

Expressibility of the alternating layered ansatz for quantum computation

Kouhei Nakaji and Naoki Yamamoto

Department of Applied Physics and Physico-Informatics & Quantum Computing Center, Keio University, Hiyoshi 3-14-1, Kohoku, Yokohama, 223-8522, Japan

The hybrid quantum-classical algorithm is actively examined as a technique applicable even to intermediate-scale quantum computers. To execute this algorithm, the hardware efficient ansatz is often used, thanks to its implementability and expressibility; however, this ansatz has a critical issue in its trainability in the sense that it generically suffers from the so-called gradient vanishing problem. This issue can be resolved by limiting the circuit to the class of shallow alternating layered ansatz. However, even though the high trainability of this ansatz is proved, it is still unclear whether it has rich expressibility in state generation. In this paper, with a proper definition of the expressibility found in the literature, we show that the shallow alternating layered ansatz has almost the same level of expressibility as that of hardware efficient ansatz. Hence the expressibility and the trainability can coexist, giving a new designing method for quantum circuits in the intermediate-scale quantum computing era.

1 Introduction

Recent rapid progress in the hardwares for quantum computing stimulates researchers to develop new techniques that utilize even noisy intermediate-scale quantum (NISQ) [1] device for real applications such as machine learning and quantum chemistry. Especially, several hybrid quantum-classical algorithms have been actively examined, as a means to reduce the computational cost required on quantum computing part. The variational quantum eigensolver (VQE) [2], which trains the parameterized quantum circuit via a classical computation to decrease a given cost function, is typically considered as such a hybrid algorithm.

The critical point in this strategy is in the difficulty to design a suitable and implementable circuit ansatz that, after the training process, may produce an exact or well-approximating solution of a given problem. The hardware efficient ansatz (HEA) [3] is often used mainly because of the implementability on a hardware; this is a relatively shallow circuit ansatz, whose parameters are embedded in the angles of single-qubit rotation gates. However, it was proven in [4] that, when those parameters are randomly chosen, the gradient of a standard cost function vanishes, meaning that the update of the cost function often gets stuck in the learning process before reaching the minimum.

For solving this vanishing gradient problem, several approaches have been proposed. Reference [5] showed numerically that, with a special type of initializing method of the HEA, the vanishing gradient problem does not occur in an ansatz with $\mathcal{O}(1)$ qubits. Also, a quantum analogue of the natural gradient have been proposed in [6, 7], the original classical version of which is often used to avoid similar vanishing gradient problems in neural networks. In this paper, we focus on the third approach given by Ref. [8] that provides a method for devising a specific structure of the HEA ansatz, called the Alternating Layered Ansatz (ALT), which in fact provably does not suffer from the vanishing gradient problem. By definition, the class of ALT is included in that of HEA; the difference is that, while a HEA consists of multiple layers of single-qubit rotation gates and entanglers that in principle combines *all* qubits in each layer, the entangling gates contained in an ALT is restricted to entangle only *local* qubits in each layer. With this setting, the authors in [8] derived a strict lower bound of the variance of the gradient for an ALT with its parameters randomly chosen (more precisely, the ensemble of unitary matrices corresponding to each circuit block is 2-design) under the condition that the cost function is local (that is, the cost function is composed of local functions of only a small number of local qubits). By using this lower bound, it was also shown that the vanishing gradient problem can be resolved if the number of layers is of the order $\mathcal{O}(\text{poly}(\log n))$ where n is the total number of qubits, or roughly speaking if the circuit is shallow.

Then an important question arises; does ALT have a sufficient expressive power (expressibility) for generating a rich class of states, which contains the optimal or a well-approximating state? Because the set of ALTs is a subclass of that of HEA, one might argue that the expressibility of ALT could be much lower than that of the HEA; if this is the case, the ALT circuit may not generate a desired state even though the learning process is smoothly running. Thus, it is worth examining the expressibility of ALT in order to assess the practicality of this ansatz in executing the hybrid quantum-classical algorithm. In this paper, we study this problem using the expressibility measure introduced in [9] and show that, fortunately, the class of shallow ALTs has the same level of expressibility as that of HEAs. Therefore, the expressibility and the trainability of a quantum circuit can coexist, which means that the existing HEA found in the literature can be basically replaced with a simpler ALT without degradation of expressibility while acquiring a better trainability. That is, the ALT might be taken as a new standard ansatz in NISQ computing era.

The structure of this paper is as follows. Section 2 is the preliminary, giving the definition of expressibility and the ansatzes. In Sec. 3, we show both theoretically and numerically that the expressibility of ALT is as high as that of HEA. In Sec. 4, we show how much the value of expressibility is reflected to the result of VQE. Finally, we conclude with some remarks in Sec. 5.

2 Preliminaries

In this section, we define indicators of the expressibility and introduce some circuit ansatzes.

2.1 Indicators of the expressibility

Following Ref. [9], we define the expressibility of a given circuit, by the randomness of states generated from the circuit, in terms of the frame potential and the Kullback-Leibler (KL) divergence.

Frame Potential

To define the expressibility of a given circuit ansatz C , let us consider the deviation of the state distribution generated by C from the Haar distribution, as follows;

$$\mathcal{A}^{(t)}(C) = \left\| \int_{\text{Haar}} (|\psi\rangle\langle\psi|)^{\otimes t} d\psi - \int_{\Theta} (|\psi_{\theta}\rangle\langle\psi_{\theta}|)^{\otimes t} d\theta \right\|_{HS},$$

where \int_{Haar} denotes the integration over the state $|\psi\rangle$ distributed with respect to the Haar measure, and $\|\cdot\|_{HS}$ is the Hilbert Schmidt distance. Also, $|\psi_{\theta}\rangle$ is the state generated by the ansatz C characterized by the parameter $\theta \in \Theta$, e.g., $|\psi_{\theta}\rangle = U_C(\theta)|0\rangle$, $\theta \in \Theta$, where $U_C(\theta)$ is the unitary operator corresponding to C and $|0\rangle$ is an initial state. Then we call that the ansatz C with smaller $\mathcal{A}^{(t)}(C)$ has a higher expressibility. This definition is justified by the following reason. That is, because the state $|\psi\rangle$ generated from the Haar distribution can in principle represent an arbitrary state, the condition $\mathcal{A}^{(t)}(C) \approx 0$ implies that the ansatz C can generate almost all states possibly including the optimal solution (e.g., the ground state in VQE); also, in this case the states generated from C are almost equally distributed, which is particularly favorable if little is known about the problem.

To compute $\mathcal{A}^{(t)}(C)$, we instead focus on the following t -th generalized frame potential [10] of C :

$$\mathcal{F}^{(t)}(C) = \int_{\Phi} \int_{\Theta} |\langle\psi_{\phi}|\psi_{\theta}\rangle|^{2t} d\phi d\theta, \quad (1)$$

where both Θ and Φ represent the same set of parameters of C . In the present paper, we simply call it the t -th frame potential. In particular, the t -th frame potential of N -dimensional states distributed with respect to the Haar measure, is given by $\mathcal{F}_{\text{Haar}}^{(t)}(N) = \int_{\text{Haar}} \int_{\text{Haar}} |\langle\psi|\psi'\rangle|^{2t} d\psi d\psi'$. The point of introducing the frame potential is because these quantities are linked to $\mathcal{A}^{(t)}(C)$ in the following form; that is, for an arbitrary positive integer t , it holds

$$\mathcal{F}^{(t)}(C) - \mathcal{F}_{\text{Haar}}^{(t)}(N) = \mathcal{A}^{(t)}(C) \geq 0. \quad (2)$$

The equality in the last inequality holds if and only if the ensemble of $|\psi_{\theta}\rangle$ is a state t -design [11–13]. Thus, the ansatz C with smaller $\mathcal{F}^{(t)}(C)$ has a higher expressibility. Also $\mathcal{F}^{(t)}(C)$ is lower bounded

by $\mathcal{F}_{\text{Haar}}^{(t)}(N)$, meaning that the frame potential can be used as an indicator for quantifying the non-uniformity in the state distribution. In Sec. 3.1, we calculate $\mathcal{F}^{(1)}(C)$ and $\mathcal{F}^{(2)}(C)$ for several ansatz C .

KL-Divergence

Note that the frame potential (1) is the t -th moment of the fidelity $F = |\langle \psi_\theta | \psi_\phi \rangle|^2$, where the circuit parameters $\theta \in \Theta$ and $\phi \in \Phi$ are randomly sampled from the circuit ansatz C . Hence the probability distribution of F , denoted by $P(C, F)$, contains more information for quantifying the randomness of C than $\mathcal{F}^{(t)}(C)$, and thereby the following measure can be used to quantify the expressibility of C :

$$\mathcal{E}(C) = D_{\text{KL}}(P(C, F) \| P_{\text{Haar}}(F)) = \int_0^1 P(C, F) \log \frac{P(C, F)}{P_{\text{Haar}}(F)} dF, \quad (3)$$

where $D_{\text{KL}}(q \| p)$ is the KL divergence between q and p . Also $P_{\text{Haar}}(F)$ is the probability distribution of the fidelity $F = |\langle \psi | \psi' \rangle|^2$, where $|\psi\rangle$ and $|\psi'\rangle$ are sampled according to the Haar measure. In Ref. [14], $P_{\text{Haar}}(F) = (N-1)(1-F)^{N-2}$ is derived, where N is the dimension of Hilbert space. Because in general $D_{\text{KL}}(q \| p) = 0$ iff $q = p$, the ansatz C with smaller value of $\mathcal{E}(C)$ has a higher expressibility. Thus, $\mathcal{E}(C)$ can also be used as an indicator for quantifying the non-uniformity of an ansatz.

Lastly recall that the t -th moment of $P(C, F)$ and $P_{\text{Haar}}(F)$ are $\mathcal{F}^{(t)}(C)$ and $\mathcal{F}_{\text{Haar}}^{(t)}(N)$, respectively. Thus, if the values of $\mathcal{F}^{(t)}(C)$ is close to $\mathcal{F}_{\text{Haar}}^{(t)}(N)$, the value of $\mathcal{E}(C)$ is close to zero.

2.2 Ansatzes

Here we describe the three types of ansatzes investigated in this paper.

Hardware Efficient Ansatz

The HEA circuit consists of multiple layers of parametrized single qubit gates and entanglers which entangle all qubits. In the following, let $C_{\text{HEA}}^{\ell, n}$ be the class of HEA with ℓ layers where each layer contains n -qubits.

Alternating Layered Ansatz

The ALT introduced in [8] also consists of multiple layers but the components of each layer are different from those of the HEA as follows. That is, each layer has some separated blocks, where each block has parametrized single-qubit rotation gates and fixed entanglers that entangle all qubits inside the block. The probability distributions of those angle parameters are independent in all blocks and all layers. In this paper, we further restrict the class of ALT as follows. First, as in HEA, the entire circuit is composed of ℓ layers, where each layer contains n qubits. Then, we assume that, in the odd-number-labeled layers, each block contains m qubits, so that m is an even number and n/m is an integer. In other words, the odd-number-labeled layers contain n/m blocks which operate on $\{1, \dots, m\}$, $\{m+1, \dots, 2m\}$, ..., and $\{n-m+1, \dots, n\}$ qubits. As for the even-number-labeled layers, they contain $n/m + 1$ blocks which operate on $\{1, \dots, m/2\}$, $\{m/2+1, \dots, 3m/2\}$, ..., and $\{n-m/2+1, \dots, n\}$ qubits; that is, the first and the last block operate on $m/2$ qubits, while the others operate on m qubits. In the following, we use $C_{\text{ALT}}^{\ell, m, n}$ to denote the class of ALT with the above-defined indices.

In [8], it is proved that the vanishing gradient problem can be avoided if the following conditions are satisfied; (i) each term of the cost Hamiltonian, $H = \sum_k H_k$ is local, meaning that each term H_k is composed of less than m neighboring qubits, (ii) the ensemble of unitary matrices in each block is 2-design, and (iii) the number of layers, ℓ , is of the order $O(\text{poly}(\log n))$. Note that each block needs to be deep enough for the distribution of corresponding unitary matrices to be close to 2-design.

