Quantum Science and Technology

OPEN ACCESS

RECEIVED
29 November 2023

REVISED
21 March 2024

ACCEPTED FOR PUBLICATION
30 April 2024

PUBLISHED
13 May 2024

PAPER

Sequential optimal selections of single-qubit gates in parameterized quantum circuits

Kaito Wada[1], **Rudy Raymond**[2,3,4], **Yuki Sato**[3,5] and **Hiroshi C Watanabe**[3,6,*]

[1] Department of Applied Physics and Physico-Informatics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan
[2] IBM Quantum, IBM Japan 19-21 Nihonbashi Hakozaki-cho, Chuo-ku, Tokyo 103-8510, Japan
[3] Quantum Computing Center, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan
[4] Department of Computer Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[5] Toyota Central R&D Labs., Inc., Koraku Mori Building 10F, 1-4-14 Koraku, Bunkyo-ku, Tokyo 112-0004, Japan
[6] Department of Chemistry, Faculty of Science, Kyushu University, 744 Motooka Nishi-ku, Fukuoka 819-0395, Japan
[*] Author to whom any correspondence should be addressed.

E-mail: hcwatanabe@chem.kyushu-univ.jp

Keywords: quantum computation, variational quantum algorithm, parametrized quantum circuit, optimization problem

Abstract

In variational quantum algorithms, it is important to balance conflicting requirements of expressibility and trainability of a parameterized quantum circuit (PQC). However, appropriate PQC designs are not necessarily trivial. Here, we propose an algorithm for optimizing the PQC structure, where single-qubit gates are sequentially replaced by the optimal ones via diagonalization of a matrix whose elements are evaluated on slightly modified circuits. This replacement leads to a better approximation of target states with limited circuit depth. Furthermore, we clarify the existence of a barren plateau in the sequential optimization in terms of the spectrum concentration of the matrix, which defines the cost landscape with respect to changes in the target gate. Then, we rigorously show the concentration is no faster than polynomials in the number of qubits when an n-qubit PQC depth is $O(\log n)$ using local observables. Finally, numerical experiments are provided to show the convergence of our method which is faster than classical optimizers on both simulators and a real device. Our results provide evidences for sequential optimizers as better alternatives to optimize PQCs on near-term quantum devices.

1. Introduction

Recent progress in quantum hardware accelerates the development of classical-quantum hybrid algorithms for near-term quantum advantage. Variational quantum algorithm (VQA) is one of the most notable hybrid algorithms; this is a versatile methodology applicable to quantum chemical calculation [1–6], combinatorial optimization problem [7–10], dynamics simulation [11, 12], time evolving simulation [13, 14], principle component analysis [15, 16], linear [17, 18] and nonlinear [19, 20] system solvers, and so on. VQA, especially variational quantum eigensolver (VQE) [1], also draws attention as a promising subroutine to prepare an initial state for other hybrid algorithms in the era of early fault-tolerant quantum computing [21]. For instance, a shallow-depth state preparation in VQA combined with the subsequent moderate-depth quantum operations such as Hamiltonian time evolution will lead to successful calculations for industry-relevant problems [22–25].

The key component in VQAs is a parameterized quantum circuit (PQC), also called ansatz. A target quantum state is generated on a PQC by classically optimizing the parameters along with a cost function [26, 27]. Hence, appropriate designs of PQCs are essential to express the quantum states of interest. PQCs are classified into physics-based ansatz and heuristic ansatz. A physics-based ansatz, including a unitary coupled cluster [28] and Hamiltonian variational ansatz [29, 30], can approximate the state of interest at high precision by limiting the Hilbert space spanned by the ansatz to the target neighborhood. However, these circuits are too deep. Since the circuit depth is practically limited due to the hardware noise, it is desirable to

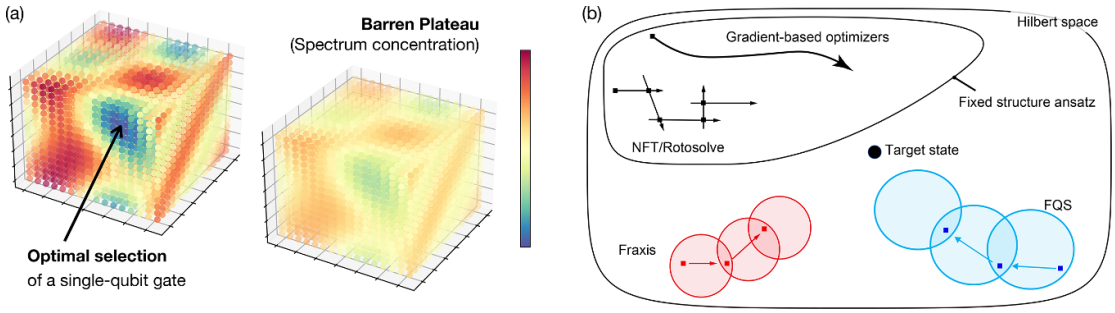


Figure 1. Overview of this work. (a) The cost landscape due to a parametrized single-qubit gate in PQCs. In our algorithm, we use quaternion representation of a general single-qubit gate, which has three degrees of freedom, and this representation allows us to select an optimal gate from all possible ones to minimize the VQA cost function. Also, our method is subjected to barren plateaus on deep circuits but can avoid them in shallow circuits and local cost functions. (b) Schematic diagram of VQA with FQS in comparison with other optimizers. The blue and red circles correspond to the expressibility of the single-qubit gate of FQS and Fraxis, respectively. Since FQS (Fraxis) has full (partial) degrees of freedom in a single-qubit gate, the FQS (blue) circles are larger than those of Fraxis (red). FQS and Fraxis can find the optimal state in its circle that drifts stepwise reflecting the circuit structure optimization.

save the depth not only to maintain VQA performance itself but also for the subsequent quantum operations on early fault-tolerant devices. In contrast, a heuristic ansatz puts more weight on hardware efficiency, which generally results in shallower circuits albeit with the uncertainty to express the target states.

A typical strategy to mitigate the uncertainty is systematically increasing the circuit depth. However, increasing depth can cause the gradient of the cost function with regards to parameters of PQCs to exponentially vanish as the number of qubits grows, a phenomenon termed *barren plateau* [31–33]. The barren plateau renders gradient-based approaches useless. Note that it is also the case for other optimizers such as Nelder–Mead, Powell, and COBYLA, that do not explicitly employ gradients but cost differences [34]. Several remedies, such as layerwise-learning [35] and parameter correlation [36], have been proposed but they can only cope with noiseless conditions. The only effective strategy for the noise-induced barren plateau to date is reducing circuit depth [32]. Hence, there are two conflicting requirements of PQCs with heuristic ansatz; they should be deep enough to express target states but should be as shallow as possible to avoid both types of barren plateau.

The convergence rate in optimization is also of great interest such as for state preparation in early fault-tolerant quantum computing and *quantum relaxation* [7], where finding good-enough quantum states is more important than finding the ground state. In classical tensor networks, sequential optimizations such as the density matrix renormalization group (DMRG) are widely used because of their fast convergence. DMRG optimizes only a small subset of the tensors while freezing the rest of them at each optimization step. Hence, sequential optimization also seems to be a promising approach for quantum circuit optimization and is drawing increasing attention [37]. However, the application of classical tensor network optimizers to quantum circuits is not straightforward. Recently, there are some proposals for the sequential optimization method in quantum circuits [38–41], but their advantage in convergence over conventional gradient-based optimization has been confirmed mainly for systems less than ten qubits. Moreover, little is known about both the theoretical analysis regarding barren plateau and relative advantages among these sequential optimizations.

In this work, we propose an algorithm for optimizing the circuit structure of a PQC by sequential optimal selection of general single-qubit gates $R_{\boldsymbol{n}}(\psi)$ that rotate a qubit around a rotational axis $\boldsymbol{n}$ with an angle ψ. This is in contrast to popular circuit structures of PQCs with single-qubit gates limited to fixed rotational x, y and z axes. In our method, the $R_{\boldsymbol{n}}(\psi)$ is mapped to a unit quaternion, which transforms a VQA cost function into a solvable quadratic form. Then the quadratic form allows for an optimal selection of a single-qubit gate to locally minimize the cost function as depicted in figure 1(a), which, as a result, leads to the structure optimization of PQCs (i.e. the rotational axes in parameterized gates [39]) while maintaining the circuit depth. We call this sequential algorithm *free quaternion selection* (*FQS*).

Since the target gate for FQS application has the highest degree of freedom over the whole single-qubit gates, simultaneous optimization of the multiple degrees in the gate leads to faster convergence, compared to both the gradient-based optimizers and the existing sequential quantum optimizers (where parameterized gates are sequentially and analytically optimized without gradients) such as Rotosolve/Rotoselect [39], NFT [38], and Fraxis [40] as illustrated in figure 1(b). Actually, we confirm the fast convergence of FQS in experiments with as large as 20 qubits. Here, the Rotosolve and NFT method find an optimal single-qubit gate $R_{\boldsymbol{m}}(\psi)$ under the constraint of the fixed-axis $\boldsymbol{m}$ by using the sinusoidal structure of the VQA cost

function with respect to ψ. The Rotoselect method iterates the procedure for each element of a discrete set of axes and then adopts the axis with the lowest cost. As for (π-) Fraxis method, the rotational axis $\boldsymbol{n}$ of an angle-fixed single-qubit gate $R_{\boldsymbol{n}}(\pi)$ is optimized by matrix diagonalization which is generalized by the proposed FQS method. It turns out that the FQS generalizes all sequential quantum optimizers; Rotosolve/NFT and Fraxis can be formulated with corresponding quadratic forms derived from the original one appeared in FQS procedure.

Next, we unveil several important properties of this optimizer family with respect to barren plateau. Since the sequential quantum optimizers do not directly use gradients (and cost differences), it is not straightforward to study the properties of these optimizers with regards to barren plateau. However, the essential nature of barren plateau is the loss of trainability as the number of qubits increases, which reflects an extreme separation between the initial and the target states caused by exponential inflation of the dimension of Hilbert space. Hence, it is worth investigating the relation between loss of the trainability and the system size in the sequential quantum optimizers. To clarify this point, we introduce the spectral radius of the matrix associated with the quadratic form of each sequential optimizer as an alternative measure for trainability. The spectral radius corresponds to the cost variation by one-time application of each optimization method to a single-qubit gate.

Here, we rigorously prove that under the same conditions as assumed in their gradient-based counterparts [31, 33], the sequential optimizers will likely run into barren plateau when a PQC has sufficient expressibility. More precisely, if the circuits become sufficiently deep to achieve unitary 2-design over the whole system, the second moment of the spectral radius shows exponential decay regardless of the global or local cost. This means that the spectrum of the corresponding matrix of the quadratic form concentrates on a single value exponentially fast with respect to the number of qubits as depicted in figure 1(a). On the other hand, we also rigorously prove that when the cost functions are local observables on an alternating layered ansatz [33, 42], the exponential concentration of the spectrum cannot occur as long as the depth of the n-qubit PQCs is $\mathcal{O}(\log n)$, and as a result the barren-plateau problems can be avoided.

We further demonstrate these properties of FQS using extensive numerical simulations and an experiment on a real quantum device up to 20 qubits. While the sequential optimizers have similar resistance to barren plateau to gradient-based optimizers, they show faster convergence for larger system, which shed light on sequential optimizers as better alternatives to the gradient-based ones.

2. Theory

2.1. Free quaternion selection for variational quantum algorithm

We first show that cost functions for VQAs are formulated as a quadratic form on a single-qubit gate. To this end, we use a quaternion representation of general single-qubit gates:

$$R_{\boldsymbol{n}}(\psi) := \mathrm{e}^{-\mathrm{i}\frac{\psi}{2}\boldsymbol{n}\cdot\vec{\sigma}} = \boldsymbol{q}\cdot\vec{\varsigma}. \tag{1}$$

The rotational axis $\boldsymbol{n}$ and angle ψ in the single-qubit gate are encoded together in a four-dimensional parameter $\boldsymbol{q}$ with the unit norm, that is, a unit quaternion. The four-component operator $\vec{\varsigma} := (I, -iX, -iY, -iZ)$ is an extension of the Pauli matrices $\vec{\sigma} = (X, Y, Z)$ with the 1-qubit identity I; the derivation of the relation (1) is provided in appendix A. Hereafter, we write $R_{\boldsymbol{n}}(\psi)$ as $R(\boldsymbol{q})$ for simplicity.

We aim to solve an optimization task with the following objective function

$$\sum_{k=1}^{K} \mathrm{tr}\left[\rho_k U\left(\{\boldsymbol{q}_d\}_{d=1}^{D}\right)^{\dagger} H_k U\left(\{\boldsymbol{q}_d\}_{d=1}^{D}\right)\right], \tag{2}$$

where ρ_k is an n-qubit initial state from a training set, and H_k is some observable. Here, U is an n-qubit PQC comprised of D parameterized single-qubit gates $\{R(\boldsymbol{q}_d)\}_{d=1}^{D}$ and parameter-free gates such as CNOT gate. In the following, without loss of generality, we focus on a single expected value in the objective function (i.e. $K = 1$, which can simply be considered as the minimization of energy for the Hamiltonian $H := H_1$ and the input state $\rho_{\mathrm{in}} := \rho_1$). The extension of our algorithm to the whole objective function is trivial due to the linearity.

For the energy minimization, we focus on a sequential optimization regarding $R(\boldsymbol{q}_d)$ where all parameters are fixed except for the dth single-qubit gate as shown in figure 2. The energy expectation is written as

$$\begin{aligned}\langle H\rangle(\boldsymbol{q}_d) &:= \mathrm{tr}\left[\rho_{\mathrm{in}} U_1^{\dagger} R(\boldsymbol{q}_d)^{\dagger} U_2^{\dagger} H U_2 R(\boldsymbol{q}_d) U_1\right], \\ &= \mathrm{tr}\left[\rho'_{\mathrm{in}} R(\boldsymbol{q}_d)^{\dagger} H' R(\boldsymbol{q}_d)\right],\end{aligned} \tag{3}$$

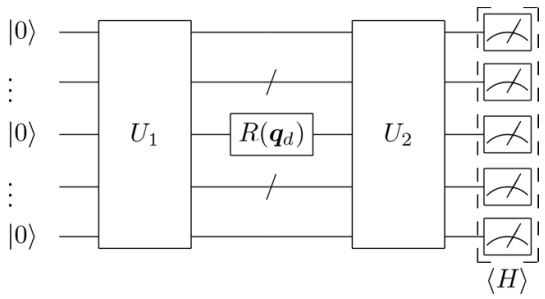


Figure 2. A quantum circuit to optimize the single-qubit gate $R(\boldsymbol{q}_d)$. The variational parameters in U_1 and U_2 are fixed during the optimization of the gate of interest. Our algorithm sequentially replaces $R(\boldsymbol{q}_d)$ in the circuit by the optimal one, which is computed by the diagonalization of the corresponding matrix S whose elements can be evaluated from the circuit by replacing $R(\boldsymbol{q}_d)$ with ten different gates as described in appendix C.

where U_1 and U_2 are the quantum circuits before and after $R(\boldsymbol{q}_d)$, respectively. H' and $\rho'_{\rm in}$ are defined as

$$H' := U_2^\dagger H U_2, \quad \rho'_{\rm in} := U_1 \rho_{\rm in} U_1^\dagger. \tag{4}$$

Here we omit the subscript d for simplicity. Substituting equation (1) into equation (3), we can obtain the following quadratic form

$$\langle H \rangle (\boldsymbol{q}) = \boldsymbol{q}^\top S \boldsymbol{q}, \tag{5}$$

where superscript $\top$ denotes a transpose operation. Denoting each element of the four-component operator $\vec{\varsigma}$ as ς_μ $(\mu = 0,1,2,3)$, $S = (S_{\mu\nu})$ is a 4×4 real-symmetric matrix whose elements are defined as

$$S_{\mu\nu} := \frac{1}{2} \mathrm{tr}\left[\rho'_{\rm in}\left(\varsigma_\mu^\dagger H' \varsigma_\nu + \varsigma_\nu^\dagger H' \varsigma_\mu\right)\right], \tag{6}$$

and more explicitly

$$S = \begin{pmatrix} \mathrm{tr}\left(H'\rho'_{\rm in}\right) & \frac{i}{2}\mathrm{tr}\left(H'\left[\rho'_{\rm in}, X\right]\right) & \frac{i}{2}\mathrm{tr}\left(H'\left[\rho'_{\rm in}, Y\right]\right) & \frac{i}{2}\mathrm{tr}\left(H'\left[\rho'_{\rm in}, Z\right]\right) \\ \cdot & \mathrm{tr}\left(H'X\rho'_{\rm in}X\right) & \frac{1}{2}\left[\mathrm{tr}\left(H'X\rho'_{\rm in}Y + H'Y\rho'_{\rm in}X\right)\right] & \frac{1}{2}\left[\mathrm{tr}\left(H'X\rho'_{\rm in}Z + H'Z\rho'_{\rm in}X\right)\right] \\ \cdot & \cdot & \mathrm{tr}\left(H'Y\rho'_{\rm in}Y\right) & \frac{1}{2}\left[\mathrm{tr}\left(H'Y\rho'_{\rm in}Z + H'Z\rho'_{\rm in}Y\right)\right] \\ \cdot & \cdot & \cdot & \mathrm{tr}\left(H'Z\rho'_{\rm in}Z\right) \end{pmatrix}, \tag{7}$$

where $[\cdot,\cdot]$ denotes the commutation relation. See appendix B for the derivation of the quadratic form.

The matrix S can be obtained by running and measuring ten quantum circuits, each corresponding to the element in the upper diagonal of S, as detailed in appendix C. Note that the minimization of the quadratic form is exactly achieved by calculating the eigenvector corresponding to the lowest eigenvalue of S. In addition, this optimization over the whole $SU(2)$ is a generalization of other sequential optimizers [38–40], which optimize only a part of $SU(2)$. To clarify this point, we show they can be derived from our general framework in section 2.2. Notice that a *special* FQS, which applies to the objective functions in a special form, was proposed for time-evolving simulation [14]. Our formulation is also regarded as an extension of that special FQS.

Since FQS can select the optimal gate from the whole $SU(2)$ for minimizing the energy expectation in equation (3), it can incorporate the multi-parameter correlation in $SU(2)$ and thus achieves better performance for optimizing PQCs than other sequential optimizers. Rotoselect, Rotosolve, and their variants [38, 39, 43–46] can be used to optimize a general single-qubit gate by first decomposing the multi-parameter gate using ZYZ-decomposition [47] to obtain three single-parameter gates. In Rotosolve/select and their variants, each parameter of the decomposed gate is optimized locally in contrast to FQS and Fraxis, where a general single-qubit gate is decomposed into two Fraxis gates defined as $R_{\boldsymbol{n}}(\pi)$. Fraxis [40] updates these gates individually but simultaneously optimizes two parameters within the gate.

