efficient measure for the expressivity of variational quantum algorithms,” Phys. Rev. Lett., vol. 128, no. 8, 2022, Art. no. 080506, doi: 10.1103/PhysRevLett.128.080506.

Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, “Connecting ansatz expressibility to gradient magnitudes and barren plateaus,” PRX Quantum, vol. 3, no. 1, 2022, Art. no. 010313, doi: 10.1038/s41570-021-00583-6.

J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, “Barren plateaus in quantum neural network training landscapes,” Nature Commun., vol. 9, no. 1, pp. 1–6, 2018, doi: 10.1038/s41467-018-07090-4.

K. Zhang, Min-Hsiu Hsieh, L. Liu, and D. Tao, “Toward trainability of deep quantum neural networks,” 2021, arXiv:2112.15002.

H.-K. Zhang, C. Zhu, G. Liu, and X. Wang, “Exponential hardness of optimization from the locality in quantum neural networks,” in Proc. AAAI Conf. Artif. Intell., vol. 38, no. 15, 2024, pp. 16741–16749.

S. Sim, P. D. Johnson, and A. Aspuru-Guzik, “Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms,” Adv. Quantum Technol., vol. 2, no. 12, 2019, Art. no. 1900070, doi: 10.1002/qute.201900070.

K. Nakaji and N. Yamamoto, “Expressibility of the alternating layered ansatz for quantum circuits,” Quantum, vol. 5, 2021, Art. no. 434, doi: 10.22331/q-2021-04-19-434.

K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, “Trainability of dissipative perceptron-based quantum neural networks,” Phys. Rev. Lett., vol. 128, no. 18, May 2022, Art. no. 180505, doi: 10.1103/PhysRevLett.128.180505.

Carlos Ortiz Marrero, Mária Kieferová, and N. Wiebe, “Entanglement-induced barren plateaus,” PRX Quantum, vol. 2, no. 4, Oct. 2021, Art. no. 040316, doi: 10.1038/s41570-021-00482-5.

E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, “An initialization strategy for addressing barren plateaus in parametrized quantum circuits,” Quantum, vol. 3, Mar. 2019, Art. no. 214, doi: 10.22331/q-2019-12-09-214.

G. Verdon et al., “Learning to learn with quantum neural networks via classical neural networks,” Jul. 2019, arXiv:1907.05415, doi: 10.48550/arXiv.1907.05415.

H.-Y. Liu, T.-P. Sun, Y.-C. Wu, Y.-J. Han, and G.-P. Guo, “Mitigating barren plateaus with transfer-learning-inspired parameter initialization,” New J. Phys., vol. 25, no. 1, 2023, Art. no. 013039, doi: 10.1088/1367-2630/acb8c6.

M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, “Cost function dependency of barren plateaus in shallow parametrized quantum circuits,” Nature Commun., vol. 12, no. 1, Mar. 2021, Art. no. 1791, doi: 10.1038/s41467-021-21728-w.

M. Kieferová, O. M. Carlos, and N. Wiebe, “Quantum generative training using Rényi divergences,” 2021, arXiv:2106.09567, doi: 10.48550/arXiv.2106.09567.

A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, “Absence of barren plateaus in quantum convolutional neural networks,” Phys. Rev. X, vol. 11, no. 4, Oct. 2021, Art. no. 041011, doi: 10.1103/PhysRevX.11.041011.

M. Cerezo et al., “Does provable absence of barren plateaus imply classical simulatability? Or, why we need to rethink variational quantum computing,” 2023, arXiv:2312.09121, doi: 10.48550/arXiv.2312.09121.

M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, Cambridge, U.K.: Cambridge Univ. Press, 2010, doi: 10.1017/CBO9780511976667.
Jiaxin Huang received the B.Sc. degree in computer science and mathematics from the Hong Kong University of Science and Technology, Hong Kong, in 2023. He is currently working toward the Ph.D. degree in quantum computation with the University of Hong Kong, Hong Kong. He was an Intern Researcher with Baidu Inc., Beijing, China, when participating in this project. After that, he worked as a Research Assistant with the Institute for Quantum Computing, University of Waterloo, Waterloo, ON, Canada. His research interests include quantum machine learning, quantum error correction, and fault tolerance.

Xin Wang (Senior Member, IEEE) received the doctorate degree in quantum information from the University of Technology Sydney, Ultimo, NSW, Australia, in 2018. From 2018 to 2019, he was a Hartree Postdoctoral Fellow with the Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, MD, USA, and from 2019 to 2023, a Staff Researcher with the Institute for Quantum Computing, at Baidu Research, Beijing, China. He is currently an Associate Professor with the Thrust of Artificial Intelligence, Information Hub, The Hong Kong University of Science and Technology, Guangzhou, China. He is also an Editor of Quantum. His research interests include broad range of perspectives of quantum information science, including quantum communication, entanglement theory, quantum software, near-term quantum algorithms, and quantum machine learning.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.