Tensor Product Ansatz

In addition to HEA and ALT, we here introduce the class of tensor product ansatz (TEN) as a relatively weak ansatz. This ansatz also consists of ℓ layers, where each layer contains n qubits, and each layer contains n/m blocks (n/m is assumed to be an integer), which contain single-qubit

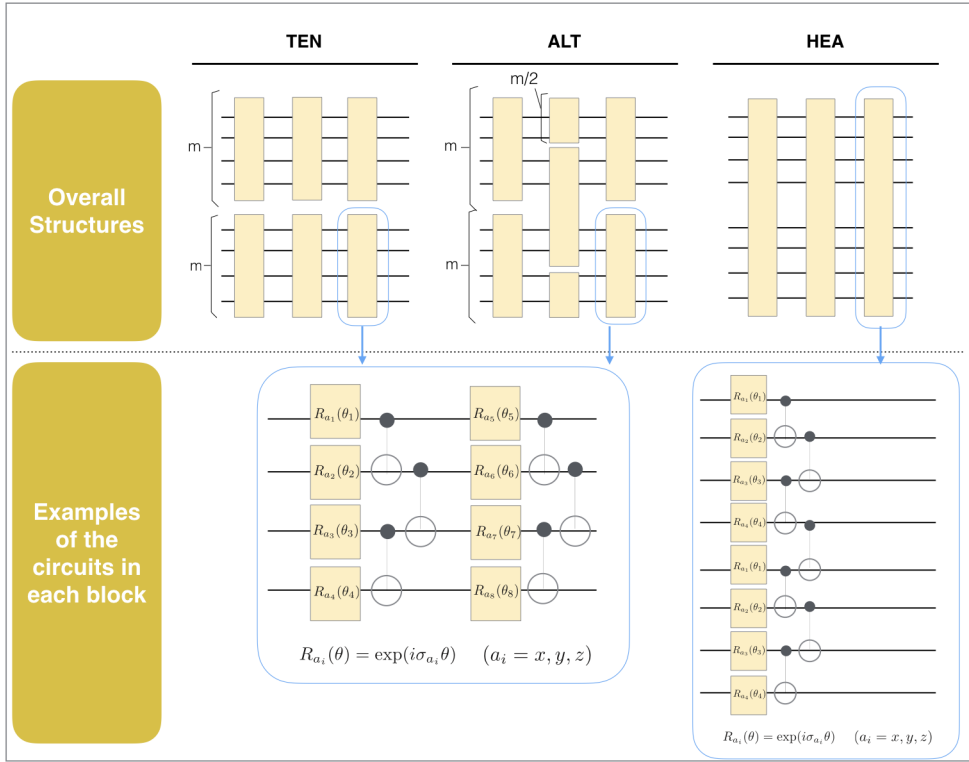


Figure 1: Examples of TEN, ALT, and HEA introduced in Section 2.2. The upper figures are the overall structures of the ansatzes in the case $n = 8$ and $\ell = 3$ (and $m = 4$ for TEN and ALT). The circuits inside the blocks are exemplified in the lower figure, where σ_{a_i} ($a_i \in \{x, y, z\}$) is the Pauli matrix operating on a qubit and $\theta_i \in [0, 2\pi]$ is the parameter.

rotation gates and entanglers combining all qubits in the block. Throughout all the layers, the blocks operate on $\{1, \dots, m\}$, $\{m+1, \dots, 2m\}$, ..., and $\{n-m+1, \dots, n\}$ qubits. Thus, TEN always generates a product state of the form $|\psi_1\rangle \otimes \dots \otimes |\psi_{n/m}\rangle$ where each state is composed of m qubits. Let $C_{\text{TEN}}^{\ell, m, n}$ be the class of TEN described above.

In Fig. 1, we show examples of the structures of these ansatzes.

3 Expressibility of the circuit ansatzes

In this section, we give some analytical expressions as well as upper bounds of the first and the second frame potentials of the three ansatzes introduced in Section 2.2, showing that the shallow ALT has almost the same expressibility as that of HEA. This result will be further confirmed by a numerical simulation in terms of the KL-divergence.

3.1 Analytical expression of the frame potential of the ansatzes

First of all, to compute the frame potentials of each ansatz, we assume that the ensemble of the unitary matrices corresponding to $C_{\text{HEA}}^{\ell, n}$ is 2-design. Similarly, the ensemble of the unitary matrices corresponding to each block of $C_{\text{ALT}}^{\ell, m, n}$ and $C_{\text{TEN}}^{\ell, m, n}$ are assumed to be 2-design. These assumptions are also adopted in the discussion of [4] and [8]. Such ensemble of unitary matrices can be generated by randomly choosing the parameters of the circuit having a specific structure.

Before going into the detail, we show some integration formulae for random unitary matrices [15]. First, if the ensemble of $n \times n$ unitary matrices $\{U\}$ is 1-design, the following formula holds:

$$\int_{\text{1-design}} dU U_{ij} U_{mk}^* = \frac{\delta_{im} \delta_{jk}}{n}, \quad (4)$$

where $\int_{1\text{-design}} dU$ is the integral over the 1-design ensemble of the unitary matrices. Second, if the ensemble of $n \times n$ unitary matrices $\{U\}$ is 2-design, the following formula holds:

$$\int_{2\text{-design}} dU U_{i_1 j_1} U_{i_2 j_2} U_{i'_1 j'_1}^* U_{i'_2 j'_2}^* = \sum_{k=1}^4 \lambda_k^{(n)} \Delta_{i_1 j_1 i_2 j_2 i'_1 j'_1 i'_2 j'_2}^k, \quad (5)$$

where

$$\begin{aligned} \lambda_1^{(n)} &= \lambda_2^{(n)} = \frac{1}{2^{2n} - 1}, \quad \lambda_3^{(n)} = \lambda_4^{(n)} = -\frac{1}{(2^{2n} - 1)2^n}, \\ \Delta_{i_1 j_1 i_2 j_2 i'_1 j'_1 i'_2 j'_2}^1 &= \delta_{i_1 i'_1} \delta_{j_1 j'_1} \delta_{i_2 i'_2} \delta_{j_2 j'_2}, \\ \Delta_{i_1 j_1 i_2 j_2 i'_1 j'_1 i'_2 j'_2}^2 &= \delta_{i_1 i'_2} \delta_{j_1 j'_2} \delta_{i_2 i'_1} \delta_{j_2 j'_1}, \\ \Delta_{i_1 j_1 i_2 j_2 i'_1 j'_1 i'_2 j'_2}^3 &= \delta_{i_1 i'_1} \delta_{j_1 j'_2} \delta_{i_2 i'_2} \delta_{j_2 j'_1}, \\ \Delta_{i_1 j_1 i_2 j_2 i'_1 j'_1 i'_2 j'_2}^4 &= \delta_{i_1 i'_2} \delta_{j_1 j'_1} \delta_{i_2 i'_1} \delta_{j_2 j'_2}, \end{aligned} \quad (6)$$

and $\int_{2\text{-design}} dU$ is the integral over the 2-design ensemble of the unitary matrices. These formulae are effectively used to derive the theorems shown below.

The First Frame Potential

For the first frame potentials, the following theorem holds.

Theorem 1. *If the ensemble of the unitary matrices corresponding to $C_{\text{HEA}}^{\ell,n}$ and the ensemble of the unitary matrices corresponding to each block of $C_{\text{ALT}}^{\ell,m,n}$ and $C_{\text{TEN}}^{\ell,m,n}$ are 2-design, then the following equalities hold:*

$$\mathcal{F}^{(1)}(C_{\text{HEA}}^{\ell,n}) = \mathcal{F}^{(1)}(C_{\text{ALT}}^{\ell,m,n}) = \mathcal{F}^{(1)}(C_{\text{TEN}}^{\ell,m,n}) = \mathcal{F}_{\text{Haar}}^{(1)}(2^n). \quad (7)$$

The equality, $\mathcal{F}^{(1)}(C_{\text{HEA}}^{\ell,n}) = \mathcal{F}_{\text{Haar}}^{(1)}(2^n)$, can be readily proved from the assumption of the theorem, because, if the ensemble of the unitary matrices corresponding to the circuit is 2-design, the ensemble of the states generated by the circuit is a state 2-design (and therefore a state 1-design). For the other equalities, the proof is given in Appendix. Note that, accordingly, the ensembles of the states generated by $C_{\text{HEA}}^{\ell,n}$, $C_{\text{ALT}}^{\ell,m,n}$, and $C_{\text{TEN}}^{\ell,m,n}$ are all 1-design.

The Second Frame Potential

The second frame potential of the Haar random circuits, $\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$, can be computed as

$$\mathcal{F}_{\text{Haar}}^{(2)}(2^n) = \int_0^1 dF F^2 (2^n - 1)(1 - F)^{2^n - 2} = \frac{1}{(2^n + 1)2^{n-1}}. \quad (8)$$

Then, for $C_{\text{HEA}}^{\ell,n}$ and $C_{\text{TEN}}^{\ell,m,n}$, the following theorem holds.

Theorem 2. *If the ensemble of the unitary matrices corresponding to $C_{\text{HEA}}^{\ell,n}$ and the ensemble of the unitary matrices corresponding to each block of $C_{\text{TEN}}^{\ell,m,n}$ are 2-design, then the following equalities hold:*

$$\mathcal{F}^{(2)}(C_{\text{HEA}}^{\ell,n}) = \mathcal{F}_{\text{Haar}}^{(2)}(2^n), \quad (9)$$

$$\mathcal{F}^{(2)}(C_{\text{TEN}}^{\ell,m,n}) = 2^{\frac{n}{m}-1} \cdot \frac{2^n + 1}{(2^m + 1)^{\frac{n}{m}}} \mathcal{F}_{\text{Haar}}^{(2)}(2^n). \quad (10)$$

The equality (9) is readily proved from the assumption of the theorem, because, as we mentioned above, if the ensemble of the unitary matrices corresponding to the circuit is 2-design, the ensemble of the states generated by the circuit is a state 2-design. For Eq.(10), we give the proof in Appendix. From this theorem we find that $\mathcal{F}^{(2)}(C_{\text{TEN}}^{\ell,m,n})$ is always larger than $\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$; in particular, $\mathcal{F}^{(2)}(C_{\text{TEN}}^{\ell,m,n}) \simeq 2^{n/m-1} \mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ for large n , meaning that the expressibility of TEN is much smaller than that of HEA in the sense of the frame potential.

As for ALT, it is difficult to obtain an explicit formula like the case of HEA and TEN. Hence in Theorem 3 below, we provide a formula for computing the values of $\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n})$ and $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n})$; the computation methods for the other ℓ s are left for future work. Before stating the theorem, below we define a 16-dimensional vector $\mathbf{a}(2, m)$, a 16×16 matrix $B(2, m)$, a 64-dimensional vector $\mathbf{a}(3, m)$, and a 64×64 matrix $B(3, m)$. Given integers $k_a, k_b \in \{1, 2, 3, 4\}$, the $(4(k_a - 1) + k_b)$ -th component of the vector $\mathbf{a}(2, m)$ is defined as

$$\mathbf{a}(2, m)_{4(k_a-1)+k_b} = \int_{2\text{design}} dP dQ \sqrt{\lambda_{k_a}^{(m)} \lambda_{k_b}^{(m)}} \Delta^{(k_a, k_b)}(P, Q), \quad (11)$$

where $\Delta^{(k_a, k_b)}(P, Q)$ is the function of $m/2 \times m/2$ unitary matrices P and Q :

$$\Delta^{(k_a, k_b)}(P, Q) = \sum_{\substack{uu'ii' \\ jj'll'}} \sum_{pp'qq'} \Delta_{u0u'0i0i'0}^{k_a} \Delta_{p0p'0q0q'0}^{k_b} P_{ju} P_{j'u'} P_{li}^* P_{l'i'}^* Q_{lp} Q_{l'p'} Q_{jq}^* Q_{j'q'}^*. \quad (12)$$

Next, given integers $k_a, k_b, k_c \in \{1, 2, 3, 4\}$, the $(16(k_a - 1) + 4(k_b - 1) + k_c)$ -th component of the vector $\mathbf{a}(3, m)$ is defined as

$$\mathbf{a}(3, m)_{16(k_a-1)+4(k_b-1)+k_c} = \int_{2\text{design}} dP dQ \sqrt{\lambda_{k_a}^{(m)} \lambda_{k_b}^{(m)} \lambda_{k_c}^{(m)}} \Delta^{(k_a, k_b, k_c)}(P, Q), \quad (13)$$

where $\Delta^{(k_a, k_b, k_c)}(P, Q)$ is a function of $m/2 \times m/2$ unitary matrices P and Q :

$$\Delta^{(k_a, k_b, k_c)}(P, Q) = \sum_{\substack{uu'ii' \\ jj'll'}} \sum_{rr'tt' \\ pp'qq'} \Delta_{u0u'0i0i'0}^{k_a} \Delta_{jl'j'l'tr't'r'}^{k_b} \Delta_{p0p'0q0q'0}^{k_c} P_{tu} P_{t'u'} P_{ji}^* P_{j'i'}^* Q_{lp} Q_{l'p'} Q_{rq}^* Q_{r'q'}^*. \quad (14)$$

Also, given integers $k_a, k_b, k_c, k_d \in \{1, 2, 3, 4\}$, the $(4(k_a - 1) + 4k_b, 4(k_c - 1) + k_d)$ -th component of the matrix $B(m, 2)$ is defined as

$$B(2, m)_{4(k_a-1)+k_b, 4(k_c-1)+k_d} = \int_{2\text{design}} dP dQ \sqrt{\lambda_{k_a}^{(m)} \lambda_{k_b}^{(m)}} \sqrt{\lambda_{k_c}^{(m)} \lambda_{k_d}^{(m)}} \Delta^{(k_a, k_b, k_c, k_d)}(P, Q), \quad (15)$$

where $\Delta^{(k_a, k_b, k_c, k_d)}(P, Q)$ is a function of $m \times m$ matrices P and Q :

$$\begin{aligned} \Delta^{(k_a, k_b, k_c, k_d)}(P, Q) = & \sum_{\substack{u_2 u'_2 i_2 i'_2 \\ j_2 j'_2 l_2 l'_2}} \sum_{p_2 p'_2 q_2 q'_2} \sum_{\substack{u_3 u'_3 i_3 i'_3 \\ j_3 j'_3 l_3 l'_3}} \sum_{p_3 p'_3 q_3 q'_3} \Delta_{u_2 0 u'_2 0 i_2 0 i'_2 0}^{k_a} \Delta_{p_2 0 p'_2 0 q_2 0 q'_2 0}^{k_b} \Delta_{u_3 0 u'_3 0 i_3 0 i'_3 0}^{k_c} \Delta_{p_3 0 p'_3 0 q_3 0 q'_3 0}^{k_d} \\ & \times P_{j_3 u_3}^{j_3 u_3} P_{j'_3 u'_3}^{j'_3 u'_3} P_{l_2 i_2}^{* l_3 i_3} P_{l'_2 i'_2}^{* l'_3 i'_3} Q_{l_2 p_2}^{l_3 p_3} Q_{l'_2 p'_2}^{l'_3 p'_3} Q_{j_2 q_2}^{* j_3 q_3} Q_{j'_2 q'_2}^{* j'_3 q'_3} \end{aligned} \quad (16)$$