2.2. Unification of sequential quantum optimizers

FQS generalizes all known sequential optimizers for single-qubit gates of PQCs, such as, Rotosolve (NFT), Rotoselect, Fraxis; those methods can be regarded as special cases of FQS.

For NFT, a single-qubit gate is restricted to a fixed axis $\boldsymbol{m}$ such as $U_{\mathrm{NFT}} := R_{\boldsymbol{m}}(\psi)$. Then the corresponding objective function in quadratic form is

$$\mathrm{tr}\left[\rho'_{\mathrm{in}} U^{\dagger}_{\mathrm{NFT}} H' U_{\mathrm{NFT}}\right] = \boldsymbol{c}^{\top} \begin{pmatrix} S_{00} & \vec{S}_0 \cdot \boldsymbol{m} \\ \vec{S}_0 \cdot \boldsymbol{m} & \boldsymbol{m}^{\top} \tilde{S} \boldsymbol{m} \end{pmatrix} \boldsymbol{c}, \tag{8}$$

where $\boldsymbol{c} := (\cos\psi/2, \sin\psi/2)^{\top}$ and $\vec{S}_0 := (S_{01}, S_{02}, S_{03})$. $\tilde{S}$ denotes the lower right 3×3 part of the S matrix. The derivation of the quadratic form is detailed in appendix B. The real symmetric matrix in equation (8) can be regarded as a contraction of the FQS matrix S with respect to the rotational axis $\boldsymbol{m}$ reducing its degree of freedom to three. In Rotoselect, three contracted matrices in equation (8) are constructed for $\boldsymbol{m} \in \{(1,0,0), (0,1,0), (0,0,1)\}$, and the lowest eigenvalue is selected after three separate diagonalization procedures. Since the three matrices share S_{00}, the total number of circuit evaluations can be reduced to seven.

As for Fraxis, the target gate $U_{\mathrm{Fraxis}} := R_{\boldsymbol{n}}(\pi)$ is simply expressed by the quaternion $\boldsymbol{q} = (0, \boldsymbol{n})$. Thus, substituting $\boldsymbol{q} = (0, \boldsymbol{n})$ into the objective function, we can reproduce the previous results derived in [40] as follows

$$\mathrm{tr}\left[\rho'_{\mathrm{in}} U^{\dagger}_{\mathrm{Fraxis}} H' U_{\mathrm{Fraxis}}\right] = \boldsymbol{n}^{\top} \tilde{S} \boldsymbol{n}. \tag{9}$$

Note that the θ-Fraxis of [40], in which the rotation angle is fixed to arbitrary values θ instead of π, can be regarded as minimizing equation (5) for $\boldsymbol{q}' := (q_1, q_2, q_3) \in \mathbb{R}^3$ under the constraint $|\boldsymbol{q}'|^2 = 1 - q_0^2$, which results in solving simultaneous equations rather than diagonalization.

It is worth noting that the required number of circuit evaluations for each sequential quantum optimizer coincides with the degrees of freedom for the real-symmetric matrix of respective methods, i.e. $3 = 1{+}2$ circuit evaluations for NFT/Rotosolve, $6 = 1{+}2{+}3$ circuit evaluations for Fraxis, and $10 = 1{+}2{+}3{+}4$ circuit evaluations for FQS.

2.3. Barren plateaus in sequential quantum optimizers

We have shown that the energy landscape of all sequential optimizers can be derived from the eigenvalues of the S matrices whose elements are computed from circuit evaluations of slightly modified PQCs. Sequential quantum optimizers can effectively obtain a better single-qubit gate minimizing the cost function as long as the eigenvalues of S are *significantly far* from degenerate. Hence, it is useful to quantify degeneracy of S for evaluation of performance of sequential quantum optimizers. To this end, we introduce a centered matrix $S_{\mathrm{c}}^{(p)}$ defined as,

$$S_{\mathrm{c}}^{(p)} := S^{(p)} - \frac{\mathrm{tr}\left[S^{(p)}\right]}{p} I_{p\times p}, \tag{10}$$

where $I_{p\times p}$ denotes a $p \times p$ identity matrix. Here, $S^{(p)}$ is a $p \times p$ real-symmetric matrix for each sequential quantum optimizer, and the degrees of freedom of target single-qubit gate is $p-1$ e.g. $p=2$ (Rotosolve/NFT), $p=3$ (Fraxis), and $p=4$ (FQS). Henceforth, we omit the superscript p for simplicity because it is clear from the context. Since the mean of the eigenvalues of S_{c} is zero, the spectral radius of S_{c} indicates the spread of the spectrum of S from the mean of eigenvalues of S. More precisely, defining $\hat{\lambda}(S)$ as the eigenvalue of S with the largest distance from $\mathrm{tr}[S]/p$, which is the mean of eigenvalues of S, the spectral radius r of S_{c} is equivalent to

$$r(S_{\mathrm{c}}) \equiv \left|\hat{\lambda}(S) - \frac{\mathrm{tr}[S]}{p}\right|. \tag{11}$$

In the following, we provide two theorems on the behavior of the spectrum of each real-symmetric matrix in equations (5), (8) and (9) on the condition that the parameters in U_1 and U_2, as in figure 2, are randomly initialized, which is standard for studying barren plateaus in gradient-based methods [31, 33]. The proof of the theorems are proved in appendix D.

2.3.1. Sequential quantum optimizers catching barren plateaus

An upper bound of equation (11) that decays exponentially with the number of qubits hints at the existence of flat energy landscape (barren plateaus). Indeed, we reveal that the flat landscape can happen under similar conditions assumed in [31] i.e. if the circuits become sufficiently deep to achieve unitary 2-design, the second moment of the spectral radius of the centered matrix shows exponential shrink regardless of the global or local cost.

Theorem 1. *Suppose that the quantum circuits U_1 and U_2 are randomly and independently generated. If either U_1 or U_2 forms a unitary t-design with $t \geqslant 2$, the second moment of the spectral radius r of the centered matrix equation (10) is upper bounded as,*

$$\mathbb{E}_{U_1,U_2}\left[r(S_{\rm c})^2\right] \leqslant \begin{cases} \frac{p^2\,{\rm tr}\left[H^2\right]\Delta\rho_{{\rm in},2}}{2(2^{2n}-1)} + \frac{\Delta\rho_{{\rm in},2}}{4p(2^{2n}-1)}\sum_{\mu,\nu=0}^{p-1}\left(p(-1)^{1-\delta_{\mu\nu}}-2\right){\rm tr}\left[\mathbb{E}_{U_2}\left[H'\varsigma_\mu\varsigma_\nu^\dagger H'\varsigma_\nu\varsigma_\mu^\dagger + {\rm h.c.}\right]\right] \\ \frac{p^2\,{\rm tr}\left[\rho_{\rm in}^2\right]\Delta H_2}{2(2^{2n}-1)} + \frac{\Delta H_2}{4p(2^{2n}-1)}\sum_{\mu,\nu=0}^{p-1}\left(p(-1)^{1-\delta_{\mu\nu}}-2\right){\rm tr}\left[\mathbb{E}_{U_1}\left[\rho_{\rm in}'\varsigma_\mu^\dagger\varsigma_\nu\rho_{\rm in}'\varsigma_\nu^\dagger\varsigma_\mu + {\rm h.c.}\right]\right] \end{cases}, \tag{12}$$

where the first case corresponds to U_1 being a t-design, and the second case corresponds to U_2 being a t-design. Here, $\mathbb{E}_{U_1,U_2}[\cdot]$ is defined as the expectation over the random quantum circuits U_1 and U_2, $\delta_{\mu\nu}$ denotes the Kronecker delta, h.c. means the Hermite conjugate of the preceding term, and $\Delta\rho_{{\rm in},2}$, ΔH_2 are defined as

$$\Delta\rho_{{\rm in},2} := {\rm tr}\left[\rho_{\rm in}^2\right] - \frac{1}{2^n},$$
$$\Delta H_2 := {\rm tr}\left[H^2\right] - \frac{{\rm tr}^2[H]}{2^n}. \tag{13}$$

In particular, in the case of $p = 2$ (Rotosolve/NFT), the right hand side of equation (12) is equivalent to the variance of gradient of the objective function with respect to the rotation angle in a target gate, which is calculated in [31]. Therefore, we can rewritten equation (12) with $p = 2$ as follows:

$$\mathbb{E}_{U_1,U_2}\left[r\left(S_{\rm c}^{(2)}\right)^2\right] \leqslant {\rm Var}\left[\frac{\partial}{\partial\theta}\langle H\rangle\right],$$

where $\langle H\rangle$ denotes the energy expectation for the quantum state in figure 2 with $R = {\rm e}^{-{\rm i}\theta\boldsymbol{n}\cdot\vec{\sigma}}$. This theorem means that if either U_1 or U_2 has sufficient expressibility i.e. unitary 2-design over n-qubit, the spectrum of S matrix concentrates on a single value and its deviation is exponentially small with respect to the number of qubits. As a result, this spectrum concentration implies that the energy of the output quantum state becomes insensitive exponentially on the selection of single-qubit gates.

2.3.2. Sequential quantum optimizers avoiding barren plateaus

Although theorem 1 assumes the relatively deep circuit forming unitary t-design with $t \geqslant 2$, we are interested in shallower circuits due to the existence of noise-induced barren plateau. Here, we show that, in contrast to theorem 1, there are shallow circuits avoiding the exponential shrink of the spectral radius when the target Hamiltonian is local.

To this end, let us consider an n-qubit alternating layered ansatz U [33, 42] with m-qubit parametrized unitary blocks, as depicted in figure 3(a). Each block consists of parametrized single-qubit gates and parameter-free gates. Suppose the ansatz consists of L layers, and each layer contains ξ blocks (i.e. $n = \xi m$). We define S_k $(k = 1, 2, \cdots, \xi)$ as the m-qubit subsystem on which the kth block from the top in the final layer acts. In the following, we focus on a block W in the lth layer and a parameterized single-qubit gate R in the block as in figure 3(b). Moreover, we define the forward light-cone $\mathcal{L}$ of the block W as a series of gates that has at least one input qubit causally connected to the output qubits of W. As well, the backward light-cone $\mathcal{L}_{\rm B}$ of W is defined in the reverse direction of $\mathcal{L}$.

Theorem 2. *Suppose that the whole quantum circuit U is an n-qubit alternating layered ansatz with m-qubit blocks as described in figure 3. Here, we focus on a block W in lth layer and a parameterized single-qubit gate R in the block. We assume that the quantum circuits $W_{\rm A}, W_{\rm B} \subset W$, which are located after and before the target gate respectively, and the other blocks form a local 2-design independently. In addition, the Hamiltonian H is assumed to be m-local such as*

$$H = c_0 I^{\otimes n} + \sum_i c_i h_i,\ c_0, c_i \in \mathbb{R},\ (\exists i, c_i \neq 0), \tag{14}$$

where h_i is a tensor product of Pauli matrices that acts non-trivially on at most m-qubit. (The same assumption was used in [33].) Then, the second moment of the spectral radius r of the centered matrix equation (10) satisfies

$$\mathbb{E}_{U_1,U_2}\left[r(S_{\rm c})^2\right] \geqslant \frac{(p+2)(p-1)\,2^{m(l+1)-1}}{p\left(2^{2m}-1\right)^2\left(2^m+1\right)^{L+l}}\sum_{i\in i_{\mathcal{L}}}\sum_{\substack{(k,k')\in k_{\mathcal{L}_{\rm B}} \\ k'\geqslant k}} c_i^2\,\epsilon(\rho_{k,k'})\,\epsilon(h_i), \tag{15}$$

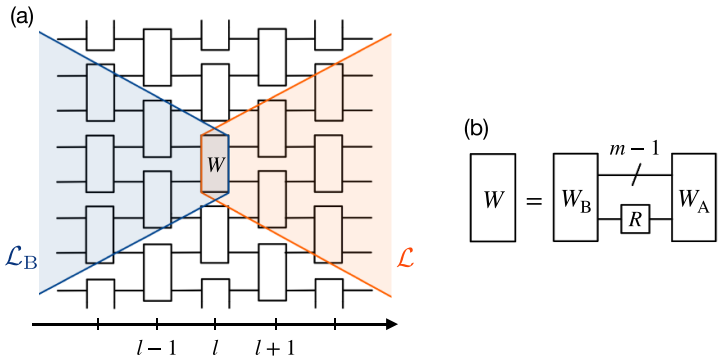


Figure 3. Schematic diagram of the alternating layered ansatz with m-qubit untiary blocks. In theorem 2, we focus on a single-qubit gate R in a block W of the lth layer as depicted in (b), where W_A and W_B are a set of unitary gates located after and before the target single-qubit gate. $\mathcal{L}$ ($\mathcal{L}_B$) denotes the forward (backward) light-cone with respect to W. Here, U_1 in figure 2 includes the gates in $\mathcal{L}$ and W_A, and U_2 includes the other gates.

where L denotes the total number of layers. Here, $i_{\mathcal{L}}$ is the set of i indices whose associated operators h_i act on qubits in the forward light-cone $\mathcal{L}$ of W, and $k_{\mathcal{L}_B}$ is the set of k indices whose associated subsystems S_k are in the backward light-cone $\mathcal{L}_B$ of W. The quantum state $\rho_{k,k'}$ is the reduced density matrix of the input state ρ_{in} on $S_kS_{k+1}\cdots S_{k'}$, and the function $\epsilon(M)$ for a matrix M is defined as $\epsilon(M) = D_{HS}(M, \mathrm{tr}(M)\mathbf{1}/d_M)$ where D_{HS} is the Hilbert–Schmidt distance and d_M is the dimension of the matrix M.

We prove this theorem by firstly deriving the relation between the lower bound of the spectral radius and the variance of the gradient regarding θ in the case $R = \mathrm{e}^{-\mathrm{i}\theta \boldsymbol{n}\cdot\vec{\sigma}}$, under the condition of theorem 2. More precisely, we prove the following relation:

$$\mathbb{E}_{U_1,U_2}\left[r(S_c)^2\right] \geqslant \frac{(p+2)(p-1)}{4p} \times \mathrm{Var}\left[\frac{\partial}{\partial\theta}\langle H\rangle\right].$$

Then, combining the previous result in [33], we complete the proof of theorem 2; see appendix D. This theorem means the lower bound of the second moment of the spectral radius does not vanish exponentially fast when we employ shallow L, which leads to similar properties as the gradient-based counterparts [33]. More precisely, if at least one term $c_i^2\epsilon(\rho_{k,k'})\epsilon(h_i)$ vanishes no faster than $\Omega(1/\mathrm{poly}(n))$, and if the number of layers L is $\mathcal{O}(\log(n))$, then

$$\mathbb{E}_{U_1,U_2}\left[r(S_c)^2\right] = \Omega\left(\frac{1}{\mathrm{poly}(n)}\right) \tag{16}$$

holds. This guarantees trainability of sequential quantum optimizers in the beginning of optimization. On the other hand, if at least one term $c_i^2\epsilon(\rho_{k,k'})\epsilon(h_i)$ vanishes no faster than $\Omega(1/2^{\mathrm{poly}(\log(n))})$, and if the number of layer L is $\mathcal{O}(\mathrm{poly}(\log(n))$ then,

$$\mathbb{E}_{U_1,U_2}\left[r(S_c)^2\right] = \Omega\left(\frac{1}{2^{\mathrm{poly}(\log(n))}}\right) \tag{17}$$

holds. This suggests a transition region of trainability where the spread of the spectrum decays faster than polynomial but slower than exponential. Note that Cerezo *et al* [33] showed a detailed analysis of variance of gradient with respect to angles of fixed-axis rotation gates. In contrast, for a global cost function, the spectral radius is expected to shrink exponentially even on a constant-depth alternating layered ansatz in line with the gradient-based methods as [33].

There is a new finding in equation (15) which is not available from the previous study on the gradient; the lower bound depends on the parameter p. We can see this from the simplest case when $p = 2$ for Rotosolve (NFT). In this case, the lower bound of the spectral radius appears to be twofold of that of the gradient-based method in [33]. This is probably because the spectral radius measures all possible energy variations that can occur by replacing the target gate, while the gradient-based analysis focuses on only a local response along a slight move in the parameter space.

Also, the lower bound implies the larger the value of p the better the chance to escape from barren plateaus. Although the values of p are limited to 2 (NFT), 3 (Fraxis), and 4 (FQS) for single-qubit gates in this study, our spectral radius analysis can be extended to a multi-qubit gate with $p > 4$ that seems to grow at least polynomially in the number of qubits. Such an extension of our method is actually an important open

problem, but it would be achieved by imposing an appropriate structure of parameterized gate to the method in the previous paper [41] or a more detailed analysis of the unified expressions of sequential optimizers.

3. Numerical simulations

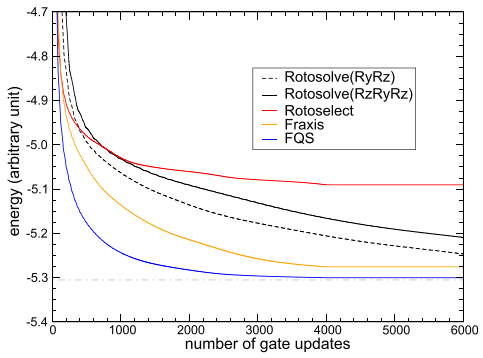
Comparing with both other sequential quantum optimizers and a gradient-based optimizer, we show the advantages of FQS in VQE for the mixed field Ising model and fidelity maximization, where the former has a local Hamiltonian as the cost function while the latter is a global cost. In addition, we demonstrate that our method can avoid the barren plateau (more precisely, the exponential concentration of spectrum of S) by numerically evaluating the second moment of spectral radius. The detailed setup for the numerical simulations is provided in appendix E. Note that in the case of sequential optimization, the order of gate (parameter) updates is arbitrary. It is not obvious how the results depend on the update order. We tested three update protocols (vertical, horizontal, random updates) using a 5-qubit Ising model before the main experiments. Refer to the appendix G and figure 10 for the detailed update protocol and benchmark results. The results showed that vertical and horizontal update have comparable performance, while random update converged faster, but led to the local minima. Henceforth, the results in the rest of this manuscript are calculated based on vertical update. In this paper, an updating cycle of all single qubit gates is referred to as a sweep.

3.1. Mixed field Ising model

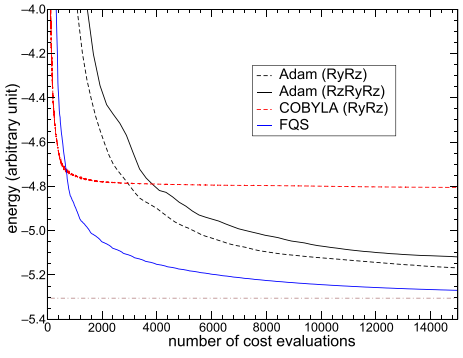
Figure 4 shows averaged trajectories of independent VQE simulations for 5-qubit (in the panels (a) and (b)) and 20-qubit (in the panels (c) and (d)) mixed-field Ising model based on the 5-layer alternating layered ansatz. While we employed statevector simulator in the 5-qubit experiments, in the 20-qubit experiments the expectation of Hamiltonian was evaluated with $61\,440 = 60 \times 1024$ shots in total. In figure 4(a), we compare the convergence efficiency of FQS with those of the other sequential quantum optimizers with respect to the number of gate updates under an ideal condition without statistical error, namely shot noise. Although the number of circuit evaluations to update a single gate is not necessarily consistent among the optimizers, the number of gate updates may be a fair index because the practical wall times for a single gate update by these optimizers would be comparable if parallel circuit evaluations were allowed.

Considering the gate expressibility, it may appear to be reasonable that FQS reached the better solution than Rotosolve and Fraxis. We note, however, the efficiency of FQS is not necessarily trivial. To verify it, we decomposed a general single-qubit gate, which we term an FQS gate, into equivalent three fixed-axis rotation gates $R_z(\phi)R_y(\vartheta)R_z(\lambda)$ and sequentially optimized the three gates by Rotosolve. In this case, the optimization turned out to be far slower not only than FQS but also than Fraxis. The ansatz with the three-gate decomposition was also slower than the *RyRz* ansatz in optimization. Actually, this deceleration of the *RzRyRz* ansatz compared to *RyRz* ansatz is consistent with the known dilemma; higher expressibility leads to lower trainability [48]. It is a contrast that FQS can maintain high optimization efficiency regardless of its high expressibility. We can clearly attribute this efficiency of FQS to the incorporated parameter correlation within a single qubit gate, which has three degrees of freedom. This insight is consistent with the fact that Fraxis also outperforms the sequential optimization for the *RyRz* ansatz by Rotosolve, because Fraxis incorporates correlation of two degrees of freedom. As shown in figure 4(c), we also emphasize that these advantages of FQS are still distinct even under a practical condition with finite number of shots and scalable up to as large as 20 qubits. Further investigation has shown that shot noise causes FQS to lose its advantage over Fraxis and Rotosolve when comparing the converged VQE energy [49]. In fact, figure 4(c) shows that FQS and Fraxis give comparable results in terms of convergence energy, but still have an advantage over Rotosolve. Furthermore, with respect to the speed of convergence, FQS has a distinct advantage. Hence, we confirmed that FQS is able to demonstrate its performance under shot noise well enough for practical use.

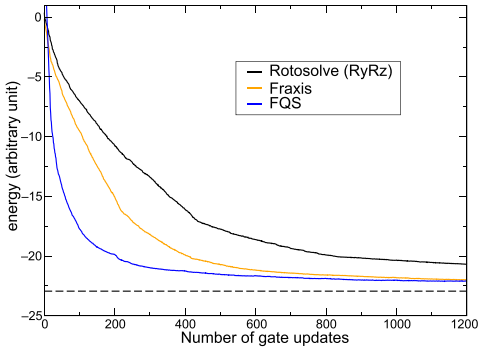
The comparison of performance between FQS and classical optimizers is presumably in great demand. Among the classical optimizers, ADAM and COBYLA are employed here as representative gradient-based and gradient-free optimizers, respectively. COBYLA has been widely employed in VQA, particularly for computation on a real device. Adam can perform efficient stochastic optimization under statistical errors that come from the finite number of measurements. Figures 4(b) and (d) show a comparison of optimization efficiency between FQS and the classical optimizers [50] where the horizontal axis represents the number of expectation evaluations. Note that it is not straightforward to compare FQS and classical optimizers in a fair manner, because their apparent performances vary depending on the assumed hardware. Although parallel computing is principally possible for the gradient-based optimizers, the required number of the cost evaluation for each optimization step is $\mathcal{O}(D)$, where D is the total number of gates in ansatz. This is contrast to FQS that requires a constant number of the cost evaluation i.e. at most 10 circuits for a single gate update. As the number of parameterized gates increases, we did not find parallel computing practical for the



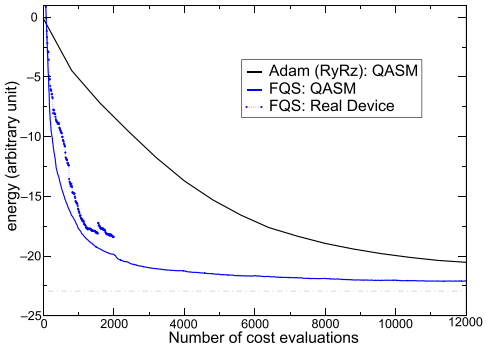
(a) Comparison with the other sequential optimizers (5-qubit)



(b) Comparison with classical optimizers (5-qubit)



(c) Comparison with the other sequential optimizers (20-qubit)

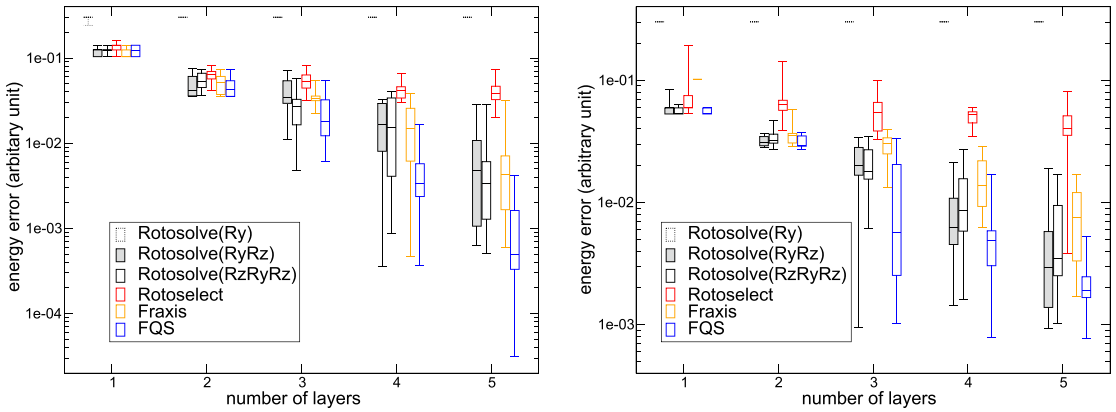


(d) Comparison with a gradient-based optimizer (20-qubit)

Figure 4. Optimization trajectories of the VQE resulting energy for 1-dimensional mixed field Ising model on the 5-layer alternating layered ansatz. The top and bottom panels show the results in a 5-qubit system with the statevector simulator and 20-qubit system with the QASM simulator and a real device(ibm_kawasaki), respectively. The trajectories in each panel are average over 20 (10) independent runs whose initial states were randomly generated in 5-qubit (20-qubit) simulations except for that from the real device where the single trajectory is presented. The Alternating Layered ansatz with five layers was employed for all the simulation except for that on the real device where the two layers was adopted. Left panels (a) and (c): comparison with the other sequential optimizers, where the horizontal axis represents the number of gate updates. Right panels (b) and (d): comparison of FQS with classical optimizers, where the horizontal axis represents the number of expectation evaluations. The learning rate parameter of 0.1 was employed for the Adam optimizer. The horizontal dash-dotted line in gray color represents the exact ground energy.

gradient-based optimizations. Henceforth we employ the number of expectation evaluations as the measure for comparison evaluations, which is not consistent with that of figures 4(a) and (c). Figure 4(b) clearly shows the FQS advantage over COBYLA and Adam optimizer. Although COBYLA achieves the fastest convergence, the optimisation is often trapped local minima. The slower convergence of the Adam with the *RzRyRz* gates can be attributed to the same reason for Rotosolve in figure 4(a). The efficiency with Adam optimization appears to be diminished when the system size increases to 20 qubits and simulation on the quantum assembly language (QASM) simulator is employed, which is contrast to FQS that maintains high efficiency (figures 4(b) and (d)). We further carried out VQE calculations with FQS on a real quantum device, ibm_kawasaki, where the Hamiltonian was evaluated with a consistent number of shots in the QASM simulation. Because of severe limitations in real device usage, we employed a two-layer ansatz as well as usage of the optimal parameter configuration for a single-qubit gates configuration proposed in [49], which is inconsistent with other results from simulators using five layers. A shallower PQC generally leads to a faster convergence rate of VQA, while the energy level at the convergence destination is usually worse. Therefore, this result from the real device cannot be directly compared to those from simulator. Despite the inclusion of noise, however, the result from the real device shows that the convergence speed is comparable to that of the 5-layer QASM simulator. Furthermore, the convergence destination is at the same level as Adam. This indicates that FQS performs well under noise on the real device.

Figure 5 shows the optimized energy after 100 sweeps based on (a) alternating layered ansatz in figures 9(a) and (b) grouped-layer ansatz with cyclic entangler in the 5-qubit mixed field Ising model with statevector simulator in figure 9(b). Although the 100 sweeps may not necessarily be sufficient for rigorous convergence, the energy update by single sweep is too small to expect the drastic improvement by further iterations. Figure 5 provides several insights. Generally, the performance of optimization methods may become more distinct as the circuit layer increases. FQS consistently showed the best performance among all



(a) Alternating layered ansatz (b) Grouped-layer ansatz with cyclic entangler

Figure 5. Boxplots of the resulting energy of VQE for 5-qubit 1-dimensional mixed field Ising model after 100 sweeps. The 20 independent VQEs were conducted from randomly-generated initial states. The vertical axis represents the difference between the obtained energy and the exact ground energy.

the sequential optimizers on all the ansätze (alternating layered, cyclic, and ladder entanglers, see also figure 11 in appendix H). Taking the higher expressibility and correlation among parameters of an FQS gate into account, the systematic difference in performance according to the number of parameters to be optimized apparently seems to be promising. Indeed, we confirmed a systematic improvement associated with the degrees of freedom of the target gate i.e. FQS optimizes three degrees, Fraxis optimizes two degrees and Rotosolve optimizes one degree. On the other hand, it should be also noted that the Rotosolve applications to a series of the fixed-axis gates (*RyRz* and *RzRyRz*) were not better than FQS even though they have equivalent expressibility. In some cases, on the contrary, Rotosolve applying to a series of *RyRz* gates showed better performance than Rotosolve to *RzRyRz* gates that possess full expressibility for a single-qubit as well as an FQS gate. This result suggests that the higher circuit expressibility alone may not be sufficient for successful optimization, and on the contrary, in some cases it may disturb further optimization if correlation among parameters is not considered. Note that barren plateaus do not account this hindrance because the FQS application that has equivalent expressibility to *RzRyRz* showed better performance. Hence, we suppose that Rotosolve optimization for *RzRyRz* gates is likely to be trapped at local minima and/or saddle points, while FQS may be more resilient due to its incorporating parameter correlation.

3.2. Fidelity maximization

As another example to verify FQS performance, we conducted fidelity maximization on 5-qubit system, where the infidelity with a reference state was regarded as the cost. Figure 6 shows the results after 100 sweeps of 40 independent runs with statevector simulator. We confirmed that FQS also showed better performance than the others, where the advantage was more distinct as the number of layers L increased. This observation is consistent with the previous results on a local Hamiltonian.

3.3. Spectral radius and noise-free barren plateau

To verify our theorem on the spectrum of FQS matrix on an alternating layered ansatz, we evaluated the second moment of the spectral radius of the centered FQS matrix with use of statevector simulator in comparison of the global cost and local cost. When we employed the infidelity as a global cost function, the second moment was independent of the number of layers, but exponentially scaled according to the number of qubits as shown in figure 7(a). Since the reference states generated with random unitary U_{Haar} from the unitary group with respect to the Haar measure are written as $|\psi_{\mathrm{ref}}\rangle = U_{\mathrm{Haar}}|0\rangle^{\otimes n}$, the fidelity can be regarded as the projection measurement on $|0\rangle^{\otimes n}$ with the input ansatz appended a randomly-generated unitary:

$$|\langle\psi_{\mathrm{FQS}}|\psi_{\mathrm{ref}}\rangle|^2 = \langle\psi_{\mathrm{FQS}}|U_{\mathrm{Haar}}|0\rangle\langle 0|^{\otimes n}U_{\mathrm{Haar}}^{\dagger}|\psi_{\mathrm{FQS}}\rangle. \tag{18}$$

As a result, the condition employed in theorem 1 holds, which is consistent with the present experiments. In the case of the local cost function, the exponential decay of the second moment can be confirmed in the large limit of the number of layers in figure 7(b). However, the second moment exhibited the transition from a constant value to exponentially small values as the number of layers increases. Therefore, the sequential quantum optimizers are able to circumvent noise-free barren plateau using an alternating layered ansatz with

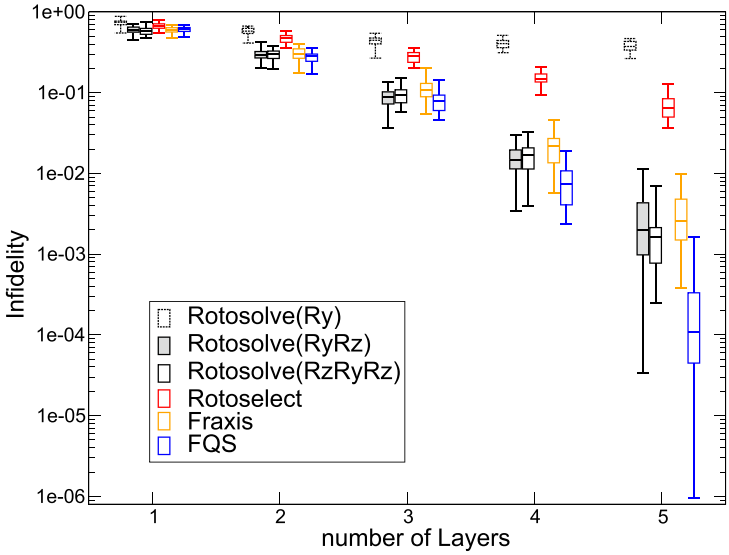
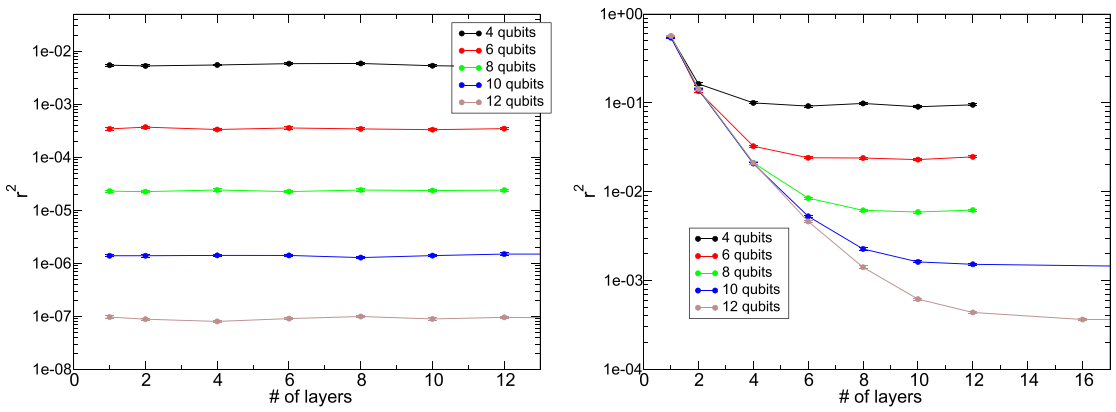


Figure 6. Boxplots of the VQE results for fidelity maximization as a function of the number of circuit layers. The alternating layered ansatz in figure 9(a) was employed. The box plot denotes quantiles consisting of independent 40 resulting energies with randomly-generated initial parameter set and target states taken from the Haar distribution.



(a) Global cost (b) Local cost

Figure 7. The second moment of spectral radius of the centered FQS matrix. The FQS matrix is evaluated for the single-qubit gate acting on the first qubit in the first layer of the alternating layered ansatz. The second moment is evaluated with 1000 samples based on the randomly initialized circuits with respect to the parameterized single-qubit gates. We employed the infidelity with the state generated using Haar measure on n-qubit unitary as the global cost. For the local cost function, we employed an expectation value of the Hamiltonian $H = Z \otimes I^{\otimes n-1}$.

limited number of layers for local cost functions. We remark that although each block in the alternating layered ansatz used in the present experiments does not form unitary 2-design, the results seem to be consistent with the consequence of theorem 2. Therefore, the unitary 2-design in the block and even alternating layered ansatz may not be necessarily required to circumvent barren plateau.

Furthermore, we also observed the transition region of the second moment of the spectral radius with a different type of alternating layered ansatz (see figure 12 in appendix H). We also note that, at present, barren plateau has been analytically proven assuming the randomly initialized conditions. Although it is the case at the beginning of the optimization, the randomness does not stand anymore during the optimization, and thus it is not trivial how the optimization proceeds after random initializations. For further understanding, we carried out the additional optimizations with the 5-layer alternating layered ansatz varying the number of qubits.

Figure 8 shows the resulting errors for the global cost become exponentially larger as the number of qubits increases, which is in line with the spectral radius in the beginning of the optimizations as in figure 7(a). In contrast, the results for local cost did not exhibit the exponential deterioration of trainability, which is also consistent with figure 7(b), where the relative error balanced after 6-qubit. We expect that this

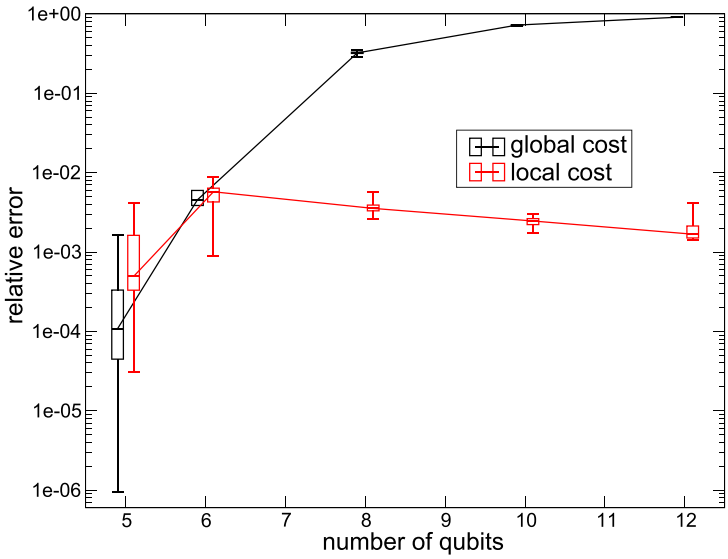


Figure 8. Boxplots for the relative errors of the obtained cost measured from the exact solution after 100 sweeps as a function of the number of qubits. As for the global cost, we executed the fidelity maximization as described in section 3.2, while the mixed field Ising Hamiltonian in equation (E.1) was employed as the local cost. For both VQEs with statevector simulator, the 5-layer alternating layered ansatz in figure 9(a) was employed. The box plots denote quantiles consisting of independent 20 optimizations, and the medians are connected by the solid line.

error may increase again, as the number of qubits increases beyond a certain threshold, because a constant layer circuit does not have sufficient expressibility for the target states, even though the optimization proceeds to some extent in the beginning. For better scalability, thus, it is required to use a high expressive circuit, e.g. alternating layered ansatz with high expressive blocks [42], keeping the number of layers small. Since our method utilizes the full expressibility of single-qubit gates to optimize the circuit structure holding the number of layers, we believe that FQS is effective to meet this requirement.