For the matrix component $M_{i,j}^{s,t}$, the upper indices correspond to the first $m/2$ qubits and the lower indices correspond to the last $m/2$. Given integers $k_a, k_b, k_c, k_d, k_e, k_f \in \{1, 2, 3, 4\}$, the $(16(k_a - 1) + 4(k_b - 1) + k_c, 16(k_d - 1) + 4(k_e - 1) + k_f)$ -th component of the matrix $B(m, 3)$ is defined as

$$\begin{aligned} B(3, m)_{16(k_a-1)+4(k_b-1)+k_c, 16(k_d-1)+4(k_e-1)+k_f} \\ = \int_{2\text{design}} dP dQ \sqrt{\lambda_{k_a}^{(m)} \lambda_{k_b}^{(m)} \lambda_{k_c}^{(m)}} \sqrt{\lambda_{k_d}^{(m)} \lambda_{k_e}^{(m)} \lambda_{k_f}^{(m)}} \Delta^{(k_a, k_b, k_c, k_d, k_e, k_f)}(P, Q), \end{aligned} \quad (17)$$

where $\Delta^{(k_a, k_b, k_c, k_d, k_e, k_f)}(P, Q)$ is a function of $m \times m$ matrices P and Q :

$$\begin{aligned} \Delta^{(k_a, k_b, k_c, k_d, k_e, k_f)}(P, Q) = & \sum_{\substack{u_2 u'_2 i_2 i'_2 \\ j_2 j'_2 l_2 l'_2}} \sum_{\substack{r_2 r'_2 t_2 t'_2 \\ p_2 p'_2 q_2 q'_2}} \sum_{\substack{u_3 u'_3 i_3 i'_3 \\ j_3 j'_3 l_3 l'_3}} \sum_{\substack{r_3 r'_3 t_3 t'_3 \\ p_3 p'_3 q_3 q'_3}} \Delta_{u_2 0 u'_2 0 i_2 0 i'_2 0}^{k_{11}^a} \Delta_{j_2 l_2 j'_2 l'_2 t_2 r_2 t'_2 r'_2}^{k_{31}^a} \Delta_{p_2 0 p'_2 q_2 0 q'_2 0}^{k_{11}^b} \\ & \times \Delta_{u_3 0 u'_3 0 i_3 0 i'_3 0}^{k_{12}^a} \Delta_{j_3 l_3 j'_3 l'_3 t_3 r_3 t'_3 r'_3}^{k_{32}^a} \Delta_{p_3 0 p'_3 q_3 0 q'_3 0}^{k_{12}^b} P_{t_2 u_2}^{t_3 u_3} P_{t'_2 u'_2}^{t'_3 u'_3} P_{j_2 i_2}^{* j_3 i_3} P_{j'_2 i'_2}^{* j'_3 i'_3} \\ & \times Q_{l_2 p_2}^{l_3 p_3} Q_{l'_2 p'_2}^{l'_3 p'_3} Q_{r_2 q_2}^{* r_3 q_3} Q_{r'_2 q'_2}^{* r'_3 q'_3}. \end{aligned} \quad (18)$$

Now we can give the theorem as follows (the proof is given in Appendix).

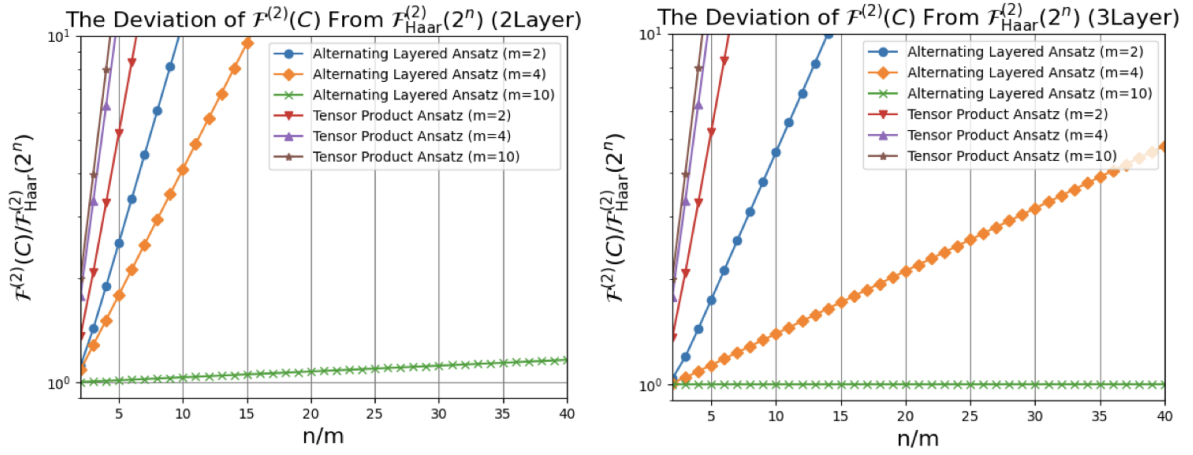


Figure 2: The values of $\mathcal{F}^{(2)}(C)/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ for ALT and TEN. The left and right figures show the case $\ell = 2$ and $\ell = 3$, respectively. The vertical axes are in the logarithmic scale. If $\mathcal{F}^{(2)}(C)/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ is close to 1, this means that the expressibility of the ansatz C is relatively high.

Theorem 3. *If the ensemble of the unitary matrices corresponding to each block of $C_{\text{ALT}}^{2,m,n}$ and $C_{\text{ALT}}^{3,m,n}$ are 2-design, then we have*

$$\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n}) = \mathbf{a}(2, m)^T B(2, m)^{\frac{n}{m}-1} \mathbf{a}(2, m), \quad (19)$$

$$\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) = \mathbf{a}(3, m)^T B(3, m)^{\frac{n}{m}-1} \mathbf{a}(3, m). \quad (20)$$

The vectors $\mathbf{a}(\ell, m)$ and the matrices $B(\ell, m)$ for $\ell = 2, 3$ are obtained by directly computing Eqs. (11), (13), (15), and (17), which then lead to $\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n})$ and $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n})$. Now our interest is in the gap of these quantities from $\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$, to see the expressibility of ALT. For this purpose, in Fig. 2 we show $\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n})/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ and $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n})/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ as a function of n/m , for several values of (m, n) . For comparison, $\mathcal{F}^{(2)}(C_{\text{TEN}}^{2,m,n})/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ and $\mathcal{F}^{(2)}(C_{\text{TEN}}^{3,m,n})/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ are shown in the figure. Recall that, if this measure takes a smaller value, this means that the corresponding ansatz has a higher expressibility. Here is the list of notable points:

- For any pair of (n, m) , it is clear that $\mathcal{F}^{(2)}(C_{\text{ALT}}^{\ell,m,n})$ is much smaller than $\mathcal{F}^{(2)}(C_{\text{TEN}}^{\ell,m,n})$ for both $\ell = 2, 3$. This means that, as expected, ALT has a much higher expressibility than TEN.
- For any pair of (n, m) , $\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n}) > \mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n})$ hold, i.e., as ℓ increases, the expressibility increases.
- For any fixed n/m , the ALT with bigger m always has a higher expressibility. For instance, the ALT with $(n, m) = (50, 10)$ has a higher expressibility than the ALT with $(n, m) = (20, 4)$. This is simply because, if the structure of the circuit (the number of division in each layer for making the block) is the same, then an ALT with bigger block components has a higher expressibility.
- For a fixed n , we have ALT with the smaller second order frame potential by taking m bigger. For instance $n = 100$, we have $\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,2,100}) > \mathcal{F}^{(2)}(C_{\text{ALT}}^{2,4,100}) > \mathcal{F}^{(2)}(C_{\text{ALT}}^{2,10,100})$. That is, for a limited number of available qubits, the ALT with less blocks has a higher expressibility.
- $\mathcal{F}^{(2)}(C_{\text{ALT}}^{\ell,m,n}) \simeq \mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ when $m = 10$ for all n/m within the figure and for both $\ell = 2, 3$. Hence the ALT composed from the blocks with $m = 10$ qubits in each layer has almost the same expressibility as the HEA without respect to the total qubits number, n . In other words, for a given HEA with fixed n , we can divide each layer into separated 10-qubits blocks to make an ALT, without decreasing the expressibility.

The last point is of particular important in our scenario. That is, we are concerned with the condition on the number m such that $\mathcal{F}^{(2)}(C_{\text{ALT}}^{\ell,m,n}) \simeq \mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ holds. The following Theorem 4 and the subsequent Corollary 1, which can be readily derived from the theorem, provide a means for evaluating such m .

Theorem 4. If the ensemble of the unitary matrices corresponding to each block of $C_{\text{ALT}}^{2,m,n}$ and $C_{\text{ALT}}^{3,m,n}$ are 2-design, then the following inequalities hold:

$$\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n}) < \left(1 + \frac{1}{2^n}\right) \left(1 + \frac{1.2}{2^m}\right)^2 \left(1 + 8 \left(\left(1 + \frac{20.8}{2^{m/2}}\right)^{\frac{n}{m}-1} - 1 \right)\right) \mathcal{F}_{\text{Haar}}^{(2)}(2^n), \quad (21)$$

$$\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) < \left(1 + \frac{1}{2^n}\right) \left(1 + \frac{1.2}{2^m}\right)^2 \left(1 + 32 \left(\left(1 + \frac{83.2}{2^{m/2}}\right)^{\frac{n}{m}-1} - 1 \right)\right) \mathcal{F}_{\text{Haar}}^{(2)}(2^n). \quad (22)$$

Corollary 1. If $m = 2a \log_2 n$ and $143/(an^{a-1} \log_2 n) < 1$,

$$\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n}) < \left(1 + \frac{1}{2^n}\right) \left(1 + \frac{1.2}{n^{2a}}\right)^2 \left(1 + \frac{143}{an^{a-1} \log_2 n}\right) \mathcal{F}_{\text{Haar}}^{(2)}(2^n). \quad (23)$$

If $m = 2a \log_2 n$ and $2288/(an^{a-1} \log_2 n) < 1$,

$$\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) < \left(1 + \frac{1}{2^n}\right) \left(1 + \frac{1.2}{n^{2a}}\right)^2 \left(1 + \frac{2288}{an^{a-1} \log_2 n}\right) \mathcal{F}_{\text{Haar}}^{(2)}(2^n). \quad (24)$$

Recall from Eq. (2) that $\mathcal{F}^{(t)}(C) \geq \mathcal{F}_{\text{Haar}}^{(t)}(N)$ holds for any ansatz C . Therefore, if $m \geq 4 \log n$ and n is enough large, Corollary 1 implies that $\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n}) \sim \mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ and $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) \sim \mathcal{F}_{\text{Haar}}^{(2)}(2^n)$. This means that the ensembles of the states generated by $C_{\text{ALT}}^{2,m,n}$ and $C_{\text{ALT}}^{3,m,n}$ are almost 2-design. Hence in this case, from Theorem 2, the expressibility of ALT is as high as that of HAE. It is worth mentioning that, when $m = O(\log_2 n)$, the vanishing gradient problem does not happen in ALT as long as the cost function is local and ℓ is small [8]. More precisely, it was shown there that the variance of the gradient of such a cost function is larger than the value proportional to $O(1/2^{m\ell})$; thus, by taking $m = O(\log_2 n)$, the variance decreases with only $O(1/\text{poly}(n))$ as a function of n , whereas in the HEA case the same variance decreases exponentially fast as n becomes large. Therefore, the expressibility and the trainability coexists in the shallow ALT with $m = O(\log_2 n)$.

3.2 Expressibility measured by KL divergence

n	Ansatz	ℓ	m	Depth of each block	# of gate parameters	Example of the circuit
4	TEN	2	2	2	16	-
		3	2	2	24	Fig. 3a
	ALT	2	2	2	16	-
		3	2	2	24	Fig. 3b
	HEA	4	-	-	16	Fig. 3c
6	TEN	2	2	2	24	-
		3	2	2	36	-
	ALT	2	2	2	24	-
		3	2	2	36	-
	HEA	6	-	-	36	-
8	TEN	2	2	2	32	-
			4	4	64	-
		3	2	2	48	-
			4	4	96	-
	ALT	2	2	2	32	-
			4	4	64	-
		3	2	2	48	-
			4	4	96	-
	HEA	8	-	-	64	-
			-	-	64	-

Table 1: Parameters chosen for computing the KL-divergence. The number of gate parameters are computed by $n \times \ell \times$ (Depth of each block) for TEN and ALT, and $n \times \ell$ for HEA.

In Subsection 3.1, we have shown that the first two moments of $P(C_{\text{ALT}}^{\ell,m,n}, F)$ and $P(C_{\text{HEA}}^{\ell,n}, F)$ are close to those of $P_{\text{Haar}}(F)$, as long as $m = O(\log_2 n)$ and the block components of the ansatzes are sufficiently random. (Recall that, if every block is completely random, then the set of HEA constitutes the Haar ensemble.) The result implies that both $P(C_{\text{ALT}}^{\ell,m,n}, F)$ and $P(C_{\text{HEA}}^{\ell,n}, F)$ are close to $P_{\text{Haar}}(F)$ itself, meaning that $P(C_{\text{ALT}}^{\ell,m,n}, F) \simeq P(C_{\text{HEA}}^{\ell,n}, F) \simeq P_{\text{Haar}}(F)$. In this subsection, to support this conjecture, we evaluate the values of KL-divergence $\mathcal{E}(C) = D_{\text{KL}}(P(C, F) \| P_{\text{Haar}}(F))$ for the case $C = C_{\text{ALT}}^{\ell,m,n}$ and $C = C_{\text{HEA}}^{\ell,n}$, in addition to $C = C_{\text{TEN}}^{\ell,m,n}$ for comparison with various sets of (ℓ, m, n) . Especially, we focus on the relationship between the values of $\mathcal{F}^{(2)}(C)$ and $\mathcal{E}(C)$, and check if $\mathcal{F}^{(2)}(C) \simeq \mathcal{F}^{(2)}(C')$ would lead to $\mathcal{E}(C) \simeq \mathcal{E}(C')$ for a fixed n .