4. Conclusions and discussion

We proposed a new algorithm for VQAs combined with the circuit structure optimization based on analytically optimal selection of a single-qubit gate. We have shown that the expectation of an observable on a quantum state prepared by a PQC can be rewritten as the solvable quadratic form on parameters of a single-qubit gate in the PQC: our algorithm utilizes the matrix diagonalization based on this quadratic form. The matrix diagonalization framework has also revealed the hierarchical relation of sequential quantum optimizers in the degrees of freedom of simultaneous optimization for the single-qubit gate i.e. Rotosolve (NFT) $\leqslant$ Rotoselect $\leqslant$ Fraxis $\leqslant$ FQS. Moreover, by introducing the spectral radius as a measure to evaluate the performance of sequential optimizers (including the existing methods), we rigorously proved the exponential concentration of spectrum of the matrix associated with the quadratic form if the circuit is too deep. On the other hand, we also proved the possibility to avoid the exponentially sharp concentration by supposing a local cost function and an alternating layered ansatz. Furthermore, numerical experiments show that sequential optimization works well for large systems.

We also remark that the hierarchical relation contains Rotoselect, which was originally proposed based on a concept of circuit structure optimization [39]. As mentioned above, the only way to maintain trainability against both barren plateau and noise-induced barren plateau is to employ shallow quantum circuits although it sacrifices the expressibilily. The circuit structure optimization fills this requirement; the ansatz is modified adapting to the cost function through the cost optimization such that the circuit can approximate the target quantum state well even with limited circuit depth. Actually, various types of circuit structure optimizations have been developed [3, 39, 51–63]. Unlike Rotoselect and our method that optimize the fixed-depth circuit structure, several circuit structure optimizations vary the circuit depth during their optimization [3, 56, 64]. For instance, since variable Ansatz (VAns) algorithm [56] consists of heuristic protocols such as insertion and simplification, it is difficult to estimate the entire computational cost. More precisely, this type of circuit structure optimizations require additional iterations to empirically confirm whether a candidate of the additional gates decreases the target cost function, which causes a large computational overhead. Since our method can certainly decrease the cost function with a few step while

fixing the circuit depth, it would provide an efficient way to check which next operations are better as well as [65], thereby suppressing the total computational cost of the variable-depth structure optimizations. Here, we make additional remarks on the hierarchical relation including out method. Rotoselect allows for the optimal selection of a (single-qubit) rotational gate among $R_x(\theta)$, $R_y(\theta)$, and $R_z(\theta)$ while the optimal rotational angle is also estimated. However, it has two drawbacks: the rotation axis is from a finite and discrete gate set, and each parameter of PQCs is updated locally. These make it prone to local optima as we numerically confirmed it in this study. To deal with the drawbacks, *Free-axis selection* (*Fraxis*) was proposed based on the representation of a single-qubit gate whose rotational axis is arbitrary 3-dimensional vector, but the rotational angle is fixed to π [40]. Since FQS is an upper compatible with these quantum optimizers, it can be also captured as the extension of the circuit structure optimization. Indeed, we demonstrated that FQS achieved higher expressibility than Rotoselect and Fraxis under constant circuit layers. In the numerical experiments, we also confirmed FQS achieved a good balance between trainability and expressibility by circuit structure optimization and incorporation of the intra-gate parameter correlation. Indeed, we demonstrated that these preferred features allow FQS to outperform a gradient-based optimizer as well as the other sequential optimizers.

In our analysis for the scalability of sequential optimization, it is notable that the degree of freedom of the target single-qubit gate affects the lower bound of the spectral radius, as in equation (15), which determines how easy the sequential optimization can avoid barren plateaus. Utilizing parameterized multi-qubit gates that have more degrees of freedom, such as proposed in [41], may result in better optimization strategies not achievable by gradient-based methods. We leave this as an interesting future work.

Our study sheds light on the remarkable performance of the sequential optimizers which have surprisingly not been as popular as the gradient-based optimizers. This may be caused by conflicting beliefs regarding their performance, particularly in the uncertainty of the robustness against barren plateau and in the issue on local minima [41, 66]. The former is because there has been no analytical study for barren plateau, while the latter has been examined only with Rotosolve that does not consider the parameter correlation. Our work has successfully provided theoretical and experimental evidences clearing out uncertainties and issues of the sequential optimizers. Actually, our experiments imply that simultaneous analytical optimizations for multiple degrees of freedom are effective in avoiding local minima. Because it is reasonable to expect that this feature holds in more general local optimization beyond the single-qubit gate optimization, although simple we think this numerical result is an important example to motivate us to further explore local optimization for quantum circuits.

Since FQS advances the circuit structure optimization from a heuristic to an analytical one, it may provide a solution to a critical problem of VQA, that is, circuit design. We hope the results are instrumental for promoting the use of sequential optimizers in VQAs.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

K W would like to thank National Institute of Information and Communications Technology (NICT) for the support through Young Researchers Lab. H C W was supported by JSPS Grant Numbers 20K03885, 23K03266 and the MEXT Quantum Leap Flagship Program Grant Number JPMXS0118067285 and JPMXS0120319794. A part of this work was performed for Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), 'Promoting the application of advanced quantum technology platforms to social issues' (Funding agency :QST). We would like to thank Dr Michihiko Sugawara, Dr Yu-ya Ohnishi, Dr Eriko Kaminishi, and Dr Naoki Yamamoto for technical discussion.

Code availability

The FQS code for this article is available in Python using Qiskit version 0.36 [67] and can be found on GitHub at https://github.com/KQCC-Chemistry/SeQpt.

Author contributions

K W conceived the original study and conducted the theoretical analysis and numerical simulations by support R R, Y S, and H C W. H C W supervised the project and organized the algorithm and simulations. All authors carried out the simulations, discussed the results and contributed to the writing of this paper.

Conflict of interest

The authors declare no competing interests.

Appendix A. Quaternion representation for single-qubit gate

A general single-qubit gate is conventionally represented as

$$R_{\boldsymbol{n}}(\psi) := \cos\left(\frac{\psi}{2}\right) I - \mathrm{i}\sin\left(\frac{\psi}{2}\right) \boldsymbol{n}\cdot\vec{\sigma}, \tag{A.1}$$

where I and $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3) = (X, Y, Z)$ denote the 1-qubit identity operator and the Pauli matrices. The parameters $\boldsymbol{n}$ and ψ correspond to a rotational axis and angle in the Bloch sphere, respectively.

Here, we show another way to parameterize the general single-qubit gate based on the well-known relationship between a single-qubit gate and a unit quaternion. Since the rotational axis $\boldsymbol{n}$ is a three-dimensional real unit vector, we can write it in the polar coordinate system with the zenith angle θ and the azimuth angle ϕ as

$$\boldsymbol{n} = \boldsymbol{n}(\theta, \phi) = (\cos\theta, \sin\theta\cos\phi, \sin\theta\sin\phi)\,. \tag{A.2}$$

Substituting equation (A.2) into equation (A.1), we obtain the quaternion representation of a single-qubit gate as

$$R_{\boldsymbol{n}(\theta,\phi)}(\psi) = \boldsymbol{q}(\psi,\theta,\phi)\cdot\vec{\varsigma} \equiv R(\boldsymbol{q})\,, \tag{A.3}$$

where a unit quaternion $\boldsymbol{q} = (q_0, q_1, q_2, q_3)$ (i.e. $\boldsymbol{q} \in \mathbb{R}^4,\ |\boldsymbol{q}| = 1$) is parameterized with (ψ, θ, ϕ) as

$$\begin{aligned} q_0 &= \cos\left(\frac{\psi}{2}\right), \\ q_1 &= \sin\left(\frac{\psi}{2}\right)\cos\theta, \\ q_2 &= \sin\left(\frac{\psi}{2}\right)\sin\theta\cos\phi, \\ q_3 &= \sin\left(\frac{\psi}{2}\right)\sin\theta\sin\phi. \end{aligned} \tag{A.4}$$

Here, $\vec{\varsigma} = (\varsigma_0, \varsigma_1, \varsigma_2, \varsigma_3)$ is an extension of the Pauli matrices defined as

$$\vec{\varsigma} := (I, -iX, -iY, -iZ)\,. \tag{A.5}$$

Equation (A.3) allows us to identify a point $\boldsymbol{q}$ on the three-dimensional spherical surface with a single-qubit gate. Note that if we focus on the conventional single-qubit rotation gate with one parameter θ for the rotational angle, such as $R_x(\theta)$, this gate can be identified with a point on the one-dimensional spherical surface, i.e. the unit circle.

We can decompose a general single-qubit gate $R(\boldsymbol{q})$ into three R_z gates and two $\sqrt{X}$ gates up to global phase [68], that is,

$$R(\boldsymbol{q}) = R_z(\phi)\sqrt{X}R_z(\theta)\sqrt{X}R_z(\lambda)\,, \tag{A.6}$$

where θ, ϕ, λ can be determined from $\boldsymbol{q}$. Since R_z gates on IBM Quantum devices are pulse-operation free, the general single-qubit gate can be implemented with only two times pulse-operation for $\sqrt{X}$.

Appendix B. Derivation of quadratic forms

Substituting the quaternion representation of single-qubit gates $R(\boldsymbol{q}) = \boldsymbol{q}\cdot\vec{\varsigma}$ into the energy (3), we obtain

$$\begin{aligned}\langle H\rangle(\boldsymbol{q}) &= \mathrm{tr}\left[\rho'_{\mathrm{in}}R(\boldsymbol{q})^{\dagger}H'R(\boldsymbol{q})\right] = \sum_{\mu,\nu=0}^{3} q_\mu q_\nu \mathrm{tr}\left[\rho'_{\mathrm{in}}\varsigma_\mu^{\dagger}H'\varsigma_\nu\right]\\ &= \sum_{\mu,\nu=0}^{3} A_{\mu\nu}q_\mu q_\nu, \quad A_{\mu\nu} := \mathrm{tr}\left[\rho'_{\mathrm{in}}\varsigma_\mu^{\dagger}H'\varsigma_\nu\right]\\ &= \frac{1}{2}\left(\sum_{\mu,\nu=0}^{3} A_{\mu\nu}q_\mu q_\nu + \sum_{\nu,\mu=0}^{3} A_{\nu\mu}q_\nu q_\mu\right) = \frac{1}{2}\left(\sum_{\mu,\nu=0}^{3} A_{\mu\nu}q_\mu q_\nu + \sum_{\mu,\nu=0}^{3} A_{\nu\mu}q_\mu q_\nu\right)\\ &= \sum_{\mu,\nu=0}^{3}\frac{A_{\mu\nu}+A_{\nu\mu}}{2}q_\mu q_\nu = \boldsymbol{q}^{\top}S\boldsymbol{q},\end{aligned} \tag{B.1}$$

where the matrix $S := (S_{\mu\nu})$ in the last line is defined by symmetrization of $A := (A_{\mu\nu})$ as follows

$$S_{\mu\nu} := \frac{A_{\mu\nu}+A_{\nu\mu}}{2} = \frac{1}{2}\mathrm{tr}\left[\rho'_{\mathrm{in}}\left(\varsigma_\mu^{\dagger}H'\varsigma_\nu + \varsigma_\nu^{\dagger}H'\varsigma_\mu\right)\right]. \tag{B.2}$$

Here, the S is obviously a real symmetric matrix since $\varsigma_\mu^{\dagger}H'\varsigma_\nu + \varsigma_\nu^{\dagger}H'\varsigma_\mu$ is a Hermitian operator.

We write $\tilde{S}$ as the lower right 3×3 part in S matrix. Then, the quadratic form for Fraxis (more precisely, π-Fraxis algorithm) [40] is represented with $\tilde{S}$. Actually, Fraxis algorithm can deal with the single-qubit gate $U_{\mathrm{Fraxis}} := R_{\boldsymbol{n}}(\pi)$ and the corresponding quaternion is $q = (0,\boldsymbol{n})^{\top}$, which is simply located on the two-dimensional spherical surface. Thus, from equation (B.1) we directly obtain the quadratic form for Fraxis as

$$\begin{aligned}&\mathrm{tr}\left[\rho'_{\mathrm{in}}U_{\mathrm{Fraxis}}^{\dagger}H'U_{\mathrm{Fraxis}}\right] = \boldsymbol{n}^{\top}\tilde{S}\boldsymbol{n}\\ &= \boldsymbol{n}^{\top}\begin{pmatrix}\mathrm{tr}\left(H'\sigma_1\rho'_{\mathrm{in}}\sigma_1\right) & \frac{1}{2}\left[\mathrm{tr}\left(H'\sigma_1\rho'_{\mathrm{in}}\sigma_2\right)+\mathrm{tr}\left(H'\sigma_2\rho'_{\mathrm{in}}\sigma_1\right)\right] & \frac{1}{2}\left[\mathrm{tr}\left(H'\sigma_1\rho'_{\mathrm{in}}\sigma_3\right)+\mathrm{tr}\left(H'\sigma_3\rho'_{\mathrm{in}}\sigma_1\right)\right]\\ \cdot & \mathrm{tr}\left(H'\sigma_2\rho'_{\mathrm{in}}\sigma_2\right) & \frac{1}{2}\left[\mathrm{tr}\left(H'\sigma_2\rho'_{\mathrm{in}}\sigma_3\right)+\mathrm{tr}\left(H'\sigma_3\rho'_{\mathrm{in}}\sigma_2\right)\right]\\ \cdot & \cdot & \mathrm{tr}\left(H'\sigma_3\rho'_{\mathrm{in}}\sigma_3\right)\end{pmatrix}\boldsymbol{n}.\end{aligned} \tag{B.3}$$

As for the NFT [38] and Rotosolve/Rotoselect [39], they can optimize an axis-fixed single-qubit gate $U_{\mathrm{NFT}} := R_{\boldsymbol{m}}(\psi)$, where $\boldsymbol{m}$ is a fixed rotational axis. Substituting the quaternion $q = (\cos\psi/2, \boldsymbol{m}\sin\psi/2)^{\top}$ corresponding to this gate into the quadratic form of FQS, we obtain

$$\begin{aligned}\mathrm{tr}\left[\rho'_{\mathrm{in}}U_{\mathrm{NFT}}^{\dagger}H'U_{\mathrm{NFT}}\right] &= S_{00}\cos^2\frac{\psi}{2} + \sum_{i=1,2,3}2S_{0i}m_i\sin\frac{\psi}{2}\cos\frac{\psi}{2} + \sum_{i,j=1,2,3}\sin^2\frac{\psi}{2}S_{ij}m_im_j\\ &= \begin{pmatrix}\cos\frac{\psi}{2} & \sin\frac{\psi}{2}\end{pmatrix}\begin{pmatrix}S_{00} & \vec{S}_0\cdot\boldsymbol{m}\\ \vec{S}_0\cdot\boldsymbol{m} & \boldsymbol{m}^{\top}\tilde{S}\boldsymbol{m}\end{pmatrix}\begin{pmatrix}\cos\frac{\psi}{2}\\ \sin\frac{\psi}{2}\end{pmatrix},\end{aligned} \tag{B.4}$$

where $\vec{S_0} := (S_{01}, S_{02}, S_{03})$. Here, Rotoselect simply solve the minimization problems of the quadratic form for different axis $\boldsymbol{m}\in\Lambda$ and select the optimal axis for minimizing the energy, where Λ is a predefined subset of rotational axis such as $\Lambda = \{(1,0,0),(0,1,0),(0,0,1)\}$.

Appendix C. Evaluation of the S elements

To determine the optimal single-qubit gate for the cost minimization, we evaluate the matrix S constructing the quadratic form equation (5). All the elements of S in equation (7) are calculated from ten expected values classified into three types as below.

Type-A

$$\mathrm{tr}\left[\rho'_{\mathrm{in}}H'\right]$$

Type-B for $k=1,2,3$,

$$\mathrm{tr}\left[\rho'_{\mathrm{in}}\left(\frac{I\pm i\sigma_k}{\sqrt{2}}\right)^{\dagger}H'\left(\frac{I\pm i\sigma_k}{\sqrt{2}}\right)\right]$$

Type-C for $(k,m)=(1,2),(1,3),(2,3)$

$$\mathrm{tr}\left[\rho'_{\mathrm{in}}\left(\frac{\sigma_k+\sigma_m}{\sqrt{2}}\right)^{\dagger}H'\left(\frac{\sigma_k+\sigma_m}{\sqrt{2}}\right)\right].$$

Here, Type-A is the $(0,0)$-the element of S and corresponds to the expected value of H on the PQC when the single-qubit gate of interest, as in figure 2, is replaced with identity. The other diagonal elements are produced by the type-A and type-B values with the following identity:

$$\mathrm{tr}\left(H'\sigma_k\rho'_{\mathrm{in}}\sigma_k\right)+\mathrm{tr}\left(H'\rho'_{\mathrm{in}}\right)=\mathrm{tr}\left[\rho'_{\mathrm{in}}\left(\frac{I-\mathrm{i}\sigma_k}{\sqrt{2}}\right)^{\dagger}H'\left(\frac{I-\mathrm{i}\sigma_k}{\sqrt{2}}\right)\right]+\mathrm{tr}\left[\rho'_{\mathrm{in}}\left(\frac{I+\mathrm{i}\sigma_k}{\sqrt{2}}\right)^{\dagger}H'\left(\frac{I+\mathrm{i}\sigma_k}{\sqrt{2}}\right)\right]. \tag{C.1}$$

Note that the Type-B corresponds to the expected value of H on the PQC when the single-qubit gate of interest, as in figure 2, is replaced with, respectively, $R_x(-\pi/2)$, $R_x(\pi/2)$, $R_y(-\pi/2)$, $R_y(\pi/2)$, $R_z(-\pi/2)$, and $R_z(\pi/2)$. In contrast, subtracting the type-B values with different signs yields the other elements in the first row directly. The remaining off-diagonal elements are produced by the type-C expected values and the already obtained diagonal elements with the following identity:

$$\begin{aligned}\frac{1}{2}\left[\mathrm{tr}\left(H'\sigma_i\rho'_{\mathrm{in}}\sigma_j\right)+\mathrm{tr}\left(H'\sigma_j\rho'_{\mathrm{in}}\sigma_i\right)\right]&=\mathrm{tr}\left[\rho'_{\mathrm{in}}\left(\frac{\sigma_i+\sigma_j}{\sqrt{2}}\right)^{\dagger}H'\left(\frac{\sigma_i+\sigma_j}{\sqrt{2}}\right)\right]\\&\quad-\frac{1}{2}\left[\mathrm{tr}\left(H'\sigma_i\rho'_{\mathrm{in}}\sigma_i\right)+\mathrm{tr}\left(H'\sigma_j\rho'_{\mathrm{in}}\sigma_j\right)\right].\end{aligned} \tag{C.2}$$

Note that the Type-C values correspond to the expected values of H on the PQC when the single-qubit gate to be optimized, as in figure 2, is replaced with, respectively, Fraxis gate $R_{\boldsymbol{n}}(\pi)$ for $\boldsymbol{n}\propto(1,1,0),(1,0,1),(0,1,1)$. All expected values of Type-A, B, and C can be evaluated with direct measurements without any control operation such as the Hadamard test. Since the degree of freedom for a 4×4 real-symmetric matrix is ten, the number of required direct measurements should be optimal.