The parameters taken for calculating the KL-divergence are summarized in Table 1. Note that the circuits are chosen to be similar to those used in Section 3.1; for TEN and ALT, the depth of the circuits inside the blocks are all set to m so that the ensemble of the unitary matrices corresponding to those circuits become close to 2-design [16, 17]; for HEA, ℓ is set to n so that the ensemble of the unitary matrices corresponding to the *whole circuits* becomes close to 2-design. It is expected that $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,2,4}) \approx \mathcal{F}^{(2)}(C_{\text{HEA}}^{4,4})$ and $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,4,8}) \approx \mathcal{F}^{(2)}(C_{\text{HEA}}^{8,8})$ are realized, because, in Fig 2, we see that $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,2,4})$ and $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,4,8})$ are almost equals to the Haar values when the ensembles of unitary matrices corresponding to each block are 2-design. Thus, we here check if $\mathcal{F}^{(2)}(C) \simeq \mathcal{F}^{(2)}(C')$ would mean $\mathcal{E}(C) \simeq \mathcal{E}(C')$ in these parameter sets. As an example of the circuit, the whole structure of $C_{\text{TEN}}^{3,2,4}$, $C_{\text{ALT}}^{3,2,4}$, and $C_{\text{HEA}}^{4,4}$ in our settings are shown in Figs. 3a, 3b, and 3c, respectively. As illustrated in the figures, each layer is composed of parametrized single qubit gates and fixed 2-qubit CNOT gates.

In each trial of computing KL-divergence, we generate 200 states. When generating a state in each trial, we randomly choose the parameters and the type of single-qubit gate of the circuit. That is, for the i -th single qubit gate $R_i(\theta_i) = \exp(\sigma_{a_i} \theta_i)$ with $a_i = \{x, y, z\}$ and $\theta_i \in [0, 2\pi]$, in each trial all a_i and θ_i are randomly chosen. Then 200 fidelity values are computed, which are then used to construct the histogram with 1000 bins to approximate the probability distribution $P(C, F)$. Note that increasing the number of generated states and the number of bins do not affect the following conclusions.

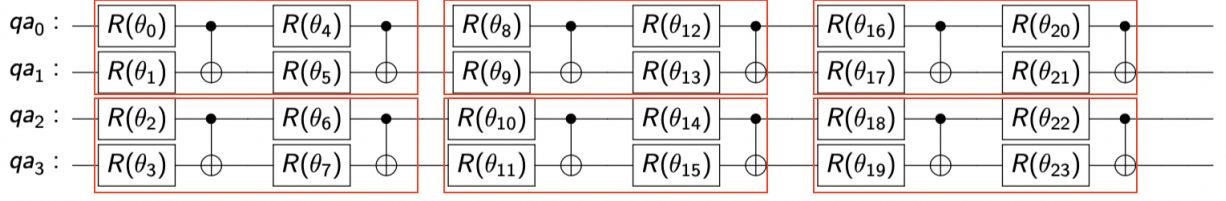
In this setting, Fig. 4 shows the KL divergences $\mathcal{E}(C_{\text{ALT}}^{\ell,m,n})$, $\mathcal{E}(C_{\text{HEA}}^{\ell,n})$, and $\mathcal{E}(C_{\text{TEN}}^{\ell,m,n})$. As a reference, we also show the values of $\mathcal{F}^{(2)}(C)/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ computed from the second moment of the fidelity distributions. Each data point and associated error bar is the average and the standard deviation of 10 trials of computation, respectively. Here is the list of points:

- For a fixed n , $\mathcal{E}(C_{\text{TEN}}^{\ell,m,n})$ is always bigger than $\mathcal{E}(C_{\text{ALT}}^{\ell,m,n})$ and $\mathcal{E}(C_{\text{HEA}}^{\ell,n})$.
- As the number of layers increases, the KL-divergence decreases for fixed (m, n) .
- For a fixed n , the tendency of the values of $\mathcal{F}^{(2)}(C)/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ is strongly correlated with that of KL-divergence.
- As expected, $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) \approx \mathcal{F}^{(2)}(C_{\text{HEA}}^{n,n})$ is realized when $(m, n) = (2, 4)$ and $(4, 8)$.
- $\mathcal{E}(C_{\text{ALT}}^{3,m,n})$ is as small as $\mathcal{E}(C_{\text{HEA}}^{n,n})$ in the parameter sets where $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) \approx \mathcal{F}^{(2)}(C_{\text{HEA}}^{n,n})$ is realized, i.e., $(m, n) = (2, 4)$ and $(4, 8)$. This result implies that the state distribution in ALT is also close to that in HEA, in the setting where the second frame potential is close to $\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$.

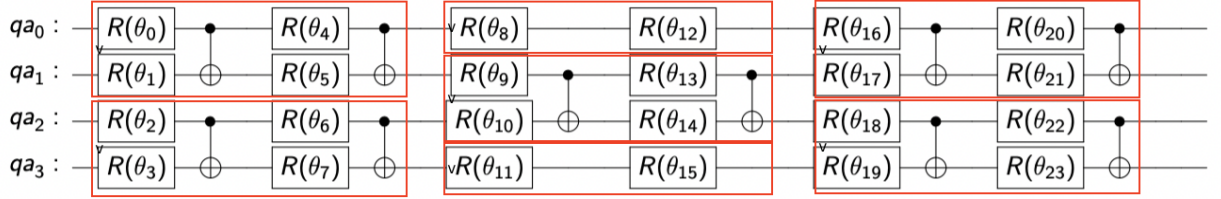
From some of the above observations, we find the strong correlation between $\mathcal{F}^{(2)}(C)/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ and $\mathcal{E}(C)$; that is, as $\mathcal{F}^{(2)}(C)$ becomes close to 1, then $\mathcal{E}(C)$ becomes close to 0. Therefore, combining the result obtained in Section 3.1, we get a clear evidence that, as far as $m = O(\log_2 n)$, $\mathcal{E}(C_{\text{ALT}}^{2,m,n}) \approx 0$ and $\mathcal{E}(C_{\text{ALT}}^{3,m,n}) \approx 0$ hold. That is, the high expressibility and trainability in ALT proven in Section 3.1 are assured also in terms of KL-divergence.

4 Application to VQE

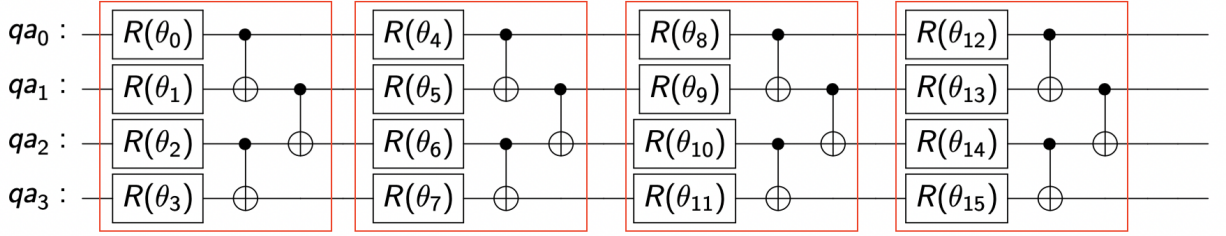
Recall that ALT was originally introduced with the motivation to resolve the vanishing gradient problem in VQE, which has been often observed when using HEA; then we were concerned with the expressibility of ALT in VQE, meaning that ALT would not offer a chance to reach the optimal solution due to the possible loss of expressibility. But we now know that this concern has been resolved under some conditions, as concluded in the previous section; that is, the expressibility and



(a) The structure of $C_{\text{TEN}}^{3,2,4}$. The red boxes correspond to blocks.



(b) The structure of $C_{\text{ALT}}^{3,2,4}$. The red boxes correspond to blocks.



(c) The structure of $C_{\text{HEA}}^{4,4}$.

Figure 3: The structures of ansatzes in our setting.

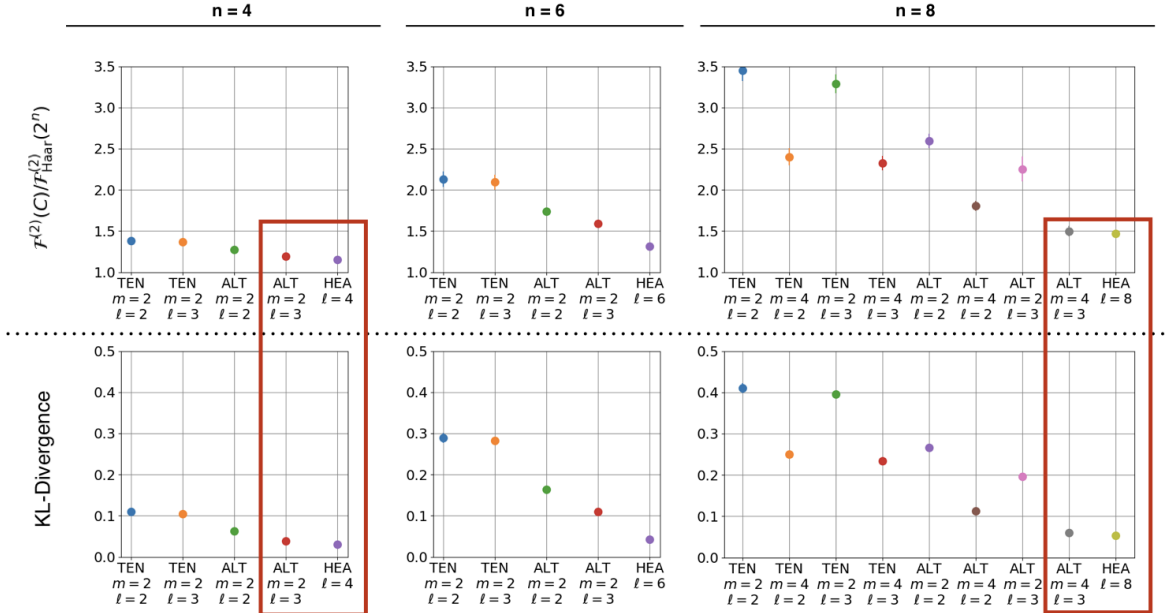


Figure 4: $\mathcal{F}^{(2)}(C)/\mathcal{F}_{\text{Haar}}^{(2)}(2^n)$ (top) and KL-divergence (bottom) for each ansatz. The sets of points with which $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) \approx \mathcal{F}^{(2)}(C_{\text{HEA}}^{3,n})$ hold are enclosed by the red rectangles.

the trainability coexists in the shallow ALT with $m = O(\log_2 n)$. This section provides a case study of VQE that implies this desirable fact.

We choose the Hamiltonian of 4-qubits Heisenberg model on a 1-dimensional lattice with periodic boundary conditions:

$$\mathcal{H} = \sum_{i=1}^4 (\sigma_x^i \sigma_x^{i+1} + \sigma_y^i \sigma_y^{i+1} + \sigma_z^i \sigma_z^{i+1}), \quad (25)$$

where σ_a^i ($a \in \{x, y, z\}$) is the Pauli matrix that operates on the i -th qubit and $\sigma_a^5 = \sigma_a^1$. The goal of VQE problem is to find the minimum eigenvalue of \mathcal{H} , by calculating the mean energy $\langle \mathcal{H} \rangle = \langle \psi_\theta | \mathcal{H} | \psi_\theta \rangle = \langle 0 | U_C(\theta)^\dagger \mathcal{H} U_C(\theta) | 0 \rangle$ via a quantum computer and then updating the parameter $\theta \in \Theta$ to decrease $\langle \mathcal{H} \rangle$ via a classical computer, in each iteration. As ansatzes, $C_{\text{TEN}}^{3,2,4}$, $C_{\text{ALT}}^{3,2,4}$, and $C_{\text{HEA}}^{4,4}$ are chosen. As indicated in Fig. 4, the values of KL divergence corresponding to these ansatzes show that $\mathcal{E}(C_{\text{TEN}}^{3,2,4}) > \mathcal{E}(C_{\text{ALT}}^{3,2,4}) \simeq \mathcal{E}(C_{\text{HEA}}^{4,4})$. That is, this ALT has the expressibility as high as that of HEA, and further, it is expected to enjoy the trainability unlike the HEA.

The simulation results are shown in Fig. 5. In the top three subfigures of Fig. 5, the blue lines and the associated error bars represent the average and the standard deviation of $\langle \mathcal{H} \rangle$ in total 100 trials, respectively. In each trial, the initial parameters of the ansatz are randomly chosen, and the optimization to decrease $\langle \mathcal{H} \rangle$ in each iteration is performed by using the Adam Optimizer with learning rate 0.001 [18]. The green line shows the theoretical minimum energy (i.e., the ground energy) of \mathcal{H} . Also the bottom subfigures of Fig. 5 show three trajectories of $\langle \mathcal{H} \rangle$ (red lines) whose energies at the final iteration step are the smallest three.

The ansatz $C_{\text{TEN}}^{3,2,4}$, which has the least expressibility in the sense of frame potentials and the KL divergence analysis, clearly gives the worst result; its least mean-energy is far above from the ground energy. This is simply because the state generated via $C_{\text{TEN}}^{3,2,4}$ cannot represent the ground state for any parameter choice. The result on $C_{\text{HEA}}^{4,4}$ is the second worst, which also does not reach the ground energy as in the case of TEN. Note that increasing the number of parameters does not change this result; that is, we also executed the simulation with $C_{\text{HEA}}^{4,6}$ that has the same number of parameters as $C_{\text{ALT}}^{3,2,4}$ but did not find a better result than that of $C_{\text{HEA}}^{4,4}$. On the other hand, $C_{\text{ALT}}^{3,2,4}$ succeeds in finding the ground energy; in fact, 5 of the total 100 trajectories generated via this ALT reach the ground energy. Hence this result implies that, in this example, the expressibility and the trainability coexist in ALT, while the latter is lacking in HEA.

This better trainability of ALT than that of HEA could be explained in terms of the “magnitude” of the gradient vector $\nabla_\theta \langle \mathcal{H} \rangle = [\partial \langle \mathcal{H} \rangle / \partial \theta_1, \dots, \partial \langle \mathcal{H} \rangle / \partial \theta_P]$, where P is the number of parameters. Now care should be taken to define an appropriate magnitude, because the focused ansatzes have different number of parameters. In this work, we regard $\partial \langle \mathcal{H} \rangle / \partial \theta_p$, ($p = 1, \dots, P$) as random variables and, based on this view, define the magnitude of $\nabla_\theta \langle \mathcal{H} \rangle$ as the mean of the absolute value of those random variables:

$$\|g(\theta)\| = \frac{1}{P} \sum_{p=1}^P \left| \frac{\partial \langle \mathcal{H} \rangle}{\partial \theta_p} \right|. \quad (26)$$

We evaluate the average of the magnitude (26) over sample trajectories, at several values of energy reached through the update of θ . For this purpose, let $\theta_C^{(i)}(t)$ denote the vector of parameters of a given ansatz C at the t -th step (number of iteration) of the i -th trajectory, and $E_C^{(i)}(t) = \langle 0 | U_C(\theta_C^{(i)}(t))^\dagger \mathcal{H} U_C(\theta_C^{(i)}(t)) | 0 \rangle$ be the energy at $\theta_C^{(i)}(t)$. Next, to define the average of $\|g(\theta_C)\|$ over the sample trajectories at the given energy value E , let $t_E^{(i)}$ be the smallest integer such that the energy of the i -th trajectory satisfies $E_C^{(i)}(t_E^{(i)}) \leq E$; in other words, $t_E^{(i)}$ represents the number of iteration such that the i -th trajectory first reaches the value E . Note that some trajectories may not reach a given value E for all the repetition of θ_C . (For example, as seen above, all trajectories of TEN never reached the value $E = -7$.) Hence, let \mathcal{I}_E be the set of index i such that the i -th trajectory reaches the value E at some point of $t_E^{(i)}$. We can now define the average of $\|g(\theta_C)\|$ as

$$\langle \|g(\theta_C)\| \rangle_E = \frac{1}{|\mathcal{I}_E|} \sum_{i \in \mathcal{I}_E} \left\| g(\theta_C^{(i)}(t_E^{(i)})) \right\|, \quad (27)$$

where $|\mathcal{I}_E|$ denotes the size of \mathcal{I}_E . Figure 6 shows Eq. (27) for the specific values of E (integers from -7 to 0) for the three ansatzes $C_{\text{TEN}}^{3,2,4}$, $C_{\text{ALT}}^{3,2,4}$, and $C_{\text{HEA}}^{4,4}$. The standard deviation of the average is

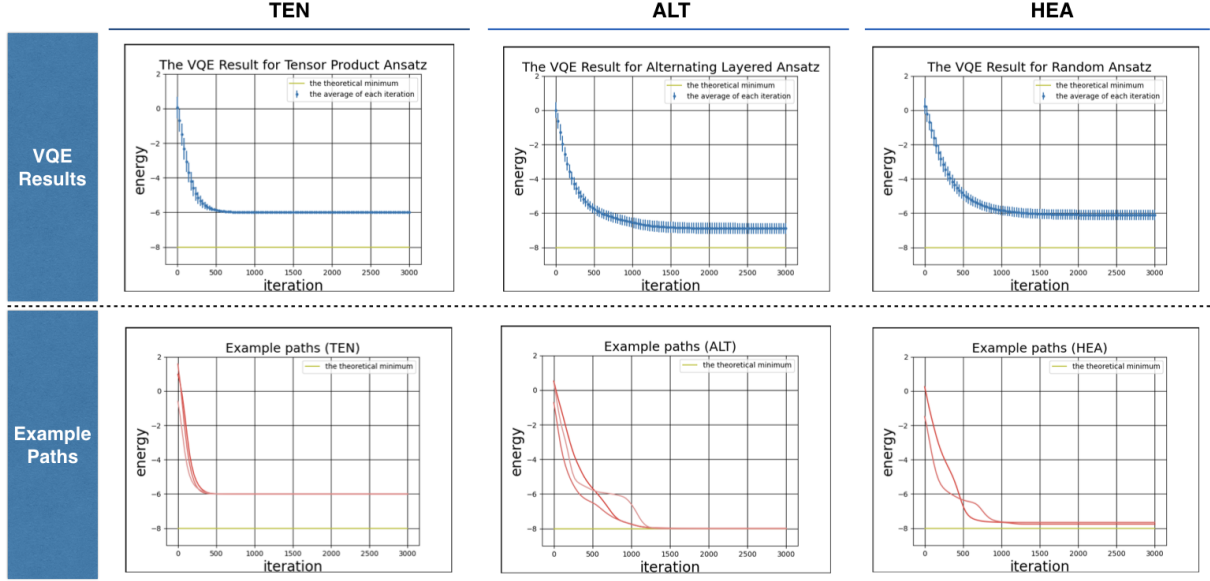


Figure 5: Top: Energy versus the iteration step in the VQE problem for the Hamiltonian (25), with the ansatz TEN (left), ALT (center), and HEA (right). The blue lines and the associated error bars represent the average and the standard deviation of the mean energies in total 100 trials, respectively; in each trial, the initial parameters of the ansatz are randomly chosen. Optimization to decrease $\langle \mathcal{H} \rangle$ in each iteration is performed by using Adam Optimizer with learning rate 0.001. Bottom: Three of 100 trajectories for each ansatz TEN (left), ALT (center), and HEA (right), indicated by red lines. The trajectories are chosen such that the energies at the final iteration step are the three smallest values.

indicated by the error bar. For instance, $\langle \|g(\theta_C)\| \rangle_E \simeq 5.6$ for the case of orange point (i.e., the case of ALT) at $E = -1$ was calculated with $|\mathcal{I}_E| = 100$; actually, all 100 trajectories become lower than $E = -1$. The figure shows that $\langle \|g(\theta_C)\| \rangle_E$ in ALT is always larger than that in HEA for all E , and this result is consistent to the theorems given in [8]. Such a larger gradient vector might enable ALT to circumvent possible flat energy landscapes and eventually realize the better trainability than HEA, but further studies are necessary to confirm this observation. Note that TEN has the largest values of $\langle \|g(\theta_C)\| \rangle_E$ when $E \geq -5$, which yet do not lead to the convergence to the global minimum due to the lack of expressibility.

5 Conclusion

This paper has examined the expressibility power of the shallow ALT, which was proposed before as a solution to the vanishing gradient problem found in various types of hybrid quantum-classical algorithms. Our conclusion is that, in addition to such a well trainability, shallow ALTs have the expressibility; that is, in the measure of the frame potential and the KL-divergence, those shallow ALTs have almost the same expressibility as HEAs, which are often used in hybrid algorithms but suffer from the vanishing gradient problem. In particular, we have proven that such expressibility holds if the number of entangled qubits in each block is of the order of the logarithm of the number of all resource qubits, which is consistent to the previous result discussing the trainability of ALT. We also provided a case study of VQE problem implying that the ALT enjoys both the expressibility and the trainability.

Although our results are limited to the case $\ell = 2, 3$, we have numerically observed that the ALT acquires even higher expressibility when making ℓ bigger. Therefore, we conjecture that the above conclusion still holds for ALT with $\ell \geq 4$. The rigorous proof is left for future work.

Acknowledgement: This work is supported by the MEXT Quantum Leap Flagship Program Grant Number JPMXS0118067285.

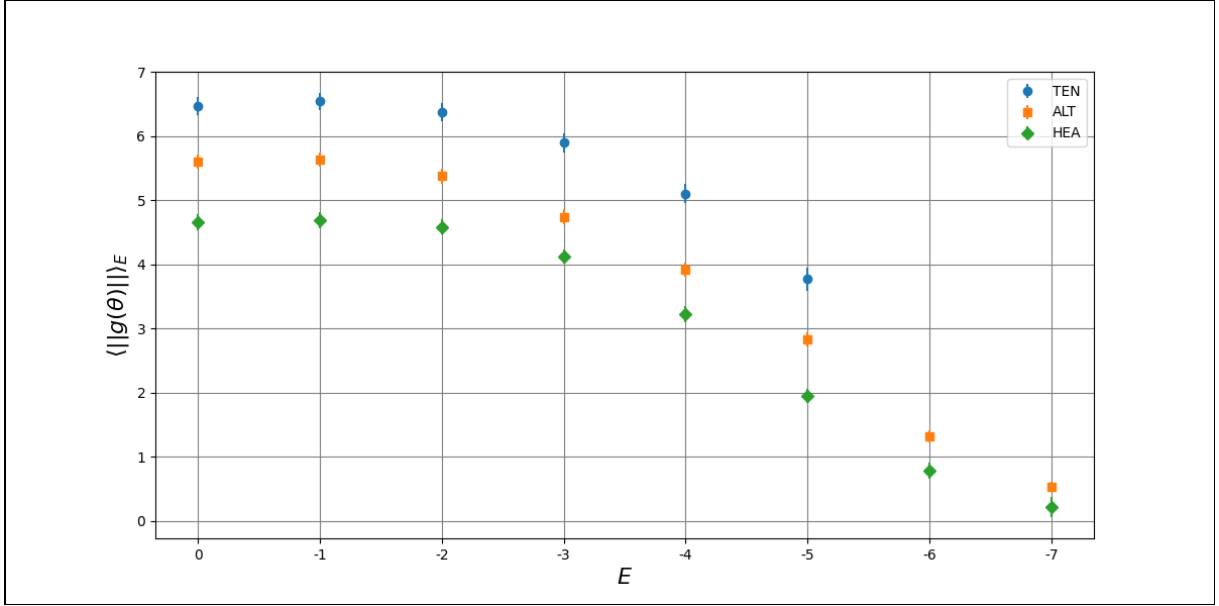


Figure 6: The average and the standard deviation of $\|g(\theta)\|$ versus E . The norm of the gradient is defined as $\|g(\theta)\| = \frac{1}{P} \sum_{p=1}^P \left| \frac{\partial \langle \mathcal{H} \rangle}{\partial \theta_p} \right|$.

A Proof of Theorems

Notation: For the unitary matrix U_a corresponding to the entire circuit, we denote the unitary matrix corresponding to the i -th layer of the circuit to be $U_a(i)$ and the unitary matrix corresponding to the j -th block in the i -th layer as $U_a(i, j)$.

A.1 Proof of Theorem 1

First, because the probability distribution of $F = |\langle \psi | \psi' \rangle|^2$ with n -qubits states $|\psi\rangle$ and $|\psi'\rangle$ taken from the Haar measure is given by $P_{\text{Haar}}(F) = (2^n - 1)(1 - F)^{2^n - 2}$, the value of $\mathcal{F}_{\text{Haar}}^{(1)}(2^n)$ is straightforwardly computed as

$$\mathcal{F}_{\text{Haar}}^{(1)}(2^n) = \int_{\text{Haar}} \int_{\text{Haar}} |\langle \psi | \psi' \rangle|^2 d\psi d\psi' = \int_0^1 dF F(2^n - 1)(1 - F)^{2^n - 2} = \frac{1}{2^n}. \quad (28)$$