Appendix D. Proof of main theorems

As derived in appendix B, each sequential quantum optimizer has a $p\times p$ real-symmetric matrix. Here, $p-1$ is the degree of freedom of the target single-qubit gate (-1 means the constraint of normalization). Note that FQS gate or general single-qubit gate $R(\boldsymbol{q})$ $(p=4)$, Fraxis gate $R_{\boldsymbol{n}}(\pi)$ $(p=3)$, and NFT gate with $\boldsymbol{m}$ axis $R_{\boldsymbol{m}}(\psi)$ $(p=2)$ are written as the following unified form

$$R^{(p)}:=\sum_{\mu=0}^{p-1}q_{\mu}^{(p)}\varsigma_{\mu}^{(p)}, \tag{D.1}$$

where $q_{\mu}^{(p)}$ is the μth element of unit vector in $\mathbb{R}^p$. Here, $\varsigma_{\mu}^{(p)}$ is an extension of the Pauli matrices such as $(\vec{\varsigma})^{(3)}=(-iX,-iY,-iZ)$ for Fraxis gate and $(\vec{\varsigma})^{(2)}=(I,-i\boldsymbol{m}\cdot\vec{\sigma})$ for NFT gate. Accordingly, each $p\times p$ real-symmetric matrix is also written as

$$S_{\mu\nu}^{(p)}:=\frac{1}{2}\mathrm{tr}\left[\rho'_{\mathrm{in}}\left(\left(\varsigma_{\mu}^{(p)}\right)^{\dagger}H'\varsigma_{\nu}^{(p)}+\left(\varsigma_{\nu}^{(p)}\right)^{\dagger}H'\varsigma_{\mu}^{(p)}\right)\right]. \tag{D.2}$$

Note that $S^{(3)}$ and $S^{(2)}$ correspond to the real-symmetric matrix in equations (B.3), (B.4), respectively. In the below, we omit the superscript p for simplicity. Moreover, it is convenient to write the elements of S as

$$\begin{aligned} S_{\mu\nu} &= \frac{1}{2} \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \mathrm{tr}\left[HU_2\varsigma_\alpha U_1\rho_{\mathrm{in}}U_1^\dagger\varsigma_\beta^\dagger U_2^\dagger\right], \quad \mu,\nu=0,1,\cdots,p-1, \\ &= \frac{1}{2} \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \mathrm{tr}\left[H^{(2)}\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\right], \quad H^{(2)} := U_2^\dagger H U_2, \quad \rho_{\mathrm{in}}^{(1)} := U_1\rho_{\mathrm{in}}U_1^\dagger. \end{aligned} \tag{D.3}$$

D.1. Proof of theorem 1

In the following, d denotes the dimension of n-qubit system i.e. $d = 2^n$. Here, we provide a proof of theorem 1. For convenience, we recall it.

Theorem 1. *Suppose that the quantum circuits U_1 and U_2 are randomly and independently generated. If either U_1 or U_2 forms a unitary t-design with $t \geqslant 2$, the second moment of the spectral radius, r, for the centered matrix $S_{\mathrm{c}} = S - \mathrm{tr}[S]I_{p\times p}/p$ ($I_{p\times p}$ denotes the $p\times p$ identity) is upper bounded as follows.*

$$\mathbb{E}_{U_1,U_2}\left[r(S_{\mathrm{c}})^2\right] \leqslant \begin{cases} \frac{p^2\mathrm{tr}[H^2]\Delta\rho_{\mathrm{in},2}}{2(d^2-1)} + \frac{\Delta\rho_{\mathrm{in},2}}{4p(d^2-1)} \sum_{\mu,\nu=0}^{p-1}\left(p(-1)^{1-\delta_{\mu\nu}}-2\right)\mathrm{tr}\left[\mathbb{E}_{U_2}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\nu\varsigma_\mu^\dagger + \mathrm{h.c.}\right]\right] \\ \frac{p^2\mathrm{tr}[\rho_{\mathrm{in}}^2]\Delta H_2}{2(d^2-1)} + \frac{\Delta H_2}{4p(d^2-1)} \sum_{\mu,\nu=0}^{p-1}\left(p(-1)^{1-\delta_{\mu\nu}}-2\right)\mathrm{tr}\left[\mathbb{E}_{U_1}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\mu^\dagger\varsigma_\nu\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu + \mathrm{h.c.}\right]\right] \end{cases}, \tag{D.4}$$

where the first case corresponds to U_1 being a t-design, and the second case corresponds to U_2 being a t-design. Here, $\mathbb{E}_{U_1,U_2}[\cdot]$ is defined as the expectation over the random quantum circuits U_1 and U_2, $\delta_{\mu\nu}$ denotes the Kronecker delta, h.c. means the Hermite conjugate of the preceding term, and $\Delta\rho_{\mathrm{in},2}, \Delta H_2$ are defined as

$$\Delta\rho_{\mathrm{in},2} := \mathrm{tr}\left[\rho_{\mathrm{in}}^2\right] - \frac{1}{d}, \quad \Delta H_2 := \mathrm{tr}\left[H^2\right] - \frac{\mathrm{tr}^2[H]}{d}. \tag{D.5}$$

Proof. Since the centered matrix S_{c} is a real-symmetric matrix regardless of quantum circuits U_1 and U_2, the following inequality holds, which comes from the maximum is at most the square root of the sum of the squares.

$$r\left(S - \frac{\mathrm{tr}[S]}{p}I_{p\times p}\right) \leqslant \left\|S - \frac{\mathrm{tr}[S]}{p}I_{p\times p}\right\|_{\mathrm{F}}, \tag{D.6}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. Thus, the second moment of the spectral radius is evaluated as

$$\begin{aligned} \mathbb{E}_{U_1,U_2}\left[r\left(S - \frac{\mathrm{tr}[S]}{p}I_{p\times p}\right)^2\right] &\leqslant \mathbb{E}_{U_1,U_2}\left[\left\|S - \frac{\mathrm{tr}[S]}{p}I_{p\times p}\right\|_{\mathrm{F}}^2\right] \\ &= \sum_{\mu\nu}\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right] - \frac{1}{p}\sum_{\mu\nu}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right]. \end{aligned} \tag{D.7}$$

First, we evaluate the r.h.s of equation (D.7) on the condition that the random quantum circuit U_1 forms a unitary t-design with $t \geqslant 2$. Using lemma 1, which is proved in the next subsection, the following identity holds,

$$\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}S_{\mu'\nu'}\right] = \frac{\mathrm{tr}^2[H]}{d^2-1}\Delta\rho_{\mathrm{in},1}\delta_{\mu\nu}\delta_{\mu'\nu'} + \frac{1}{4(d^2-1)}\Delta\rho_{\mathrm{in},2}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathbb{E}_{U_2}\left[\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]\right] \tag{D.8}$$

where we defined

$$\Delta\rho_{\mathrm{in},1} := 1 - \frac{\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]}{d}. \tag{D.9}$$

Hence, we can calculate the expectations in the final line of equation (D.7) as follows.

$$\begin{aligned}\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right] &= \frac{\mathrm{tr}^2[H]}{d^2-1}\Delta\rho_{\mathrm{in},1}\delta_{\mu\nu} \\ &\quad + \frac{1}{4(d^2-1)}\Delta\rho_{\mathrm{in},2}\sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)}\sum_{(\alpha',\beta')=(\mu,\nu),(\nu,\mu)}\mathbb{E}_{U_2}\left[\mathrm{tr}\left[H^{(2)}\varsigma_\alpha\varsigma_\beta^\dagger H^{(2)}\varsigma_{\alpha'}\varsigma_{\beta'}^\dagger\right]\right] \\ &= \frac{\mathrm{tr}^2[H]}{d^2-1}\Delta\rho_{\mathrm{in},1}\delta_{\mu\nu} + \frac{\mathrm{tr}\left[H^2\right]}{2(d^2-1)}\Delta\rho_{\mathrm{in},2} \\ &\quad + \frac{1}{4(d^2-1)}\Delta\rho_{\mathrm{in},2}\mathrm{tr}\left[\mathbb{E}_{U_2}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger + \mathrm{h.c.}\right]\right], \end{aligned} \tag{D.10}$$

$$\begin{aligned}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right] &= \frac{\mathrm{tr}^2[H]}{d^2-1}\Delta\rho_{\mathrm{in},1} \\ &\quad + \frac{1}{4(d^2-1)}\Delta\rho_{\mathrm{in},2}\sum_{(\alpha,\beta)=(\mu,\mu),(\mu,\mu)}\sum_{(\alpha',\beta')=(\nu,\nu),(\nu,\nu)}\mathbb{E}_{U_2}\left[\mathrm{tr}\left[H^{(2)}\varsigma_\alpha\varsigma_\beta^\dagger H^{(2)}\varsigma_{\alpha'}\varsigma_{\beta'}^\dagger\right]\right] \\ &= \frac{\mathrm{tr}^2[H]}{d^2-1}\Delta\rho_{\mathrm{in},1} + \frac{1}{d^2-1}\Delta\rho_{\mathrm{in},2}\mathbb{E}_{U_2}\left[\mathrm{tr}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\nu\varsigma_\mu^\dagger\right]\right] \\ &= \frac{\mathrm{tr}^2[H]}{d^2-1}\Delta\rho_{\mathrm{in},1} + \frac{1}{2(d^2-1)}\Delta\rho_{\mathrm{in},2}\mathrm{tr}\left[\mathbb{E}_{U_2}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\nu\varsigma_\mu^\dagger + \mathrm{h.c.}\right]\right]. \end{aligned} \tag{D.11}$$

Accordingly, we can evaluate the second moment of the spectral radius as

$$\begin{aligned}&\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right] - \frac{1}{p}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right] \\ &= \frac{\mathrm{tr}^2[H]}{p(d^2-1)}\Delta\rho_{\mathrm{in},1}(p\delta_{\mu\nu}-1) + \frac{\mathrm{tr}\left[H^2\right]}{2(d^2-1)}\Delta\rho_{\mathrm{in},2} \\ &\quad + \frac{\Delta\rho_{\mathrm{in},2}}{4p(d^2-1)}\left(p\mathrm{tr}\left[\mathbb{E}_{U_2}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger + \mathrm{h.c.}\right]\right] - 2\mathrm{tr}\left[\mathbb{E}_{U_2}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\nu\varsigma_\mu^\dagger + \mathrm{h.c.}\right]\right]\right) \\ &= \frac{\mathrm{tr}^2[H]}{p(d^2-1)}\Delta\rho_{\mathrm{in},1}(p\delta_{\mu\nu}-1) + \frac{\mathrm{tr}\left[H^2\right]}{2(d^2-1)}\Delta\rho_{\mathrm{in},2} \\ &\quad + \frac{\Delta\rho_{\mathrm{in},2}}{4p(d^2-1)}\left(p(-1)^{1-\delta_{\mu\nu}}-2\right)\mathrm{tr}\left[\mathbb{E}_{U_2}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\nu\varsigma_\mu^\dagger + \mathrm{h.c.}\right]\right], \end{aligned} \tag{D.12}$$

where we used $\varsigma_\mu\varsigma_\nu^\dagger = (-1)^{1-\delta_{\mu\nu}}\varsigma_\nu\varsigma_\mu^\dagger$ in the last equality, and

$$\begin{aligned}&\mathbb{E}_{U_1,U_2}\left[r\left(S - \frac{\mathrm{tr}[S]}{p}I_{p\times p}\right)^2\right] \\ &\leqslant \sum_{\mu\nu}\left(\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right] - \frac{1}{p}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right]\right) \\ &= \frac{p^2\mathrm{tr}\left[H^2\right]\Delta\rho_{\mathrm{in},2}}{2(d^2-1)} + \frac{\mathrm{tr}^2[H]\Delta\rho_{\mathrm{in},1}}{p(d^2-1)}\sum_{\mu\nu}(p\delta_{\mu\nu}-1) \\ &\quad + \frac{\Delta\rho_{\mathrm{in},2}}{4p(d^2-1)}\sum_{\mu\nu}\left(p(-1)^{1-\delta_{\mu\nu}}-2\right)\mathrm{tr}\left[\mathbb{E}_{U_2}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\nu\varsigma_\mu^\dagger + \mathrm{h.c.}\right]\right] \\ &= \frac{p^2\mathrm{tr}\left[H^2\right]\Delta\rho_{\mathrm{in},2}}{2(d^2-1)} + \frac{\Delta\rho_{\mathrm{in},2}}{4p(d^2-1)}\sum_{\mu\nu}\left(p(-1)^{1-\delta_{\mu\nu}}-2\right)\mathrm{tr}\left[\mathbb{E}_{U_2}\left[H^{(2)}\varsigma_\mu\varsigma_\nu^\dagger H^{(2)}\varsigma_\nu\varsigma_\mu^\dagger + \mathrm{h.c.}\right]\right]. \end{aligned} \tag{D.13}$$

Next, we evaluate the r.h.s of equation (D.7) on the condition that the random quantum circuit U_2 forms a unitary t-design with $t \geqslant 2$. Using lemma 2, which is also proved in the next subsection, we calculate the expectations in the final line of equation (D.7) as follows.

$$\begin{aligned}\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right] &= \frac{\mathrm{tr}^2\left[\rho_{\mathrm{in}}\right]}{d^2-1}\Delta H_1\delta_{\mu\nu} \\ &\quad + \frac{1}{4\left(d^2-1\right)}\Delta H_2 \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)} \sum_{(\alpha',\beta')=(\mu,\nu),(\nu,\mu)} \mathbb{E}_{U_1}\left[\mathrm{tr}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\varsigma_\alpha\right]\right] \\ &= \frac{1}{d^2-1}\Delta H_1\delta_{\mu\nu} + \frac{\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]}{2\left(d^2-1\right)}\Delta H_2 \\ &\quad + \frac{1}{4\left(d^2-1\right)}\Delta H_2\mathrm{tr}\left[\mathbb{E}_{U_1}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu + \mathrm{h.c.}\right]\right], \end{aligned} \tag{D.14}$$

$$\begin{aligned}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right] &= \frac{\mathrm{tr}^2\left[\rho_{\mathrm{in}}\right]}{d^2-1}\Delta H_1 \\ &\quad + \frac{1}{4\left(d^2-1\right)}\Delta H_2 \sum_{(\alpha,\beta)=(\mu,\mu),(\mu,\mu)} \sum_{(\alpha',\beta')=(\nu,\nu),(\nu,\nu)} \mathbb{E}_{U_1}\left[\mathrm{tr}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\varsigma_\alpha\right]\right] \\ &= \frac{1}{d^2-1}\Delta H_1 + \frac{1}{d^2-1}\Delta H_2\mathbb{E}_{U_1}\left[\mathrm{tr}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\mu^\dagger\varsigma_\nu\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu\right]\right] \\ &= \frac{1}{d^2-1}\Delta H_1 + \frac{1}{2\left(d^2-1\right)}\Delta H_2\mathrm{tr}\left[\mathbb{E}_{U_1}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\mu^\dagger\varsigma_\nu\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu + \mathrm{h.c.}\right]\right], \end{aligned} \tag{D.15}$$

where we defined

$$\Delta H_1 := \mathrm{tr}^2\left[H\right] - \frac{\mathrm{tr}\left[H^2\right]}{d}. \tag{D.16}$$

Accordingly, we can evaluate the second moment of the spectral radius as

$$\begin{aligned}&\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right] - \frac{1}{p}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right] \\ &= \frac{1}{p\left(d^2-1\right)}\Delta H_1\left(p\delta_{\mu\nu}-1\right) + \frac{\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]}{2\left(d^2-1\right)}\Delta H_2 \\ &\quad + \frac{\Delta H_2}{4p\left(d^2-1\right)}\left(p\,\mathrm{tr}\left[\mathbb{E}_{U_1}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu + \mathrm{h.c.}\right]\right] - 2\mathrm{tr}\left[\mathbb{E}_{U_1}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\mu^\dagger\varsigma_\nu\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu + \mathrm{h.c.}\right]\right]\right) \\ &= \frac{1}{p\left(d^2-1\right)}\Delta H_1\left(p\delta_{\mu\nu}-1\right) + \frac{\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]}{2\left(d^2-1\right)}\Delta H_2 \\ &\quad + \frac{\Delta H_2}{4p\left(d^2-1\right)}\left(p\left(-1\right)^{1-\delta_{\mu\nu}}-2\right)\mathrm{tr}\left[\mathbb{E}_{U_1}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\mu^\dagger\varsigma_\nu\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu + \mathrm{h.c.}\right]\right], \end{aligned} \tag{D.17}$$

where we used $\varsigma_\nu^\dagger\varsigma_\mu = (-1)^{1-\delta_{\mu\nu}}\varsigma_\mu^\dagger\varsigma_\nu$ in the last equality, and

$$\begin{aligned}&\mathbb{E}_{U_1,U_2}\left[r\left(S - \frac{\mathrm{tr}\left[S\right]}{p}I_{p\times p}\right)^2\right] \\ &\leqslant \sum_{\mu\nu}\left(\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right] - \frac{1}{p}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right]\right) \\ &= \frac{p^2\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]\Delta H_2}{2\left(d^2-1\right)} + \frac{\Delta H_2}{4p\left(d^2-1\right)}\sum_{\mu\nu}\left(p\left(-1\right)^{1-\delta_{\mu\nu}}-2\right)\mathrm{tr}\left[\mathbb{E}_{U_1}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\mu^\dagger\varsigma_\nu\rho_{\mathrm{in}}^{(1)}\varsigma_\nu^\dagger\varsigma_\mu + \mathrm{h.c.}\right]\right]. \end{aligned} \tag{D.18}$$

□

D.2. Proof of theorem 2

Here, we provide a proof of theorem 2. For convenience, we recall it.