Next, we provide the proof of $\mathcal{F}^{(1)}(C_{\text{ALT}}^{\ell, m, n}) = \mathcal{F}_{\text{Haar}}^{(1)}(2^n)$. Given two final states $|\phi\rangle = U_a|0\rangle$ and $|\psi\rangle = U_b|0\rangle$ generated by $C_{\text{ALT}}^{\ell, m, n}$, we have

$$\begin{aligned} \mathcal{F}^{(1)}(C_{\text{ALT}}^{\ell, m, n}) &= \int_{1\text{design}} dU_a dU_b \langle 0 | U_a^\dagger U_b | 0 \rangle \langle 0 | U_b^\dagger U_a | 0 \rangle \\ &= \int_{1\text{design}} \left(\prod_{i=1}^{\ell} dU_a(i) \right) \left(\prod_{i=1}^{\ell} dU_b(i) \right) \\ &\quad \times \langle 0 | U_a(1)^\dagger U_a(2)^\dagger \cdots U_a(\ell)^\dagger U_b(\ell) \cdots U_b(2) U_b(1) | 0 \rangle \\ &\quad \times \langle 0 | U_b(1)^\dagger U_b(2)^\dagger \cdots U_b(\ell)^\dagger U_a(\ell) \cdots U_a(2) U_a(1) | 0 \rangle \\ &= \int_{1\text{design}} \left(\prod_{i=1}^{\ell} \left(\prod_{j=1}^{k(i)} dU_a(i, j) \right) \right) \left(\prod_{i'=1}^{\ell} \left(\prod_{j'=1}^{k(i')} dU_b(i', j') \right) \right) \\ &\quad \times \langle 0 | U_a(1)^\dagger U_a(2)^\dagger \cdots U_a(\ell)^\dagger U_b(\ell) \cdots U_b(2) U_b(1) | 0 \rangle \\ &\quad \times \langle 0 | U_b(1)^\dagger U_b(2)^\dagger \cdots U_b(\ell)^\dagger U_a(\ell) \cdots U_a(2) U_a(1) | 0 \rangle, \end{aligned} \quad (29)$$

where $k(i)$ is the number of blocks in the i -th layer and each $\int dU$ is the average over the ensemble of the unitary matrix U . Because the distribution of each $U_a(i, j)$ is 2-design (and is therefore 1-

design), we can apply the formula (4) to the integrals with respect to $U_a(i, j)$. Actually, by integrating $\prod_{j=1}^{k(i)} dU_a(i, j)$ for all a in the last line of (29), we have

$$\begin{aligned}\mathcal{F}^{(1)}(C_{\text{ALT}}^{\ell, m, n}) &= \int_{1\text{design}} \left(\prod_{i=1}^{\ell} \left(\prod_{j=1}^{k(i)} dU_a(i, j) \right) \right) \left(\prod_{i'=1}^{\ell} \left(\prod_{j'=1}^{k(i')} dU_b(i', j') \right) \right) \\ &\quad \times \langle 0 | U_a(1)^\dagger U_a(2)^\dagger \cdots U_a(\ell-1)^\dagger U_a(\ell-1) \cdots U_a(2) U_a(1) | 0 \rangle \\ &\quad \times \langle 0 | U_b(1)^\dagger U_b(2)^\dagger \cdots U_b(\ell)^\dagger U_b(\ell) \cdots U_b(2) U_b(1) | 0 \rangle \\ &= \left(\frac{1}{2^m} \right)^\ell \int_{1\text{design}} \left(\prod_{i=1}^{L-1} \left(\prod_{\alpha=1}^{k(i)} dU_a^{i\alpha} \right) \right) \left(\prod_{j=1}^L \left(\prod_{\beta=1}^{k(i')} dU_b^{j\beta} \right) \right) \times 1 \\ &= \frac{1}{2^n} = \mathcal{F}_{\text{Haar}}^{(1)}(2^n).\end{aligned}\tag{30}$$

The other equality in Eq. (7) can be proved in the same manner.

A.2 Proof of Theorem 2

Similar to the first frame potential, the value of $\mathcal{F}_{\text{Haar}}^{(1)}(2^n)$ is straightforwardly computed as

$$\mathcal{F}_{\text{Haar}}^{(2)}(2^n) = \int_0^1 dF F^2 (2^n - 1) (1 - F)^{2^n - 2} = \frac{1}{2^{n-1}(2^n + 1)}.\tag{31}$$

Next, we compute $\mathcal{F}^{(2)}(C_{\text{TEN}}^{\ell, m, n})$. Given two final states $|\phi\rangle = U_a|0\rangle$ and $|\psi\rangle = U_b|0\rangle$, it is computed as

$$\begin{aligned}\mathcal{F}^{(2)}(C_{\text{TEN}}^{\ell, m, n}) &= \int_{2\text{design}} \left(\prod_{i=1}^{\ell} dU_a(i, 1) dU_b(i, 1) \right) |\langle 0 | U_a(1, 1)^\dagger U_a(2, 1)^\dagger \cdots U_a(\ell, 1)^\dagger U_b(\ell, 1) \cdots U_b(2, 1) U_b(1, 1) | 0 \rangle|^4 \\ &\quad \times \int_{2\text{design}} \left(\prod_{i=1}^{\ell} dU_a(i, 2) dU_b(i, 2) \right) |\langle 0 | U_a(1, 2)^\dagger U_a(2, 2)^\dagger \cdots U_a(\ell, 2)^\dagger U_b(\ell, 2) \cdots U_b(2, 2) U_b(1, 2) | 0 \rangle|^4 \\ &\quad \times \cdots \times \int_{2\text{design}} \left(\prod_{i=1}^{\ell} dU_a\left(i, \frac{n}{m}\right) dU_b\left(i, \frac{n}{m}\right) \right) \\ &\quad \times |\langle 0 | U_a(\ell, 1)^\dagger U_a(\ell, 2)^\dagger \cdots U_a(\ell, \ell)^\dagger U_b(\ell, \ell) \cdots U_b(\ell, 2) U_b(\ell, 1) | 0 \rangle|^4 \\ &= \left(\frac{1}{(2^m + 1)2^{m-1}} \right)^{\frac{n}{m}} = 2^{\frac{n}{m}-1} \cdot \frac{2^n + 1}{(2^m + 1)^{\frac{n}{m}}} \mathcal{F}_{\text{Haar}}^{(2)}(2^n).\end{aligned}\tag{32}$$

A.3 Proof of Theorem 3

Here we only show the computation of $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3, m, n})$. The computation of $\mathcal{F}^{(2)}(C_{\text{ALT}}^{2, m, n})$ can be done in a similar manner. The second frame potential can be expressed as follows:

$$\begin{aligned}\mathcal{F}^{(2)}(C_{\text{ALT}}^{3, m, n}) &= \int dU_a(1) dU_a(2) dU_a(3) dU_b(1) dU_b(2) dU_b(3) |\langle 0 | U_a(1)^\dagger U_a(2)^\dagger U_a(3)^\dagger U_b(3) U_b(2) U_b(1) | 0 \rangle|^4 \\ &= \int_{2\text{design}} \left(\prod_{i=1}^3 \prod_{j=1}^{k(i)} dU_a(i, j) \right) \left(\prod_{i'=1}^3 \prod_{j'=1}^{k(i')} dU_b(i', j') \right) \\ &\quad \times |\langle 0 | U_a(1, 1)^\dagger U_a(1, 2)^\dagger \cdots U_a(1, n/m)^\dagger U_a(2, 1)^\dagger U_a(2, 2)^\dagger \cdots U_a(2, n/m+1)^\dagger \\ &\quad \times U_a(3, 1)^\dagger U_a(3, 2)^\dagger \cdots U_a(3, n/m)^\dagger U_b(3, n/m) \cdots U_b(3, 2) U_b(3, 1) \\ &\quad \times U_b(2, n/m+1) \cdots U_b(2, 2) U_b(2, 1) U_b(1, n/m) \cdots U_b(1, 2) U_b(1, 1) | 0 \rangle|^4.\end{aligned}$$

Recall that $k(i)$ denotes the number of blocks in i -th layer; $k(i) = n/m$ for $i = 1, 3$ and $k(i) = n/m+1$ for $i = 2$. We can get the final formula in the theorem by integrating only the unitary matrices in the

first layer and the third layer. Executing integrals $\int_{2\text{design}} \prod_{j=1}^{n/m} dU_a(\ell, j)$ and $\int_{2\text{design}} \prod_{j'=1}^{n/m} dU_b(\ell, j')$ for $\ell = 1, 3$, we have

$$\begin{aligned}
\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) &= \int_{2\text{design}} \left(\prod_{j=1}^{n/m+1} dU_a(2, j) \right) \left(\prod_{j'=1}^{n/m+1} dU_b(2, j') \right) \\
&\quad \sum_{k_{11}^a=1}^4 \sum_{k_{12}^a=1}^4 \cdots \sum_{k_{1\frac{n}{m}}^a=1}^4 \sum_{k_{31}^a=1}^4 \sum_{k_{32}^a=1}^4 \cdots \sum_{k_{3\frac{n}{m}}^a=1}^4 \sum_{k_{11}^b=1}^4 \sum_{k_{12}^b=1}^4 \cdots \sum_{k_{1\frac{n}{m}}^b=1}^4 \\
&\quad \lambda_{k_{11}^a}^{(m)} \lambda_{k_{12}^a}^{(m)} \cdots \lambda_{k_{1\frac{n}{m}}^a}^{(m)} \lambda_{k_{31}^a}^{(m)} \lambda_{k_{32}^a}^{(m)} \cdots \lambda_{k_{3\frac{n}{m}}^a}^{(m)} \lambda_{k_{11}^b}^{(m)} \lambda_{k_{12}^b}^{(m)} \cdots \lambda_{k_{1\frac{n}{m}}^b}^{(m)} \\
&\quad \times \Delta^{(k_{11}^a, k_{31}^a, k_{11}^b)}(U_a(2, 1), U_b(2, 1)) \times \Delta^{(k_{11}^a, k_{31}^a, k_{11}^b, k_{12}^a, k_{32}^a, k_{12}^b)}(U_a(2, 2), U_b(2, 2)) \\
&\quad \times \Delta^{(k_{12}^a, k_{32}^a, k_{12}^b, k_{13}^a, k_{33}^a, k_{13}^b)}(U_a(2, 3), U_b(2, 3)) \times \cdots \\
&\quad \times \Delta^{(k_{1\frac{n}{m}}^a, k_{3\frac{n}{m}}^a, k_{1\frac{n}{m}}^b)}(U_a(2, \frac{n}{m}), U_b(2, \frac{n}{m})) \\
&\quad \times \Delta^{(k_{1\frac{n}{m}}^a, k_{3\frac{n}{m}}^a, k_{1\frac{n}{m}}^b)}(U_a(2, \frac{n}{m} + 1), U_b(2, \frac{n}{m} + 1)) \\
&= \sum_{k_{11}^a=1}^4 \sum_{k_{12}^a=1}^4 \cdots \sum_{k_{1\frac{n}{m}}^a=1}^4 \sum_{k_{31}^a=1}^4 \sum_{k_{32}^a=1}^4 \cdots \sum_{k_{3\frac{n}{m}}^a=1}^4 \sum_{k_{11}^b=1}^4 \sum_{k_{12}^b=1}^4 \cdots \sum_{k_{1\frac{n}{m}}^b=1}^4 \\
&\quad \int_{2\text{design}} dU_a(2, 1) dU_b(2, 1) \sqrt{\lambda_{k_{11}^a}^{(m)} \lambda_{k_{31}^a}^{(m)} \lambda_{k_{11}^b}^{(m)}} \Delta^{(k_{11}^a, k_{31}^a, k_{11}^b)}(U_a(2, 1), U_b(2, 1)) \\
&\quad \times \int_{2\text{design}} dU_a(2, 2) dU_b(2, 2) \sqrt{\lambda_{k_{12}^a}^{(m)} \lambda_{k_{32}^a}^{(m)} \lambda_{k_{12}^b}^{(m)}} \sqrt{\lambda_{k_{12}^a}^{(m)} \lambda_{k_{32}^a}^{(m)} \lambda_{k_{12}^b}^{(m)}} \Delta^{(k_{11}^a, k_{31}^a, k_{11}^b, k_{12}^a, k_{32}^a, k_{12}^b)}(U_a(2, 2), U_b(2, 2)) \\
&\quad \times \int_{2\text{design}} dU_a(2, 3) dU_b(2, 3) \sqrt{\lambda_{k_{12}^a}^{(m)} \lambda_{k_{32}^a}^{(m)} \lambda_{k_{12}^b}^{(m)}} \sqrt{\lambda_{k_{13}^a}^{(m)} \lambda_{k_{33}^a}^{(m)} \lambda_{k_{13}^b}^{(m)}} \Delta^{(k_{12}^a, k_{32}^a, k_{12}^b, k_{13}^a, k_{33}^a, k_{13}^b)}(U_a(2, 3), U_b(2, 3)) \\
&\quad \times \cdots \times \int_{2\text{design}} dU_a(2, \frac{n}{m}) dU_b(2, \frac{n}{m}) \sqrt{\lambda_{k_{1\frac{n}{m}}^a}^{(m)} \lambda_{k_{3\frac{n}{m}}^a}^{(m)} \lambda_{k_{1\frac{n}{m}}^b}^{(m)}} \sqrt{\lambda_{k_{1\frac{n}{m}}^a}^{(m)} \lambda_{k_{3\frac{n}{m}}^a}^{(m)} \lambda_{k_{1\frac{n}{m}}^b}^{(m)}} \\
&\quad \times \Delta^{(k_{1\frac{n}{m}}^a, k_{3\frac{n}{m}}^a, k_{1\frac{n}{m}}^b)}(U_a(2, \frac{n}{m}), U_b(2, \frac{n}{m})) \\
&\quad \times \int_{2\text{design}} dU_a(2, \frac{n}{m} + 1) dU_b(2, \frac{n}{m} + 1) \sqrt{\lambda_{k_{1\frac{n}{m}}^a}^{(m)} \lambda_{k_{3\frac{n}{m}}^a}^{(m)} \lambda_{k_{1\frac{n}{m}}^b}^{(m)}} \\
&\quad \times \Delta^{(k_{1\frac{n}{m}}^a, k_{3\frac{n}{m}}^a, k_{1\frac{n}{m}}^b)}(U_a(2, \frac{n}{m} + 1), U_b(2, \frac{n}{m} + 1)) \\
&= \mathbf{a}(3, m)^T B(3, m)^{\frac{n}{m}-1} \mathbf{a}(3, m), \tag{33}
\end{aligned}$$

where we use definitions: (13), (14), (17), and (18). For the purpose of exemplifying the computation in the first equality of (33), we show the computation when $n/m = 2$ in the following. When $n/m = 2$, the second frame potential can be computed as follows:

$$\begin{aligned}
\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,n,m}) &= \int_{2\text{design}} dU_a(1, 1) dU_a(1, 2) dU_a(2, 1) dU_a(2, 2) dU_a(2, 3) dU_a(3, 1) dU_a(3, 2) \\
&\quad dU_b(1, 1) dU_b(1, 2) dU_b(2, 1) dU_b(2, 2) dU_b(2, 3) dU_b(3, 1) dU_b(3, 2) \\
&\quad |\langle 0 | U_a^\dagger(1, 1) U_a^\dagger(1, 2) U_a^\dagger(2, 1) U_a^\dagger(2, 2) U_a^\dagger(2, 3) U_a^\dagger(3, 1) U_a^\dagger(3, 2) \\
&\quad U_b(3, 1) U_b(3, 2) U_b(2, 1) U_b(2, 2) U_b(2, 3) U_b(1, 1) U_b(1, 2) | 0 \rangle|^4
\end{aligned}$$

$$\begin{aligned}
&= \int_{2\text{design}} dU_a(1,1)dU_a(1,2)dU_a(2,1)dU_a(2,2)dU_a(2,3)dU_a(3,1)dU_a(3,2) \\
&\quad dU_b(1,1)dU_b(1,2)dU_b(2,1)dU_b(2,2)dU_b(2,3)dU_b(3,1)dU_b(3,2) \\
&\quad \sum_{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{l}, \mathbf{p}} \left(U_a^*(1,1)_{i_1 0}^{i_2 0} U_a^*(1,2)_{i_3 0}^{i_4 0} U_a^*(2,1)_{j_1 i_1} U_a^*(2,2)_{j_2 i_2}^{j_3 i_3} U_a^*(2,3)_{j_4 i_4} U_a^*(3,1)_{k_1 j_1}^{k_2 j_2} U_a^*(3,2)_{k_3 j_3}^{k_4 j_4} \right. \\
&\quad \left. U_b(3,1)_{l_1 1}^{k_2 l_2} U_b(3,2)_{l_3 1}^{k_4 l_4} U_b(2,1)_{l_1 p_1} U_b(2,2)_{l_2 p_2}^{l_3 p_3} U_b(2,3)_{l_4 p_4} U_b(1,1)_{p_1 0}^{p_2 0} U_b(1,2)_{p_3 0}^{p_4 0} \right) \times \\
&\quad \sum_{\mathbf{q}, \mathbf{r}, \mathbf{s}, \mathbf{t}, \mathbf{u}} \left(U_b^*(1,1)_{q_1 0}^{q_2 0} U_b^*(1,2)_{q_3 0}^{q_4 0} U_b^*(2,1)_{r_1 q_1} U_b^*(2,2)_{r_2 q_2}^{r_3 q_3} U_b^*(2,3)_{r_4 q_4} U_b^*(3,1)_{s_1 r_1}^{s_2 r_2} U_b^*(3,2)_{s_3 r_3}^{s_4 r_4} \right. \\
&\quad \left. U_a(3,1)_{s_1 t_1}^{s_2 t_2} U_a(3,2)_{s_3 t_3}^{s_4 t_4} U_a(2,1)_{t_1 u_1} U_a(2,2)_{t_2 u_2}^{t_3 u_3} U_a(2,3)_{t_4 u_4} U_a(1,1)_{u_1 0}^{u_2 0} U_a(1,2)_{u_3 0}^{u_4 0} \right) \times \\
&\quad \sum_{\mathbf{i}', \mathbf{j}', \mathbf{k}', \mathbf{l}', \mathbf{p}'} \left(U_a^*(1,1)_{i'_1 0}^{i'_2 0} U_a^*(1,2)_{i'_3 0}^{i'_4 0} U_a^*(2,1)_{j'_1 i'_1} U_a^*(2,2)_{j'_2 i'_2}^{j'_3 i'_3} U_a^*(2,3)_{j'_4 i'_4} U_a^*(3,1)_{k'_1 j'_1}^{k'_2 j'_2} U_a^*(3,2)_{k'_3 j'_3}^{k'_4 j'_4} \right. \\
&\quad \left. U_b(3,1)_{l'_1 1}^{k'_2 l'_2} U_b(3,2)_{l'_3 1}^{k'_4 l'_4} U_b(2,1)_{l'_1 p'_1} U_b(2,2)_{l'_2 p'_2}^{l'_3 p'_3} U_b(2,3)_{l'_4 p'_4} U_b(1,1)_{p'_1 0}^{p'_2 0} U_b(1,2)_{p'_3 0}^{p'_4 0} \right) \times \\
&\quad \sum_{\mathbf{q}', \mathbf{r}', \mathbf{s}', \mathbf{t}', \mathbf{u}'} \left(U_b^*(1,1)_{q'_1 0}^{q'_2 0} U_b^*(1,2)_{q'_3 0}^{q'_4 0} U_b^*(2,1)_{r'_1 q'_1} U_b^*(2,2)_{r'_2 q'_2}^{r'_3 q'_3} U_b^*(2,3)_{r'_4 q'_4} U_b^*(3,1)_{s'_1 r'_1}^{s'_2 r'_2} U_b^*(3,2)_{s'_3 r'_3}^{s'_4 r'_4} \right. \\
&\quad \left. U_a(3,1)_{s'_1 t'_1}^{s'_2 t'_2} U_a(3,2)_{s'_3 t'_3}^{s'_4 t'_4} U_a(2,1)_{t'_1 u'_1} U_a(2,2)_{t'_2 u'_2}^{t'_3 u'_3} U_a(2,3)_{t'_4 u'_4} U_a(1,1)_{u'_1 0}^{u'_2 0} U_a(1,2)_{u'_3 0}^{u'_4 0} \right)
\end{aligned} \tag{34}$$

where the bold symbols in the bottom the summation denote the multiple indices, e.g., $\mathbf{i} = i_1, i_2, i_3, i_4$. For the integrals $U_a(1,1), U_b(1,1), U_a(1,2), U_b(1,2)$,

$$\int_{2\text{design}} dU_a(1,1)U_a(1,1)_{u_1 0}^{u_2 0} U_a(1,1)_{u'_1 0}^{u'_2 0} U_a^*(1,1)_{i_1 0}^{i_2 0} U_a^*(1,1)_{i'_1 0}^{i'_2 0} = \sum_{k_{11}^a=1}^4 \lambda_{k_{11}^a}^{(m)} \Delta_{u_1 0 u'_1 0 i_1 0 i'_1 0}^{k_{11}^a} \Delta_{u_2 0 u'_2 0 i_2 0 i'_2 0}^{k_{11}^a}, \tag{35}$$

$$\int_{2\text{design}} dU_b(1,1)U_b(1,1)_{p_1 0}^{p_2 0} U_b(1,1)_{p'_1 0}^{p'_2 0} U_b^*(1,1)_{q_1 0}^{q_2 0} U_b^*(1,1)_{q'_1 0}^{q'_2 0} = \sum_{k_{11}^b=1}^4 \lambda_{k_{11}^b}^{(m)} \Delta_{p_1 0 p'_1 0 q_1 0 q'_1 0}^{k_{11}^b} \Delta_{p_2 0 p'_2 0 q_2 0 q'_2 0}^{k_{11}^b}, \tag{36}$$

$$\int_{2\text{design}} dU_a(1,2)U_a(1,2)_{u_3 0}^{u_4 0} U_a(1,2)_{u'_3 0}^{u'_4 0} U_a^*(1,2)_{i_3 0}^{i_4 0} U_a^*(1,2)_{i'_3 0}^{i'_4 0} = \sum_{k_{12}^a=1}^4 \lambda_{k_{12}^a}^{(m)} \Delta_{u_3 0 u'_3 0 i_3 0 i'_3 0}^{k_{12}^a} \Delta_{u_4 0 u'_4 0 i_4 0 i'_4 0}^{k_{12}^a}, \tag{37}$$

$$\int_{2\text{design}} dU_b(1,2)U_b(1,2)_{p_3 0}^{p_4 0} U_b(1,2)_{p'_3 0}^{p'_4 0} U_b^*(1,2)_{q_3 0}^{q_4 0} U_b^*(1,2)_{q'_3 0}^{q'_4 0} = \sum_{k_{12}^b=1}^4 \lambda_{k_{12}^b}^{(m)} \Delta_{p_3 0 p'_3 0 q_3 0 q'_3 0}^{k_{12}^b} \Delta_{p_4 0 p'_4 0 q_4 0 q'_4 0}^{k_{12}^b} \tag{38}$$

hold. For the integrals $U_a(3, 1), U_b(3, 1), U_a(3, 2), U_b(3, 2)$,

$$\begin{aligned} & \int_{2\text{design}} dU_a(3, 1) dU_b(3, 1) \\ & \sum_{\substack{k_1, k_2 \\ s_1, s_2}} \sum_{\substack{k'_1, k'_2 \\ s'_1, s'_2}} U_a^*(3, 1)_{k_1 j_1}^{k_2 j_2} U_b(3, 1)_{k_1 l_1}^{k_2 l_2} U_b^*(3, 1)_{s_1 r_1}^{s_2 r_2} U_a(3, 1)_{s_1 t_1}^{s_2 t_2} U_a^*(3, 1)_{k'_1 j'_1}^{k'_2 j'_2} U_b(3, 1)_{k'_1 l'_1}^{k'_2 l'_2} U_b^*(3, 1)_{s'_1 r'_1}^{s'_2 r'_2} U_a(3, 1)_{s'_1 t'_1}^{s'_2 t'_2} \\ & = \sum_{k_{31}^a=1}^4 \lambda_{k_{31}^a}^{(m)} \Delta_{j_1 l_1 j'_1 l'_1 t_1 r_1 t'_1 r'_1}^{k_{31}^a} \Delta_{j_2 l_2 j'_2 l'_2 t_2 r_2 t'_2 r'_2}^{k_{31}^a} \end{aligned} \quad (39)$$

$$\begin{aligned} & \int_{2\text{design}} dU_a(3, 2) dU_b(3, 2) \\ & \sum_{\substack{k_3, k_4 \\ s_3, s_4}} \sum_{\substack{k'_3, k'_4 \\ s'_3, s'_4}} U_a^*(3, 2)_{k_3 j_3}^{k_4 j_4} U_b(3, 2)_{k_3 l_3}^{k_4 l_4} U_b^*(3, 1)_{s_3 r_3}^{s_4 r_4} U_a(3, 2)_{s_3 t_3}^{s_4 t_4} U_a^*(3, 2)_{k'_3 j'_3}^{k'_4 j'_4} U_b(3, 2)_{k'_3 l'_3}^{k'_4 l'_4} U_b^*(3, 2)_{s'_3 r'_3}^{s'_4 r'_4} U_a(3, 2)_{s'_3 t'_3}^{s'_4 t'_4} \\ & = \sum_{k_{31}^a=1}^4 \lambda_{k_{31}^a}^{(m)} \Delta_{j_3 l_3 j'_3 l'_3 t_3 r_3 t'_3 r'_3}^{k_{32}^a} \Delta_{j_4 l_4 j'_4 l'_4 t_4 r_4 t'_4 r'_4}^{k_{32}^a} \end{aligned} \quad (40)$$

hold. Substituting (35), (36), (37), (38), (39), and (40) to (34), we get

$$\begin{aligned} \mathcal{F}^{(2)}(C_{\text{ALT}}^{3,n,m}) &= \sum_{k_{11}^a=1}^4 \sum_{k_{31}^a=1}^4 \sum_{k_{11}^b=1}^4 \sum_{k_{12}^a=1}^4 \sum_{k_{32}^a=1}^4 \sum_{k_{12}^b=1}^4 \lambda_{k_{11}^a}^{(m)} \lambda_{k_{31}^a}^{(m)} \lambda_{k_{11}^b}^{(m)} \lambda_{k_{12}^a}^{(m)} \lambda_{k_{32}^a}^{(m)} \lambda_{k_{12}^b}^{(m)} \\ & \int_{2\text{design}} dU_a(2, 1) dU_b(2, 1) \left[\sum_{\substack{u_1 u'_1 i_1 i'_1 \\ j_1 j'_1 l_1 l'_1}} \sum_{\substack{r_1 r'_1 t_1 t'_1 \\ p_1 p'_1 q_1 q'_1}} \Delta_{u_1 0 u'_1 0 i_1 0 i'_1 0}^{k_{11}^a} \Delta_{j_1 l_1 j'_1 l'_1 t_1 r_1 t'_1 r'_1}^{k_{31}^a} \Delta_{p_1 0 p'_1 0 q_1 0 q'_1 0}^{k_{11}^b} \right. \\ & U_a(2, 1)_{t_1 u_1} U_a(2, 1)_{t'_1 u'_1} U_a^*(2, 1)_{j_1 i_1} U_a^*(2, 1)_{j'_1 i'_1} U_b(2, 1)_{l_1 p_1} U_b(2, 1)_{l'_1 p'_1} U_b^*(2, 1)_{r_1 q_1} U_b^*(2, 1)_{r'_1 q'_1} \Big] \times \\ & \int_{2\text{design}} dU_a(2, 2) dU_b(2, 2) \left[\sum_{\substack{u_2 u'_2 i_2 i'_2 \\ j_2 j'_2 l_2 l'_2}} \sum_{\substack{r_2 r'_2 t_2 t'_2 \\ p_2 p'_2 q_2 q'_2}} \sum_{\substack{u_3 u'_3 i_3 i'_3 \\ j_3 j'_3 l_3 l'_3}} \sum_{\substack{r_3 r'_3 t_3 t'_3 \\ p_3 p'_3 q_3 q'_3}} \Delta_{u_2 0 u'_2 0 i_2 0 i'_2 0}^{k_{11}^a} \Delta_{j_2 l_2 j'_2 l'_2 t_2 r_2 t'_2 r'_2}^{k_{31}^a} \Delta_{p_2 0 p'_2 0 q_2 0 q'_2 0}^{k_{11}^b} \right. \\ & \Delta_{u_3 0 u'_3 0 i_3 0 i'_3 0}^{k_{12}^a} \Delta_{j_3 l_3 j'_3 l'_3 t_3 r_3 t'_3 r'_3}^{k_{32}^a} \Delta_{p_3 0 p'_3 0 q_3 0 q'_3 0}^{k_{12}^b} U_a(2, 2)_{t_2 u_2} U_a(2, 2)_{t'_2 u'_2} U_a^*(2, 2)_{j_2 i_2} U_a^*(2, 2)_{j'_2 i'_2} \\ & U_b(2, 2)_{l_2 p_2} U_b(2, 2)_{l'_2 p'_2} U_b^*(2, 2)_{r_2 q_2} U_b^*(2, 2)_{r'_2 q'_2} \Big] \times \\ & \int_{2\text{design}} dU_a(2, 3) dU_b(2, 3) \left[\sum_{\substack{u_4 u'_4 i_4 i'_4 \\ j_4 j'_4 l_4 l'_4}} \sum_{\substack{r_4 r'_4 t_4 t'_4 \\ p_4 p'_4 q_4 q'_4}} \Delta_{u_4 0 u'_4 0 i_4 0 i'_4 0}^{k_{12}^a} \Delta_{j_4 l_4 j'_4 l'_4 t_4 r_4 t'_4 r'_4}^{k_{32}^a} \Delta_{p_4 0 p'_4 0 q_4 0 q'_4 0}^{k_{12}^b} \right. \end{aligned} \quad (41)$$