Theorem 2. *Suppose that the whole quantum circuit U is an n-qubit alternating layered ansatz with m-qubit blocks as described in section 2.3.2. Here, we focus on a block W in lth layer and a parameterized single-qubit gate R in the block. We assume that the quantum circuits $W_{\mathrm{A}}, W_{\mathrm{B}} \subset W$, which are located after and before the target gate respectively, and the other blocks form a local 2-design independently. In addition, the Hamiltonian H is assumed to be m-local such as*

$$H = c_0 I^{\otimes n} + \sum_i c_i h_i,\ c_0, c_i \in \mathbb{R},\ \left(\exists i, c_i \neq 0\right), \tag{D.19}$$

where a tensor product of Pauli matrices h_i acts non-trivially on at most m-qubit. (This is the same assumption in [33].) Then, the second moment of the spectral radius is lower bounded as

$$\mathbb{E}_{U_1,U_2}\left[r\left(S-\frac{\mathrm{tr}[S]}{p}I_{p\times p}\right)^2\right]\geqslant\frac{(p+2)(p-1)2^{m(l+1)-1}}{p(2^{2m}-1)^2(2^m+1)^{L+l}}\sum_{i\in i_{\mathcal{L}}}\sum_{\substack{(k,k')\in k_{\mathcal{L}_{\mathrm{B}}}\\ k'\geqslant k}}c_i^2\epsilon(\rho_{k,k'})\epsilon(h_i)\,, \tag{D.20}$$

where L denotes the total number of layers. Here, $i_{\mathcal{L}}$ is the set of i indices whose associated operators h_i act on qubits in the forward light-cone $\mathcal{L}$ of W, and $k_{\mathcal{L}_{\mathrm{B}}}$ is the set of k indices whose associated subsystems S_k are in the backward light-cone $\mathcal{L}_{\mathrm{B}}$ of W. The quantum state $\rho_{k,k'}$ is the reduced density matrix of the input state ρ_{in} on $S_kS_{k+1}\cdots S_{k'}$, and the function $\epsilon(M)$ for a matrix M is defined as $\epsilon(M)=D_{\mathrm{HS}}(M,\mathrm{tr}(M)\mathbf{1}/d_M)$ where D_{HS} is the Hilbert–Schmidt distance and d_M is the dimension of the matrix M.

Proof. To establish the lower bound of the second moment, we begin with the following inequality, which comes from the maximum is at least the square root of the average of the sum of the squares.

$$r\left(S-\frac{\mathrm{tr}[S]}{p}I_{p\times p}\right)\geqslant\frac{1}{\sqrt{p}}\left\|S-\frac{\mathrm{tr}[S]}{p}I_{p\times p}\right\|_{\mathrm{F}}\,, \tag{D.21}$$

where $||\cdot||_{\mathrm{F}}$ denotes the Frobenius norm. Since the inequality holds for any quantum circuits U_1 and U_2 in figure 2,

$$\begin{aligned}\mathbb{E}_{U_1,U_2}\left[r\left(S-\frac{\mathrm{tr}[S]}{p}I_{p\times p}\right)^2\right]&\geqslant\frac{1}{p}\mathbb{E}_{U_1,U_2}\left[\left\|S-\frac{\mathrm{tr}[S]}{p}I_{p\times p}\right\|_{\mathrm{F}}^2\right]\\&=\frac{1}{p}\sum_{\mu\nu}\left(\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right]-\frac{1}{p}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right]\right).\end{aligned} \tag{D.22}$$

We first evaluate the expectation of $S_{\mu\nu}S_{\mu'\nu'}$ over the block of interest. From the setting, the block W containing a single-qubit gate R is decomposed as

$$W=W_{\mathrm{A}}(\mathbf{1}_{m-1}\otimes R)W_{\mathrm{B}}, \tag{D.23}$$

where $\mathbf{1}_{m-1}$ is the identity on $m-1$ qubits system, and $W_{\mathrm{A}},W_{\mathrm{B}}\subset W$ are the quantum circuits after and before R, respectively. As shown in figure 2, the quantum circuits U_1 and U_2 for an alternating layered ansatz can be written as

$$U_1=(\mathbf{1}_{\overline{w}}\otimes W_{\mathrm{B}})V_1,\quad U_2=V_2(\mathbf{1}_{\overline{w}}\otimes W_{\mathrm{A}})\,, \tag{D.24}$$

where $\mathbf{1}_{\overline{w}}$ denotes the identity over the qubits on which the block W acts trivially. Here, V_2 contains the gates in the forward light-cone $\mathcal{L}$ of W, i.e. all gates with at least one input qubit causally connected to the qubits of W as in figure 3, and V_1 contains other gates. Accordingly, we can write the elements of S matrix as

$$S_{\mu\nu}=\frac{1}{2}\sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)}\mathrm{tr}\left[HV_2(\mathbf{1}_{\overline{w}}\otimes W_{\mathrm{A}}(\mathbf{1}_{m-1}\otimes\varsigma_\alpha)W_{\mathrm{B}})V_1\rho_{\mathrm{in}}V_1^\dagger\left(\mathbf{1}_{\overline{w}}\otimes W_{\mathrm{B}}^\dagger\left(\mathbf{1}_{m-1}\otimes\varsigma_\beta^\dagger\right)W_{\mathrm{A}}^\dagger\right)V_2^\dagger\right], \tag{D.25}$$

and we obtain

$$\begin{aligned}&S_{\mu\nu}S_{\mu'\nu'}\\&=\frac{1}{4}\sum_{\substack{(\alpha,\beta)\\(\alpha',\beta')}}\mathrm{tr}\left[HV_2(\mathbf{1}_{\overline{w}}\otimes W_{\mathrm{A}}(\mathbf{1}_{m-1}\otimes\varsigma_\alpha)W_{\mathrm{B}})V_1\rho_{\mathrm{in}}V_1^\dagger\left(\mathbf{1}_{\overline{w}}\otimes W_{\mathrm{B}}^\dagger\left(\mathbf{1}_{m-1}\otimes\varsigma_\beta^\dagger\right)W_{\mathrm{A}}^\dagger\right)V_2^\dagger\right]\\&\quad\times\mathrm{tr}\left[HV_2(\mathbf{1}_{\overline{w}}\otimes W_{\mathrm{A}}(\mathbf{1}_{m-1}\otimes\varsigma_{\alpha'})W_{\mathrm{B}})V_1\rho_{\mathrm{in}}V_1^\dagger\left(\mathbf{1}_{\overline{w}}\otimes W_{\mathrm{B}}^\dagger\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta'}^\dagger\right)W_{\mathrm{A}}^\dagger\right)V_2^\dagger\right]\\&=\frac{1}{4}\sum_{\substack{(\alpha,\beta)\\(\alpha',\beta')}}\sum_{\substack{\boldsymbol{i},\boldsymbol{j}\\\boldsymbol{i'},\boldsymbol{j'}}}\mathrm{tr}\left[W_{\mathrm{A}}(\mathbf{1}_{m-1}\otimes\varsigma_\alpha)W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}W_{\mathrm{B}}^\dagger\left(\mathbf{1}_{m-1}\otimes\varsigma_\beta^\dagger\right)W_{\mathrm{A}}^\dagger H^{(2)}_{\boldsymbol{ji}}\right]\\&\quad\times\mathrm{tr}\left[W_{\mathrm{A}}(\mathbf{1}_{m-1}\otimes\varsigma_{\alpha'})W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{i'j'}}W_{\mathrm{B}}^\dagger\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta'}^\dagger\right)W_{\mathrm{A}}^\dagger H^{(2)}_{\boldsymbol{j'i'}}\right].\end{aligned} \tag{D.26}$$

Here, we defined

$$\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}} := \mathrm{tr}_{\overline{w}}\left[\left(|\boldsymbol{j}\rangle\langle\boldsymbol{i}|\otimes\mathbf{1}_{\omega}\right)V_1\rho_{\mathrm{in}}V_1^{\dagger}\right],\quad H^{(2)}_{\boldsymbol{ji}} := \mathrm{tr}_{\overline{w}}\left[\left(|\boldsymbol{i}\rangle\langle\boldsymbol{j}|\otimes\mathbf{1}_{w}\right)V_2^{\dagger}HV_2\right], \tag{D.27}$$

where $\mathrm{tr}_{\overline{\omega}}[\cdot]$ means the partial trace over the qubits which are not in W, and $\{|\boldsymbol{i}\rangle\}$ denotes the computational basis on $(n-m)$-qubit system.

Since W_{A} forms a local 2-design, we first compute the expectation of equation (D.26) over W_{A} as

$$\begin{aligned}
&\mathbb{E}_{W_{\mathrm{A}}}\left[S_{\mu\nu}S_{\mu'\nu'}\right]\\
&= \frac{1}{4\left(2^{2m}-1\right)}\sum_{\substack{(\alpha,\beta)\\(\alpha',\beta')}}\sum_{\boldsymbol{i},\boldsymbol{j}\\\boldsymbol{i}',\boldsymbol{j}'}\left(\mathrm{tr}\left[H^{(2)}_{\boldsymbol{ji}}\right]\mathrm{tr}\left[H^{(2)}_{\boldsymbol{j}'\boldsymbol{i}'}\right]-\frac{\mathrm{tr}\left[H^{(2)}_{\boldsymbol{ji}}H^{(2)}_{\boldsymbol{j}'\boldsymbol{i}'}\right]}{2^m}\right)\\
&\quad\times\mathrm{tr}\left[\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta}^{\dagger}\varsigma_{\alpha}\right)W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}W_{\mathrm{B}}^{\dagger}\right]\mathrm{tr}\left[\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta'}^{\dagger}\varsigma_{\alpha'}\right)W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}W_{\mathrm{B}}^{\dagger}\right]\\
&\quad+\frac{1}{4\left(2^{2m}-1\right)}\sum_{\substack{(\alpha,\beta)\\(\alpha',\beta')}}\sum_{\boldsymbol{i},\boldsymbol{j}\\\boldsymbol{i}',\boldsymbol{j}'}\left(\mathrm{tr}\left[H^{(2)}_{\boldsymbol{ji}}H^{(2)}_{\boldsymbol{j}'\boldsymbol{i}'}\right]-\frac{\mathrm{tr}\left[H^{(2)}_{\boldsymbol{ji}}\right]\mathrm{tr}\left[H^{(2)}_{\boldsymbol{j}'\boldsymbol{i}'}\right]}{2^m}\right)\\
&\quad\times\mathrm{tr}\left[\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta'}^{\dagger}\varsigma_{\alpha}\right)W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}W_{\mathrm{B}}^{\dagger}\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta}^{\dagger}\varsigma_{\alpha'}\right)W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}W_{\mathrm{B}}^{\dagger}\right].
\end{aligned}\tag{D.28}$$

Here, we employed the following integration formula from the Weingarten calculus [33]

$$\begin{aligned}
\mathbb{E}_{W}\,\mathrm{tr}\left[WAW^{\dagger}BWCW^{\dagger}D\right] &= \frac{1}{2^{2m}-1}\left(\mathrm{tr}\left[A\right]\mathrm{tr}\left[C\right]\mathrm{tr}\left[BD\right]+\mathrm{tr}\left[AC\right]\mathrm{tr}\left[B\right]\mathrm{tr}\left[D\right]\right)\\
&\quad-\frac{1}{2^m\left(2^{2m}-1\right)}\left(\mathrm{tr}\left[AC\right]\mathrm{tr}\left[BD\right]+\mathrm{tr}\left[A\right]\mathrm{tr}\left[B\right]\mathrm{tr}\left[C\right]\mathrm{tr}\left[D\right]\right),
\end{aligned}\tag{D.29}$$

where W is Haar-distributed on the unitary group of degree 2^m, and A, B, C and D are arbitrary linear operators on an m-qubit system. Noting that

$$\sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)}\mathrm{tr}\left[\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta}^{\dagger}\varsigma_{\alpha}\right)W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}W_{\mathrm{B}}^{\dagger}\right]=2\delta_{\mu\nu}\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\right]\tag{D.30}$$

holds, then the first term of equation (D.28) is independent of W_{B}. Since W_{B} also forms a local 2-design, a part of the second term of equation (D.28) is evaluated as

$$\begin{aligned}
&\sum_{\substack{(\alpha,\beta)\\(\alpha',\beta')}}\mathbb{E}_{W_{\mathrm{B}}}\,\mathrm{tr}\left[\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta'}^{\dagger}\varsigma_{\alpha}\right)W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}W_{\mathrm{B}}^{\dagger}\left(\mathbf{1}_{m-1}\otimes\varsigma_{\beta}^{\dagger}\varsigma_{\alpha'}\right)W_{\mathrm{B}}\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}W_{\mathrm{B}}^{\dagger}\right]\\
&=\frac{1}{2^{2m}-1}\sum_{\substack{(\alpha,\beta)\\(\alpha',\beta')}}\mathrm{tr}\left[\mathbf{1}_{m-1}\otimes\varsigma_{\beta}^{\dagger}\varsigma_{\alpha'}\varsigma_{\beta'}^{\dagger}\varsigma_{\alpha}\right]\left(\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}\right]-\frac{\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}\right]}{2^m}\right)\\
&\quad+\frac{1}{2^{2m}-1}\sum_{\substack{(\alpha,\beta)\\(\alpha',\beta')}}\mathrm{tr}\left[\mathbf{1}_{m-1}\otimes\varsigma_{\beta}^{\dagger}\varsigma_{\alpha'}\right]\mathrm{tr}\left[\mathbf{1}_{m-1}\otimes\varsigma_{\beta'}^{\dagger}\varsigma_{\alpha}\right]\left(\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}\right]-\frac{\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}\right]}{2^m}\right)\\
&=\frac{4\cdot 2^m}{2^{2m}-1}\delta_{\mu\nu}\delta_{\mu'\nu'}\left(\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}\right]-\frac{\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}\right]}{2^m}\right)\\
&\quad+\frac{8\cdot 2^{2(m-1)}}{2^{2m}-1}\left(\delta_{\mu'\mu}\delta_{\nu'\nu}+\delta_{\mu\nu'}\delta_{\mu'\nu}\right)\left(\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}\right]-\frac{\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{ij}}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},\boldsymbol{i}'\boldsymbol{j}'}\right]}{2^m}\right),
\end{aligned}\tag{D.31}$$

where we used the formula equation (D.29) in the first equality. In the second equality, we used the following relation, which can be derived from direct calculation, as

$$\sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)}\sum_{(\alpha',\beta')=(\mu',\nu'),(\nu',\mu')}\mathrm{tr}\left[\varsigma_{\beta}^{\dagger}\varsigma_{\alpha'}\right]\mathrm{tr}\left[\varsigma_{\beta'}^{\dagger}\varsigma_{\alpha}\right]=8\left(\delta_{\mu'\mu}\delta_{\nu'\nu}+\delta_{\mu\nu'}\delta_{\mu'\nu}\right).\tag{D.32}$$

Taking the expectation over W_{B} and substituting equations (D.30), (D.31) back into equation (D.28), we obtain

$$
\begin{aligned}
&\mathbb{E}_{W_{\mathrm{A}},W_{\mathrm{B}}}\left[S_{\mu\nu}S_{\mu'\nu'}\right] \\
&= \frac{\delta_{\mu\nu}\delta_{\mu'\nu'}}{2^{2m}-1}\sum_{\substack{i,j\\ i',j'}}\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},i'j'}\right]\left(\mathrm{tr}\left[H^{(2)}_{ji}\right]\mathrm{tr}\left[H^{(2)}_{j'i'}\right]-\frac{\mathrm{tr}\left[H^{(2)}_{ji}H^{(2)}_{j'i'}\right]}{2^m}\right) \\
&\quad+\frac{2^m\delta_{\mu\nu}\delta_{\mu'\nu'}}{\left(2^{2m}-1\right)^2}\sum_{\substack{i,j\\ i',j'}}\left(\mathrm{tr}\left[H^{(2)}_{ji}H^{(2)}_{j'i'}\right]-\frac{\mathrm{tr}\left[H^{(2)}_{ji}\right]\mathrm{tr}\left[H^{(2)}_{j'i'}\right]}{2^m}\right)\left(\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},i'j'}\right]-\frac{\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\rho^{(1)}_{\mathrm{in},i'j'}\right]}{2^m}\right) \\
&\quad+\frac{2\cdot 2^{2m}}{4\left(2^{2m}-1\right)^2}\left(\delta_{\mu'\mu}\delta_{\nu'\nu}+\delta_{\mu\nu'}\delta_{\mu'\nu}\right) \\
&\quad\times\sum_{\substack{i,j\\ i',j'}}\left(\mathrm{tr}\left[H^{(2)}_{ji}H^{(2)}_{j'i'}\right]-\frac{\mathrm{tr}\left[H^{(2)}_{ji}\right]\mathrm{tr}\left[H^{(2)}_{j'i'}\right]}{2^m}\right)\left(\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\rho^{(1)}_{\mathrm{in},i'j'}\right]-\frac{\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},i'j'}\right]}{2^m}\right) \\
&= \delta_{\mu\nu}\delta_{\mu'\nu'}T^{(1,2)}+\frac{2\cdot 2^{2m}}{4(2^{2m}-1)^2}\left(\delta_{\mu'\mu}\delta_{\nu'\nu}+\delta_{\mu\nu'}\delta_{\mu'\nu}\right)\sum_{\substack{i,j\\ i',j'}}\Delta H^{i'j'}_{ij}\,\Delta\rho^{i'j'}_{ij}\,. \qquad \text{(D.33)}
\end{aligned}
$$

Here, we defined

$$
\Delta H^{i'j'}_{ij} := \mathrm{tr}\left[H^{(2)}_{ji}H^{(2)}_{j'i'}\right]-\frac{\mathrm{tr}\left[H^{(2)}_{ji}\right]\mathrm{tr}\left[H^{(2)}_{j'i'}\right]}{2^m}, \qquad \text{(D.34)}
$$

$$
\Delta\rho^{i'j'}_{ij} := \mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\rho^{(1)}_{\mathrm{in},i'j'}\right]-\frac{\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},i'j'}\right]}{2^m}, \qquad \text{(D.35)}
$$

$$
\begin{aligned}
T^{(1,2)} :=& \frac{1}{2^{2m}-1}\sum_{\substack{i,j\\ i',j'}}\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},i'j'}\right]\left(\mathrm{tr}\left[H^{(2)}_{ji}\right]\mathrm{tr}\left[H^{(2)}_{j'i'}\right]-\frac{\mathrm{tr}\left[H^{(2)}_{ji}H^{(2)}_{j'i'}\right]}{2^m}\right) \\
&+\frac{2^m}{\left(2^{2m}-1\right)^2}\sum_{\substack{i,j\\ i',j'}}\Delta H^{i'j'}_{ij}\left(\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\right]\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},i'j'}\right]-\frac{\mathrm{tr}\left[\rho^{(1)}_{\mathrm{in},ij}\rho^{(1)}_{\mathrm{in},i'j'}\right]}{2^m}\right). \qquad \text{(D.36)}
\end{aligned}
$$

Before proceeding to evaluate the expectation over V_1 and V_2, we calculate the summation in equation (D.22) as follows.

$$
\sum_{\mu\nu}\mathbb{E}_{W_{\mathrm{A}},W_{\mathrm{B}}}\left[S^2_{\mu\nu}\right] = pT^{(1,2)}+\frac{2\cdot 2^{2m}}{4\left(2^{2m}-1\right)^2}\left(p^2+p\right)\sum_{\substack{i,j\\ i',j'}}\Delta H^{i'j'}_{ij}\,\Delta\rho^{i'j'}_{ij}\,, \qquad \text{(D.37)}
$$

$$
\sum_{\mu\nu}\mathbb{E}_{W_{\mathrm{A}},W_{\mathrm{B}}}\left[S_{\mu\mu}S_{\nu\nu}\right] = p^2T^{(1,2)}+\frac{2^{2m}}{\left(2^{2m}-1\right)^2}p\sum_{\substack{i,j\\ i',j'}}\Delta H^{i'j'}_{ij}\,\Delta\rho^{i'j'}_{ij}\,. \qquad \text{(D.38)}
$$

Accordingly,

$$
\begin{aligned}
\sum_{\mu\nu}\left(\mathbb{E}_{U_1,U_2}\left[S^2_{\mu\nu}\right]-\frac{1}{p}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right]\right) &= \mathbb{E}_{V_1,V_2}\left[\sum_{\mu\nu}\mathbb{E}_{W_{\mathrm{A}},W_{\mathrm{B}}}\left[S^2_{\mu\nu}\right]-\frac{1}{p}\sum_{\mu\nu}\mathbb{E}_{W_{\mathrm{A}},W_{\mathrm{B}}}\left[S_{\mu\mu}S_{\nu\nu}\right]\right] \\
&= \frac{2\cdot 2^{2m}(p+2)(p-1)}{4\left(2^{2m}-1\right)^2}\sum_{\substack{i,j\\ i',j'}}\mathbb{E}_{V_1}\left[\Delta\rho^{i'j'}_{ij}\right]\mathbb{E}_{V_2}\left[\Delta H^{i'j'}_{ij}\right]. \qquad \text{(D.39)}
\end{aligned}
$$

Finally, we evaluate the expectations over V_1, V_2 in equation (D.39), which can be calculated basically with a series of integration for the m-qubit blocks in V_1, V_2. In the same assumption of ours, the previous study [33] has showed the following inequality holds:

$$\sum_{\substack{i,j\\ i',j'}} \mathbb{E}_{V_1}\left[\Delta\rho_{ij}^{i'j'}\right]\mathbb{E}_{V_2}\left[\Delta H_{ij}^{i'j'}\right] \geqslant \frac{2^{m(l-1)}}{\left(2^m+1\right)^{L+l}} \sum_{i\in i_{\mathcal{L}}} \sum_{\substack{(k,k')\in k_{\mathcal{L}_{\mathrm{B}}}\\ k'\geqslant k}} c_i^2 \epsilon\left(\rho_{k,k'}\right)\epsilon\left(h_i\right), \tag{D.40}$$

where we recall that L is the total number of layers, and the block W of interest is in the lth layer. Combining this inequality with equation (D.39), we establish the lower bound for the second moment of spectral radius as

$$\begin{aligned}\mathbb{E}_{U_1,U_2}\left[r\left(S-\frac{\mathrm{tr}\left[S\right]}{p}I_{p\times p}\right)^2\right] &\geqslant \frac{1}{p}\sum_{\mu\nu}\left(\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}^2\right]-\frac{1}{p}\mathbb{E}_{U_1,U_2}\left[S_{\mu\mu}S_{\nu\nu}\right]\right)\\ &= \frac{1}{p}\frac{2\cdot 2^{2m}\left(p+2\right)\left(p-1\right)}{4\left(2^{2m}-1\right)^2}\sum_{\substack{i,j\\ i',j'}}\mathbb{E}_{V_1}\left[\Delta\rho_{ij}^{i'j'}\right]\mathbb{E}_{V_2}\left[\Delta H_{ij}^{i'j'}\right]\\ &\geqslant \frac{\left(p+2\right)\left(p-1\right)2^{m(l+1)-1}}{p\left(2^{2m}-1\right)^2\left(2^m+1\right)^{L+l}}\sum_{i\in i_{\mathcal{L}}}\sum_{\substack{(k,k')\in k_{\mathcal{L}_{\mathrm{B}}}\\ k'\geqslant k}} c_i^2\epsilon\left(\rho_{k,k'}\right)\epsilon\left(h_i\right).\end{aligned} \tag{D.41}$$

□

D.3. Proof of some useful lemmas

In this subsection, we provide some lemmas for the proof of theorem 1.

Lemma 1. *Suppose that the random quantum circuits U_1 and U_2 are independent. If U_1 forms a unitary t-design with $t\geqslant 2$, the expectation of the product of two matrix elements for S can be evaluated as*

$$\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}S_{\mu'\nu'}\right] = \frac{\mathrm{tr}^2\left[H\right]}{d^2-1}\Delta\rho_{\mathrm{in},1}\delta_{\mu\nu}\delta_{\mu'\nu'} + \frac{1}{4\left(d^2-1\right)}\Delta\rho_{\mathrm{in},2}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathbb{E}_{U_2}\left[\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]\right]. \tag{D.42}$$

Here, the summation runs over the set of μ,ν as follows

$$\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')} = \sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)}\sum_{(\alpha',\beta')=(\mu',\nu'),(\nu',\mu')}. \tag{D.43}$$

Proof. Substituting equation (D.3) into the l.h.s of equation (D.42) and omitting the expectation over U_2, we obtain

$$\begin{aligned}&\mathbb{E}_{U_1}\left[S_{\mu\nu}S_{\mu'\nu'}\right]\\ &= \frac{1}{4}\sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)}\sum_{(\alpha',\beta')=(\mu',\nu'),(\nu',\mu')}\mathbb{E}_{U_1}\,\mathrm{tr}\left[H^{(2)}\varsigma_\alpha U_1\rho_{\mathrm{in}}U_1^\dagger\varsigma_\beta^\dagger\right]\mathrm{tr}\left[H^{(2)}\varsigma_{\alpha'}U_1\rho_{\mathrm{in}}U_1^\dagger\varsigma_{\beta'}^\dagger\right]\\ &= \frac{1}{4}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\sum_{i,j}\mathbb{E}_{U_1}\,\mathrm{tr}\left[U_1\rho_{\mathrm{in}}U_1^\dagger\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha|\boldsymbol{i}\rangle\langle\boldsymbol{j}|U_1\rho_{\mathrm{in}}U_1^\dagger\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}|\boldsymbol{j}\rangle\langle\boldsymbol{i}|\right],\end{aligned} \tag{D.44}$$

where $\{|\boldsymbol{i}\rangle\}$ denotes the computational basis on n-qubit system. If U_1 forms a unitary t-design with $t\geqslant 2$, we can evaluate

$$\begin{aligned}&\sum_{i,j}\mathbb{E}_{U_1}\,\mathrm{tr}\left[U_1\rho_{\mathrm{in}}U_1^\dagger\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha|\boldsymbol{i}\rangle\langle\boldsymbol{j}|U_1\rho_{\mathrm{in}}U_1^\dagger\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}|\boldsymbol{j}\rangle\langle\boldsymbol{i}|\right]\\ &= \frac{1}{d^2-1}\left(\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\right]\mathrm{tr}\left[\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]+\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]\right)\\ &\quad-\frac{1}{d\left(d^2-1\right)}\left(\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\right]\mathrm{tr}\left[\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]+\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]\right),\end{aligned} \tag{D.45}$$

where we used the formula in equation (D.29). Substituting equation (D.45) back into equation (D.44), we obtain

$$\begin{aligned}
&\mathbb{E}_{U_1}\left[S_{\mu\nu}S_{\mu'\nu'}\right] \\
&= \frac{1}{4}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\frac{1}{d^2-1}\left(\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\right]\mathrm{tr}\left[\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]\left(1-\frac{\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]}{d}\right)\right) \\
&\quad + \frac{1}{4}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\frac{1}{d^2-1}\left(\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]\left(\mathrm{tr}\left[\rho_{\mathrm{in}}^2\right]-\frac{1}{d}\right)\right) \\
&= \frac{1}{4}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\frac{1}{d^2-1}\left(\Delta\rho_{\mathrm{in},1}\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\right]\mathrm{tr}\left[\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]+\Delta\rho_{\mathrm{in},2}\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]\right).
\end{aligned} \tag{D.46}$$

In addition, the first term in equation (D.46) can be further simplified as

$$\begin{aligned}
\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\right]\mathrm{tr}\left[\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right] &= \left(\sum_{(\alpha,\beta)}\mathrm{tr}\left[H^{(2)}\varsigma_\alpha\varsigma_\beta^\dagger\right]\right)\left(\sum_{(\alpha',\beta')}\mathrm{tr}\left[H^{(2)}\varsigma_{\alpha'}\varsigma_{\beta'}^\dagger\right]\right) \\
&= 4\delta_{\mu\nu}\mathrm{tr}\left[H^{(2)}\right]\delta_{\mu'\nu'}\mathrm{tr}\left[H^{(2)}\right] \\
&= 4\delta_{\mu\nu}\delta_{\mu'\nu'}\mathrm{tr}^2\left[H\right].
\end{aligned} \tag{D.47}$$

Finally, we obtain

$$\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}S_{\mu'\nu'}\right] = \frac{\mathrm{tr}^2\left[H\right]}{d^2-1}\Delta\rho_{\mathrm{in},1}\delta_{\mu\nu}\delta_{\mu'\nu'}+\frac{1}{4\left(d^2-1\right)}\Delta\rho_{\mathrm{in},2}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathbb{E}_{U_2}\left[\mathrm{tr}\left[\varsigma_\beta^\dagger H^{(2)}\varsigma_\alpha\varsigma_{\beta'}^\dagger H^{(2)}\varsigma_{\alpha'}\right]\right]. \tag{D.48}$$

□

Lemma 2. *Suppose that the random quantum circuits U_1 and U_2 are independent. If U_2 forms a unitary t-design with $t \geqslant 2$, the expectation of the product of two matrix elements for S can be evaluated as*

$$\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}S_{\mu'\nu'}\right] = \frac{\mathrm{tr}^2\left[\rho_{\mathrm{in}}\right]}{d^2-1}\Delta H_1\delta_{\mu\nu}\delta_{\mu'\nu'}+\frac{1}{4\left(d^2-1\right)}\Delta H_2\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathbb{E}_{U_1}\left[\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\right]. \tag{D.49}$$

Here, the summation runs in the same way as lemma 1.

Proof. This proof follows the same flow in lemma 1.

$$\begin{aligned}
&\mathbb{E}_{U_2}\left[S_{\mu\nu}S_{\mu'\nu'}\right] \\
&= \frac{1}{4}\sum_{(\alpha,\beta)=(\mu,\nu),(\nu,\mu)}\sum_{(\alpha',\beta')=(\mu',\nu'),(\nu',\mu')}\mathbb{E}_{U_2}\mathrm{tr}\left[U_2^\dagger H U_2\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\right]\mathrm{tr}\left[U_2^\dagger H U_2\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right] \\
&= \frac{1}{4}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\sum_{i,j}\mathbb{E}_{U_2}\mathrm{tr}\left[U_2\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger U_2^\dagger|i\rangle\langle j|HU_2\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger U_2^\dagger|j\rangle\langle i|H\right].
\end{aligned} \tag{D.50}$$

If U_2 forms a unitary t-design with $t \geqslant 2$, we obtain

$$\begin{aligned}
&\sum_{i,j}\mathbb{E}_{U_2}\mathrm{tr}\left[U_2\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger U_2^\dagger|i\rangle\langle j|HU_2\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger U_2^\dagger|j\rangle\langle i|H\right] \\
&= \frac{1}{d^2-1}\left(\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\right]\mathrm{tr}\left[\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\mathrm{tr}^2\left[H\right]+\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\mathrm{tr}\left[H^2\right]\right) \\
&\quad - \frac{1}{d\left(d^2-1\right)}\left(\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\mathrm{tr}^2\left[H\right]+\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\right]\mathrm{tr}\left[\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\mathrm{tr}\left[H^2\right]\right),
\end{aligned} \tag{D.51}$$

where we used the formula in equation (D.29). Thus, this leads to

$$\begin{aligned}
&\mathbb{E}_{U_2}\left[S_{\mu\nu}S_{\mu'\nu'}\right] \\
&= \frac{1}{4}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\frac{1}{d^2-1}\left(\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\right]\mathrm{tr}\left[\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\left(\mathrm{tr}^2\left[H\right]-\frac{\mathrm{tr}\left[H^2\right]}{d}\right)\right) \\
&\quad+\frac{1}{4}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\frac{1}{d^2-1}\left(\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\left(\mathrm{tr}\left[H^2\right]-\frac{\mathrm{tr}^2\left[H\right]}{d}\right)\right) \\
&= \frac{1}{4}\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\frac{1}{d^2-1}\left(\Delta H_1\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\right]\mathrm{tr}\left[\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]+\Delta H_2\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\right). \qquad \text{(D.52)}
\end{aligned}$$

In addition, the first term in equation (D.52) can be further simplified as

$$\begin{aligned}
\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\right]\mathrm{tr}\left[\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right] &= \left(\sum_{(\alpha,\beta)}\mathrm{tr}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_\alpha\right]\right)\left(\sum_{(\alpha',\beta')}\mathrm{tr}\left[\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\varsigma_{\alpha'}\right]\right) \\
&= 4\delta_{\mu\nu}\mathrm{tr}\left[\rho_{\mathrm{in}}^{(1)}\right]\delta_{\mu'\nu'}\mathrm{tr}\left[\rho_{\mathrm{in}}^{(1)}\right] \\
&= 4\delta_{\mu\nu}\delta_{\mu'\nu'}. \qquad \text{(D.53)}
\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}
&\mathbb{E}_{U_1,U_2}\left[S_{\mu\nu}S_{\mu'\nu'}\right] \\
&= \frac{1}{4\left(d^2-1\right)}\mathbb{E}_{U_1}\left[\Delta H_1\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\right]\mathrm{tr}\left[\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\right] \\
&\quad+\frac{1}{4\left(d^2-1\right)}\mathbb{E}_{U_1}\left[\Delta H_2\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\right] \\
&= \frac{1}{d^2-1}\Delta H_1\delta_{\mu\nu}\delta_{\mu'\nu'}+\frac{1}{4\left(d^2-1\right)}\Delta H_2\sum_{(\alpha,\beta)}\sum_{(\alpha',\beta')}\mathbb{E}_{U_1}\left[\mathrm{tr}\left[\varsigma_\alpha\rho_{\mathrm{in}}^{(1)}\varsigma_\beta^\dagger\varsigma_{\alpha'}\rho_{\mathrm{in}}^{(1)}\varsigma_{\beta'}^\dagger\right]\right]. \qquad \text{(D.54)}
\end{aligned}$$

□

Appendix E. Setup for numerical simulations

E.1. 1-dimensional mixed field Ising model

We carried out VQE optimizations for the 1-dimensional n-qubit mixed field Ising model whose Hamiltonian is

$$H = J\sum_{i=1}^{n}Z^{(i)}Z^{(i+1)}+h\sum_{i=1}^{n}\left(\mathrm{Y}^{(i)}+Z^{(i)}\right), \qquad \text{(E.1)}$$

where we employed $n=5,20$. The superscript i denotes the index of each site, and $J=1, h=1/\sqrt{2}$. We employed the periodic boundary condition, that is, $Z^{(n+1)}=Z^{(1)}$. The employed ansaetze are shown in figure 9.

In the FQS optimization for 5-qubit system, we prepared 20 initial parameter sets in the state-random manner, where the respective single-qubit gates were initialized based on the Haar random distribution. As for 20-qubit system, a rotational axis $\boldsymbol{n}$ of each FQS gate $R_{\boldsymbol{n}}(\psi)$ was initialized with a uniformly random distribution on the Bloch sphere while a rotational angle ψ was randomly chosen from a uniform distribution in the interval $[0, 0.3\pi)$, respectively. Similarly, in the Fraxis optimizations, the rotational axis $\boldsymbol{n}$ in $R_{\boldsymbol{n}}(\psi)$ of each single-qubit gate was sampled from the uniform probability distribution on the Bloch sphere, while the rotational angles were fixed to π. As for Rotoselect, the initial rotational axes were randomly selected from X, Y and Z axes, while the rotational angle were randomly initialized in both Rotosolve (NFT) and Rotoselect.

In the case of Rotosolve(*RzRyRz*) (or Rotosolve(*RyRz*)), we replaced R in figure 9 with three fixed-axis rotation gates $R_z(\phi)R_y(\vartheta)R_z(\lambda)$ (or *RyRz*) gates and sequentially optimized the gates by Rotosolve. In a single sweep, all gates were updated once in ascending order of the gate set index as labeled in figures 9(a) and (b).

For comparison, we also conducted the conventional gradient-based VQE, where we replaced FQS gates by either *RzRyRz* or *RyRz* gates. Note that we employed the learning rate of 0.1 as a hyper-parameter for Adam, which appears to be rather larger than in its conventional usage in VQA, but provided modest results compared to 0.01 and 0.001 in statevector simulations as shown figure 13 in appendix H.

E.2. Fidelity maximization

To benchmark the FQS application to a global cost function, we conducted fidelity maximization simulation. In the optimization, infidelity was employed as the cost function as

$$1 - \mathrm{tr}\left[\rho'_{\mathrm{in}} R(\boldsymbol{q}_d)^{\dagger} \rho_{\mathrm{ref}} R(\boldsymbol{q}_d)\right]. \tag{E.2}$$

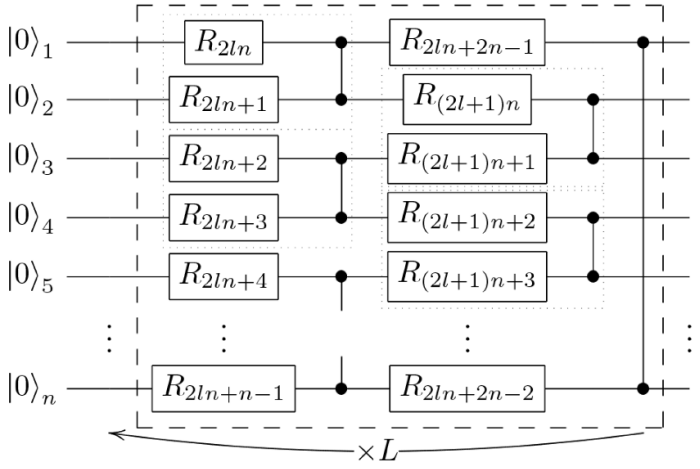
where ρ_{ref} was a reference state that were independently prepared with Haar random generator in Qiskit [67] in each optimization.

E.3. Spectral radius and noise-free barren plateau

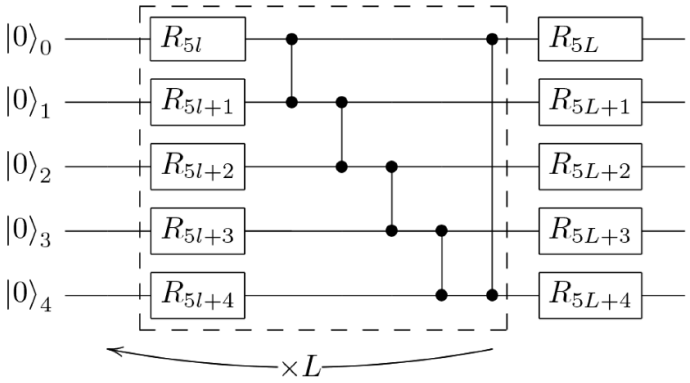
We evaluated the second moment of the spectral radius of the centered FQS matrix equation (10) in comparison of a global cost and a local cost functions. The FQS matrix was evaluated for the single-qubit gate acting on the first qubit in the first layer on the alternating layered ansatz. The second moment is calculated with 1000 samples based on the randomly initialized circuits with respect to the parameterized single-qubit gates in the state-random manner. As the global cost, we employed infidelity with a random reference state in equation (E.2). On the other hand, we used the 1-local Hamiltonian defined as $H = Z \otimes I^{\otimes n-1}$ as the local cost.