$$\begin{aligned} & U_a(2, 3)_{t_4 u_4} U_a(2, 3)_{t'_4 u'_4} U_a^*(2, 3)_{j_4 i_4} U_a^*(2, 3)_{j'_4 i'_4} U_b(2, 3)_{l_4 p_4} U_b(2, 3)_{l'_4 p'_4} U_b^*(2, 3)_{r_4 q_4} U_b^*(2, 3)_{r'_4 q'_4} \Big] \\ & = \sum_{k_{11}^a=1}^4 \sum_{k_{31}^a=1}^4 \sum_{k_{11}^b=1}^4 \sum_{k_{12}^a=1}^4 \sum_{k_{32}^a=1}^4 \sum_{k_{12}^b=1}^4 \lambda_{k_{11}^a}^{(m)} \lambda_{k_{31}^a}^{(m)} \lambda_{k_{11}^b}^{(m)} \lambda_{k_{12}^a}^{(m)} \lambda_{k_{32}^a}^{(m)} \lambda_{k_{12}^b}^{(m)} \\ & \int_{2\text{design}} dU_a(2, 1) dU_b(2, 1) \Delta^{(k_{11}^a, k_{31}^a, k_{11}^b)}(U_a(2, 1), U_b(2, 1)) \times \\ & \int_{2\text{design}} dU_a(2, 2) dU_b(2, 2) \Delta^{(k_{11}^a, k_{31}^a, k_{11}^b, k_{12}^a, k_{32}^a, k_{12}^b)}(U_a(2, 2), U_b(2, 2)) \times \\ & \int_{2\text{design}} dU_a(2, 3) dU_b(2, 3) \Delta^{(k_{12}^a, k_{32}^a, k_{12}^b)}(U_a(2, 3), U_b(2, 3)), \end{aligned} \quad (43)$$

which is the right hand side of the first equality in (33) when $n/m = 2$.

A.4 Proof of Theorem 4

Similar to Theorem 3, we only show the inequality for $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n})$ here. The inequality for $\mathcal{F}^{(2)}(C_{\text{ALT}}^{2,m,n})$ can be shown in the same manner. In the process of showing the final inequality of the theorem, we expand $\mathbf{a}(3, m)$ and $B(3, m)$ as the sum of a vector/matrix whose components are $O(1)$ and a vector/matrix whose components are $O(1/2^{m/2})$.

To evaluate Eq. (13)¹, we can expand $\mathbf{a}(3, m)$ as

$$\mathbf{a}(3, m) = \frac{1}{2^m} \left(\mathbf{v}_0 + \frac{1.2}{2^{m/2}} \mathbf{v}_1 \right), \quad (44)$$

where

$$\mathbf{v}_{0i} = \begin{cases} 1 & i = 1 \quad (k_a, k_b, k_c = 1), \\ 1 & i = 22 \quad (k_a, k_b, k_c = 2), \\ 0 & \text{otherwise,} \end{cases} \quad |\mathbf{v}_{1i}| < 1. \quad (45)$$

Also, evaluating Eq. (17), we can expand $B(3, m)$ as

$$B(3, m) = \frac{1}{2^{2m}} \left(D + \frac{1.3}{2^{m/2-6}} X \right), \quad (46)$$

where

$$D_{ij} = \begin{cases} 1 & i = 1, j = 1 \quad (k_a, k_b, k_c, k_d, k_e, k_f = 1), \\ 1 & i = 22, j = 22 \quad (k_a, k_b, k_c, k_d, k_e, k_f = 2), \\ 0 & \text{otherwise,} \end{cases} \quad |X_{ij}| < \frac{1}{64}. \quad (47)$$

With $\alpha = n/m$, let $g_\alpha^k(X, D)$ be as the set of matrices expressed by $\prod_{i=1}^\alpha R_i$ where $R_i = D$ or X and the number of X s in $\{R_i\}$ is k . For example, $XDXX \in g_4^3(X, D)$ and $XDDDD \in g_4^1(X, D)$. Then, $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n})$ is expanded as

$$\begin{aligned} \mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) &= \frac{1}{2^{2m}} \left(\mathbf{v}_0^T + \frac{1.2}{2^{m/2}} \mathbf{v}_1^T \right) \frac{1}{2^{2n-2m}} \left(D + \frac{1.3}{2^{m/2-6}} X \right)^{\alpha-1} \left(\mathbf{v}_0 + \frac{1.2}{2^{m/2}} \mathbf{v}_1 \right) \\ &= \frac{1}{2^{2n}} \left(\mathbf{v}_0^T D^{\alpha-1} \mathbf{v}_0 + \left(\frac{2.4}{2^{m/2}} \right) \mathbf{v}_1^T D^{\alpha-1} \mathbf{v}_0 + \left(\frac{1.2^2}{2^m} \right) \mathbf{v}_1^T D^{\alpha-1} \mathbf{v}_1 \right) \\ &\quad + \frac{1}{2^{2n}} \sum_{k=1}^{\alpha-1} \left(\frac{1.3}{2^{m/2-6}} \right)^k \sum_{i=1}^{\alpha-1} \left(\mathbf{v}_0^T g_{\alpha i}^k \mathbf{v}_0 + \left(\frac{2.4}{2^{m/2}} \right) \mathbf{v}_1^T g_{\alpha i}^k \mathbf{v}_0 + \left(\frac{1.2^2}{2^m} \right) \mathbf{v}_1^T g_{\alpha i}^k \mathbf{v}_1 \right), \end{aligned} \quad (48)$$

where $g_{\alpha i}^k$ ($i = 1, 2 \dots \alpha C_k$) is an element of $g_\alpha^k(X, D)$. For an arbitrary $g \in g_\alpha^k(X, D)$ with $k \geq 1$,

$$\mathbf{v}_r^T g \mathbf{v}_s < (1, 1, \dots, 1) \begin{pmatrix} 1 & 0 \dots & 0 \\ 0 & 1 \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \dots & 1 \end{pmatrix}^{\alpha-k} \begin{pmatrix} \frac{1}{64} & \frac{1}{64} \dots & \frac{1}{64} \\ \frac{1}{64} & \frac{1}{64} \dots & \frac{1}{64} \\ \vdots & \vdots & \vdots \\ \frac{1}{64} & \frac{1}{64} \dots & \frac{1}{64} \end{pmatrix}^k \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = 64 \quad (49)$$

holds. For D and $\mathbf{v}_0, \mathbf{v}_1$

$$\mathbf{v}_1^T D^{\frac{n}{m}-1} \mathbf{v}_s < \sum_{i=1}^{64} (1 \cdot (\delta_{i1} + \delta_{i22}) \cdot 1) = 2 \quad (50)$$

$$\mathbf{v}_0^T D^{\frac{n}{m}-1} \mathbf{v}_0 = 2 \quad (51)$$

¹The evaluation procedure is straightforward, but a lot of computation is required. Thus, instead of performing a hand-calculation, we built an algorithm to evaluate (13) for arbitrary m and derived the expansion formula by computational calculation. We also built an algorithm for evaluating (17) and derived the expansion formula by computational calculation.

holds where $r, s = 0, 1$. By using the inequalities (49), (50), and (51), the upper bound for $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n})$ is derived as follows:

$$\begin{aligned}\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) &< \frac{1}{2^{2n}} \left(2 + 2 \left(\frac{2.4}{2^{m/2}} \right) + 2 \left(\frac{1.2^2}{2^m} \right) + 64 \sum_{k=1}^{\alpha-1} \alpha_{-1} C_k \left(\frac{1.3}{2^{m/2-6}} \right)^k 1^k \left(1 + \frac{1.2}{2^{m/2}} \right)^2 \right) \\ &= \frac{1}{2^{2n-1}} \left(1 + \frac{1.2}{2^m} \right)^2 \left(1 + 32 \left(\left(1 + \frac{1.3}{2^{m/2-6}} \right)^{\alpha-1} - 1 \right) \right) \\ &= \left(1 + \frac{1}{2^n} \right) \left(1 + \frac{1.2}{2^m} \right)^2 \left(1 + 32 \left(\left(1 + \frac{83.2}{2^{m/2}} \right)^{\alpha-1} - 1 \right) \right) \mathcal{F}_{\text{Haar}}^{(2)}(2^n).\end{aligned}\quad (52)$$

A.5 Proof of Corollary 1

As in the above theorems, we only show the inequality for $\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n})$ here. When $m = 2a \log_2 n$,

$$\left(1 + \frac{83.2}{2^{m/2}} \right)^{\alpha-1} < \left[\left(1 + \frac{83.2}{2^{m/2}} \right)^{\frac{2^{m/2}}{83.2}} \right]^{\frac{83.2n}{2^{m/2}m}} < e^{\frac{83.2n}{2^{m/2}m}} = e^{\frac{41.6}{an^{a-1} \log_2 n}}. \quad (53)$$

If $41.6/(an^{a-1} \log_2 n) < 1$,

$$e^{\frac{41.6}{an^{a-1} \log_2 n}} < 1 + (e-1) \frac{41.6}{an^{a-1} \log_2 n}. \quad (54)$$

Substituting Eqs. (53), (54), and $m = 2a \log_2 n$ into (52), we get

$$\mathcal{F}^{(2)}(C_{\text{ALT}}^{3,m,n}) < \left(1 + \frac{1}{2^n} \right) \left(1 + \frac{1.2}{n^{2a}} \right)^2 \left(1 + \frac{2288}{an^{a-1} \log_2 n} \right) \mathcal{F}_{\text{Haar}}^{(2)}(2^n). \quad (55)$$

References

- [1] J. Preskill. Quantum computing in the nisc era and beyond. *Quantum*, 2:79, 2018. DOI: <https://doi.org/10.22331/q-2018-08-06-79>.
- [2] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):1–7, 2014. DOI: <https://doi.org/10.1038/ncomms5213>.
- [3] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017. DOI: <https://doi.org/10.1038/nature23879>.
- [4] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):1–6, 2018. DOI: <https://doi.org/10.1038/s41467-018-07090-4>.
- [5] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, 2019. DOI: <https://doi.org/10.22331/q-2019-12-09-214>.
- [6] J. Stokes, J. Isaac, N. Killoran, and G. Carleo. Quantum natural gradient. *Quantum*, 4:269, 2020. DOI: <https://doi.org/10.22331/q-2020-05-25-269>.
- [7] N. Yamamoto. On the natural gradient for variational quantum eigensolver. *arXiv preprint arXiv:1909.05074*, 2019.
- [8] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature communications*, 12 1:1791, 2021. DOI: <https://doi.org/10.1038/s41467-021-21728-w>.
- [9] S. Sim, P. D. Johnson, and A. Aspuru-Guzik. Expressibility and entangling capability of parametrized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2(12):1900070, 2019. DOI: <https://doi.org/10.1002/qute.201900070>.
- [10] T. Ali, A. Bhattacharyya, S. S. Haque, E. H. Kim, N. Moynihan, and J. Murugan. Chaos and complexity in quantum mechanics. *Physical Review D*, 101(2):026021, 2020. DOI: [https://doi.org/10.1007/JHEP04\(2017\)121](https://doi.org/10.1007/JHEP04(2017)121).

- [11] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, 2004. DOI: <https://doi.org/10.1063/1.1737053>.
- [12] A. Klappenecker and M. Rotteler. Mutually unbiased bases are complex projective 2-designs. In *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, pages 1740–1744. IEEE, 2005. DOI: <https://doi.org/10.1109/ISIT.2005.1523643>.
- [13] I. Bengtsson and K. Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge university press, 2017. DOI: <https://doi.org/10.1017/CBO9780511535048>.
- [14] K. Życzkowski and H.-J. Sommers. Average fidelity between random quantum states. *Physical Review A*, 71(3):032313, 2005. DOI: <https://doi.org/10.1103/PhysRevA.71.032313>.
- [15] Z. Puchała and J. Miszcza. Symbolic integration with respect to the haar measure on the unitary groups. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 65(No 1): 21–27, 2017. DOI: <https://doi.org/10.1515/bpasts-2017-0003>.
- [16] F. G. Brandao, A. W. Harrow, and M. Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346(2):397–434, 2016. DOI: <https://doi.org/10.1007/s00220-016-2706-8>.
- [17] A. Harrow and S. Mehraban. Approximate unitary t -designs by short random quantum circuits using nearest-neighbor and long-range gates. *arXiv preprint arXiv:1809.06957*, 2018.
- [18] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.