To examine the practical performance of VQE in a relation with barren plateau, we carried out the optimizations varying the number of qubits on the 5-layer alternating layered ansatz in figure 9(a). As the global cost we employed the infidelity in equation (E.2) with randomly-generated states as described in the previous section, while the local cost is the Hamiltonian of the mixed-field Ising model in equation (E.1). For both the cost functions, we executed independent 20 VQE runs starting from different initial conditions and evaluated the resulting deviation from the exact minimum value after 100 sweeps.

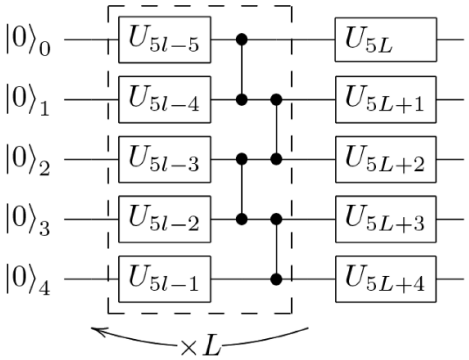
Appendix F. Ansätze



(a) Alternating layered ansatz



(b) Grouped-layer ansatz with cyclic entangler



(c) Grouped-layer PQC with ladder entangler

Figure 9. PQCs employed for numerical experiments. Each layer consists of gates in the dashed line, and the total number of layers is written as L. In sequential quantum optimization, parameterized single-qubit gates R are updated in ascending order of the subscript.

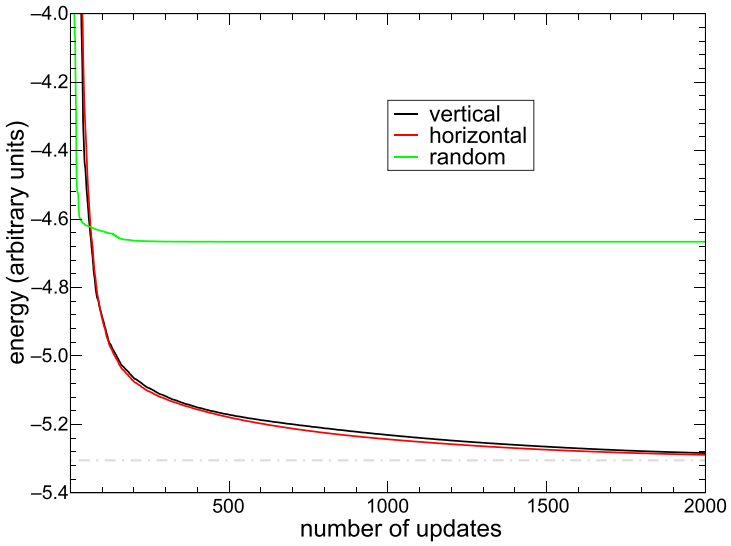


Figure 10. Optimization trajectories of the VQE resulting energy for 1-dimensional mixed field Ising model on various update orders. The trajectories were obtained on the 5-layer alternating layered ansatz using the statevector simulator. Each trajectories are average over 20 independent runs whose initial states were randomly generated. The horizontal dash-dotted line in gray color represents the exact ground energy.

Appendix G. Update order

In the case of sequential optimization, the order of gate (parameter) updates is arbitrary. To examine how the results depend on the update order, we tested three update orders (vertical, horizontal, random) using a 5-qubit Ising model and alternating layered ansatz as shown in figure (a). In the vertical update method, single-qubit gates in the same layer are updated first, and then those in the next layer are updated. In other words, single qubit gates are updated in ascending order of gate index, as shown in figure (a). In the horizontal update method, single-qubit gates acting on an identical qubit were updated first in ascending order of gate index, and then those on the next qubit were updated. In vertical and horizontal updates, an updating cycle of all single qubit gates is referred to as a sweep. In random updates, single-qubit gates to be updated are randomly selected. Because the consecutive update of an identical gate by FQS does not improve the cost, if one gate is consecutively selected, the target gate is selected again using a random number. Figure 10 showed the resulting trajectories of energy in comparison of updating order. The results showed that the vertical and horizontal updates have comparable performance, while the random update converged faster, but led to the local minima. Henceforth, the results in the rest of this manuscript are calculated based on vertical update. Given the fact that the consecutive FQS applications to a gate does not improve the cost, it is presumed that to efficiently improve cost by a single application of FQS, it is better to allow as much interval as possible between updates of a gate. In this context, it makes sense that the vertical or horizontal updates, in which all gates are updated in a fixed order and each gate is updated once per sweep, are more efficient than the random update because they guarantee a maximum update interval.

Appendix H. Additional numerical experiments

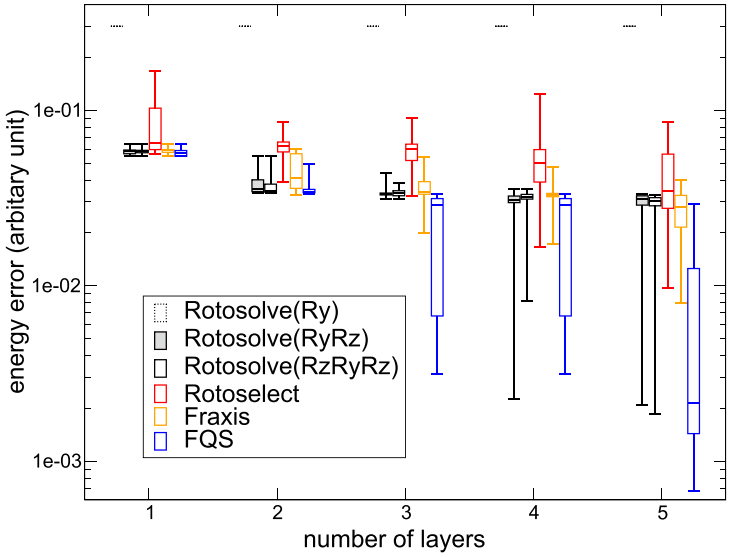


Figure 11. Boxplots of the resulting energy error of VQE for 5-qubit 1-dimensional mixed field Ising model after 100 sweeps. The grouped-layer PQC with ladder entangler in figure 9(c), which is also an alternating layered ansatz, was employed. The 20 independent VQEs were conducted from randomly-generated initial states.

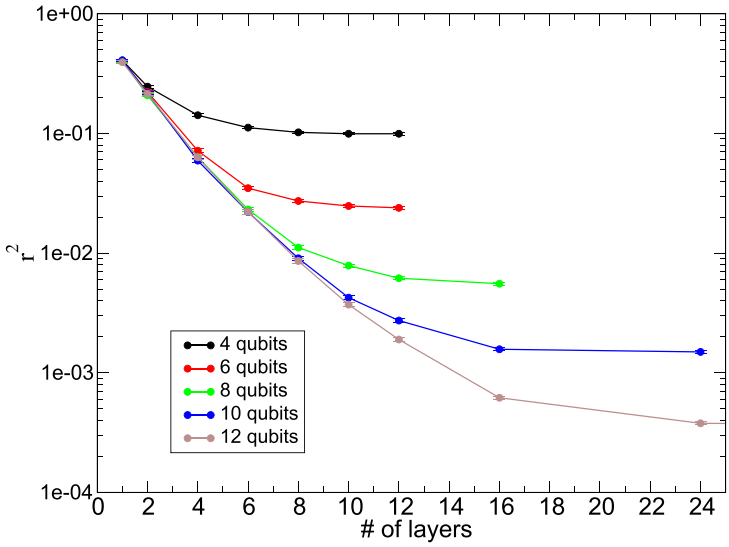


Figure 12. The second moment of spectrum radius of the centered FQS matrix. The FQS matrix was evaluated for the single-qubit gate acting on the first qubit in the first layer of the alternating layered ansatz with ladder entanglers as shown in figure 9(c). The second moment was evaluated with 1000 samples based on the randomly parameterized circuits. For the local cost function, we employed an expectation value of the Hamiltonian $H = Z \otimes I^{\otimes n-1}$.

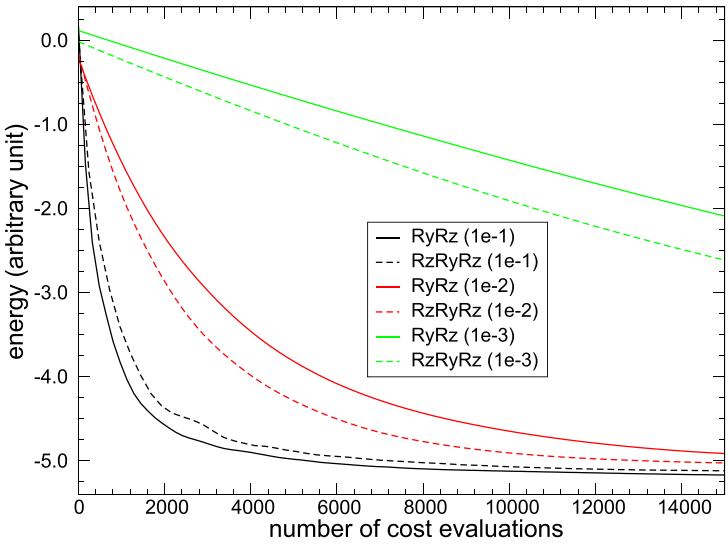


Figure 13. Comparison of the optimization trajectories of the VQE resulting energy by the Adam optimizer with different learning rates for 5-qubit 1-dimensional mixed field Ising model. Each trajectory is average over 20 independent runs. Black, red, and green lines represent the results of the learning rates of 0.1, 0.01 and 0.001, respectively. The solid and dash lines represent on the 5-layer alternating layered ansatz consisting of *RyRz* and *RzRyRz*, respectively.

ORCID iDs

Kaito Wada https://orcid.org/0000-0003-4976-4530
Rudy Raymond https://orcid.org/0000-0003-1005-6705
Yuki Sato https://orcid.org/0000-0003-2069-0355
Hiroshi C Watanabe https://orcid.org/0000-0003-4379-8633

References

[1] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 *Nat. Commun.* **5** 4213
[2] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 *Nature* **549** 242
[3] Grimsley H R, Economou S E, Barnes E and Mayhall N J 2019 *Nat. Commun.* **10** 3007
[4] Ma H, Govoni M and Galli G 2020 *npj Comput. Mater.* **6** 85
[5] Gao Q, Jones G O, Motta M, Sugawara M, Watanabe H C, Kobayashi T, Watanabe E, Ohnishi Y-Y, Nakamura H and Yamamoto N 2021 *npj Comput. Mater.* **7** 1
[6] Shikano Y, Watanabe H C, Nakanishi K M and Ohnishi Y-Y 2021 *Eur. Phys. J. Spec. Top.* **230** 1037
[7] Fuller B *et al* 2021 arXiv:2111.03167
[8] Amaro D, Rosenkranz M, Fitzpatrick N, Hirano K and Fiorentini M 2022 *EPJ Quantum Technol.* **9** 5
[9] Zoufal C, Mishmash R V, Sharma N, Kumar N, Sheshadri A, Deshmukh A, Ibrahim N, Gacon J and Woerner S 2022 arXiv:2205.03045
[10] Patti T L, Kossaifi J, Anandkumar A and Yelin S F 2023 *Quantum* **7** 1057
[11] Li Y and Benjamin S C 2017 *Phys. Rev.* X **7** 021050
[12] Yao Y-X, Gomes N, Zhang F, Wang C-Z, Ho K-M, Iadecola T and Orth P P 2021 *PRX Quantum* **2** 030307
[13] Benedetti M, Fiorentini M and Lubasch M 2021 *Phys. Rev. Res.* **3** 033083
[14] Wada K, Raymond R, Ohnishi Y-Y, Kaminishi E, Sugawara M, Yamamoto N and Watanabe H C 2022 *Phys. Rev.* A **105** 062421
[15] LaRose R, Tikku A, O'Neel-Judy É, Cincio L and Coles P J 2019 *npj Quantum Inf.* **5** 1
[16] Cerezo M, Sharma K, Arrasmith A and Coles P J 2022 *npj Quantum Inf.* **8** 1
[17] Bravo-Prieto C, LaRose R, Cerezo M, Subasi Y, Cincio L and Coles P J 2019 arXiv:1909.05820
[18] Xu X, Sun J, Endo S, Li Y, Benjamin S C and Yuan X 2021 *Sci. Bull.* **66** 2181
[19] Lubasch M, Joo J, Moinier P, Kiffner M and Jaksch D 2020 *Phys. Rev.* A **101** 010301
[20] Demirdjian R, Gunlycke D, Reynolds C A, Doyle J D and Tafur S 2022 *Quantum Inf. Process.* **21** 1
[21] Suzuki Y, Endo S, Fujii K and Tokunaga Y 2022 *PRX Quantum* **3** 010345
[22] Lin L and Tong Y 2020 *Quantum* **4** 372
[23] Zhang R, Wang G and Johnson P 2022 *Quantum* **6** 761
[24] Lin L and Tong Y 2022 *PRX Quantum* **3** 010318
[25] Zeng P, Sun J and Yuan X 2021 arXiv:2109.15304
[26] Cerezo M *et al* 2021 *Nat. Rev. Phys.* **3** 625
[27] Tilly J *et al* 2021 *Phys. Rep.* **986** 1
[28] Shen Y, Zhang X, Zhang S, Zhang J-N, Yung M-H and Kim K 2017 *Phys. Rev.* A **95** 020501
[29] Wecker D, Hastings M B and Troyer M 2015 *Phys. Rev.* A **92** 042303
[30] Wiersema R, Zhou C, de Sereville Y, Carrasquilla J F, Kim Y B and Yuen H 2020 *PRX Quantum* **1** 020319
[31] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 *Nat. Commun.* **9** 4812
[32] Wang S, Fontana E, Cerezo M, Sharma K, Sone A, Cincio L and Coles P J 2021 *Nat. Commun.* **12** 1

[33] Cerezo M, Sone A, Volkoff T, Cincio L and Coles P J 2021 *Nat. Commun.* **12** 1791
[34] Arrasmith A, Cerezo M, Czarnik P, Cincio L and Coles P J 2021 *Quantum* **5** 558
[35] Skolik A, McClean J R, Mohseni M, van der Smagt P and Leib M 2021 *Quantum Mach. Intell.* **3** 1
[36] Volkoff T and Coles P J 2021 *Quantum Sci. Technol.* **6** 025008
[37] Haghshenas R, Gray J, Potter A C and Chan G K-L 2022 *Phys. Rev.* X **12** 011047
[38] Nakanishi K M, Fujii K and Todo S 2020 *Phys. Rev. Res.* **2** 043158
[39] Ostaszewski M, Grant E and Benedetti M 2021 *Quantum* **5** 391
[40] Watanabe H C, Raymond R, Ohnishi Y-Y, Kaminishi E and Sugawara M 2023 *IEEE Trans. Quantum Eng.* **4** 1
[41] Slattery L, Villalonga B and Clark B K 2022 *Phys. Rev. Res.* **4** 023072
[42] Nakaji K and Yamamoto N 2021 *Quantum* **5** 434
[43] Vidal J G and Theis D O 2018 arXiv:1812.06323
[44] Parrish R M, Iosue J T, Ozaeta A and McMahon P L 2019 arXiv:1904.03206
[45] Koczor B and Benjamin S C 2022 *Phys. Rev. Res.* **4** 023017
[46] Wierichs D, Izaac J, Wang C and Lin C Y-Y 2022 *Quantum* **6** 677
[47] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press)
[48] Holmes Z, Sharma K, Cerezo M and Coles P J 2022 *PRX Quantum* **3** 010313
[49] Endo K, Sato Y, Raymond R, Wada K, Yamamoto N and Watanabe H C 2023 *Phys. Rev. Res.* **5** 043136
[50] Kingma D P and Ba J 2015 *3rd Int. Conf. on Learning Representations (ICLR 2015) (San Diego, CA, USA, 7–9 May 2015) (Conf. Track Proc.)*
[51] Cincio L, Subaşı Y, Sornborger A T and Coles P J 2018 *New J. Phys.* **20** 113022
[52] Rattew A G, Hu S, Pistoia M, Chen R and Wood S 2019 arXiv:1910.09694
[53] Tang H L, Shkolnikov V, Barron G S, Grimsley H R, Mayhall N J, Barnes E and Economou S E 2021 *PRX Quantum* **2** 020310
[54] Zhang Z-J, Kyaw T H, Kottmann J S, Degroote M and Aspuru-Guzik A 2021 *Quantum Sci. Technol.* **6** 035001
[55] Tkachenko N V, Sud J, Zhang Y, Tretiak S, Anisimov P M, Arrasmith A T, Coles P J, Cincio L and Dub P A 2021 *PRX Quantum* **2** 020337
[56] Bilkis M, Cerezo M, Verdon G, Coles P J and Cincio L 2021 arXiv:2103.06712
[57] Cincio L, Rudinger K, Sarovar M and Coles P J 2021 *PRX Quantum* **2** 010324
[58] Chivilikhin D, Samarin A, Ulyantsev V, Iorsh I, Oganov A and Kyriienko O 2020 arXiv:2007.04424
[59] Meng F-X, Li Z-T, Yu X-T and Zhang Z-C 2021 *IEEE Trans. Quantum Eng.* **2** 1
[60] Ostaszewski M, Trenkwalder L M, Masarczyk W, Scerri E and Dunjko V 2021 *Advances in Neural Information Processing Systems* vol 34 p 18182
[61] Du Y, Huang T, You S, Hsieh M-H and Tao D 2022 *npj Quantum Inf.* **8** 62
[62] Zhang S-X, Hsieh C-Y, Zhang S and Yao H 2022 *Quantum Sci. Technol.* **7** 045023
[63] Nakaji K *et al* 2024 arXiv:2401.09253
[64] Claudino D, Wright J, McCaskey A J and Humble T S 2020 *Front. Chem.* **8** 606863
[65] Feniou C, Claudon B, Hassan M, Courtat A, Adjoua O, Maday Y and Piquemal J-P 2023 arXiv:2306.17159
[66] Mc Keever C and Lubasch M 2023 *Phys. Rev. Res.* **5** 023146
[67] Abraham H *et al* 2019 Qiskit: an open-source framework for quantum computing *Zenodo* https://doi.org/10.5281/zenodo.2562110
[68] McKay D C, Wood C J, Sheldon S, Chow J M and Gambetta J M 2017 *Phys. Rev.* A **96** 022330