Optimal parameter configurations for sequential optimization of the variational quantum eigensolver

Katsuhiro Endo,[1,2,*] Yuki Sato,[3,2] Rudy Raymond,[4,2,5] Kaito Wada,[6] Naoki Yamamoto,[2,6] and Hiroshi C. Watanabe[2,7]

[1]*Research Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan*

[2]*Quantum Computing Center, Keio University, Hiyoshi 3-14-1, Kohoku-ku, Yokohama 223-8522, Japan*

[3]*Toyota Central R & D Labs., Inc., Koraku Mori Building 10F, 1-4-14 Koraku, Bunkyo-ku, Tokyo 112-0004, Japan*

[4]*IBM Quantum, IBM Japan, 19-21 Nihonbashi Hakozaki-cho, Chuo-ku, Tokyo 103-8510, Japan*

[5]*Department of Computer Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

[6]*Department of Applied Physics and Physico-Informatics, Keio University, Hiyoshi 3-14-1, Kohoku-ku, Yokohama 223-8522, Japan*

[7]*Department of Chemistry, Graduate School of Science, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan*

(Received 20 April 2023; accepted 12 September 2023; published 9 November 2023)

The variational quantum eigensolver (VQE) is a hybrid algorithm to find the minimum eigenvalue/vector of the given Hamiltonian by optimizing a parameterized quantum circuit (PQC) using a classical computer. Sequential optimization methods, which are often used in quantum circuit tensor networks, are popular for optimizing the parameterized gates of PQCs. In this paper, we focus on the case where the components to be optimized are single-qubit gates, in which the analytic optimization of a single-qubit gate is sequentially performed. The analytical solution is given by diagonalization of a matrix whose elements are computed from the expectation values of observables specified by a set of predetermined parameters, which we refer to as the parameter configurations. In this paper, we first show that the optimization accuracy significantly depends on the choice of parameter configurations owing to the statistical errors in the expectation values. We then identify a metric that quantifies the optimization accuracy of a parameter configuration for all possible statistical errors, named configuration overhead/cost or C-cost. We theoretically provide the lower bound of C-cost and show that, for the minimum size of parameter configurations, the lower bound is achieved if and only if the parameter configuration satisfies the so-called equiangular line condition. Finally, we provide numerical experiments demonstrating that the optimal parameter configuration exhibits the best result in several VQE problems. We hope that this general statistical methodology will enhance the efficacy of sequential optimization of PQCs for solving practical problems with near-term quantum devices.

DOI: 10.1103/PhysRevResearch.5.043136

I. INTRODUCTION

The variational quantum eigensolver (VQE) [1–3] is a classical-quantum hybrid algorithm that is implementable on near-term quantum devices, for finding the minimum eigenvalue/vector of a given Hamiltonian; the recipe is simply to prepare a parameterized quantum circuit (PQC) $U(\theta)$, also called *ansatz*, and then find a parameter θ that minimizes $\langle H\rangle = \langle\psi| U(\theta)^\dagger H U(\theta)|\psi\rangle$ with some initial state $|\psi\rangle$. Note that VQE is a class of variational quantum algorithms (VQAs) [4,5], where the cost, in general, is a nonlinear function of the expectation values of some Hamiltonians. A VQA has a wide range of applications, such as quantum chemical calculations [1,2,6], combinatorial optimization [7–9], and linear equation solvers [10–13].

The core question is how to model the PQC $U(\theta)$ and how to minimize $\langle H\rangle$ with some classical optimizer. There has been extensive investigation into this problem [4]. The *sequential optimization method* has been used in a variety of settings, such as quantum circuit tensor networks [14–18], where θ corresponds to a set of local unitaries, and they are sequentially optimized one by one. In this paper, we focus on the special type of sequential optimization method developed in Refs. [19–23]. In this framework, θ's are the parameters characterizing the set of single-qubit rotation gates such as $R_y(\theta) = e^{i\theta Y}$ (Y is the Pauli y matrix) in the case of RotoSolve [19,20]. Then the sequential optimization method takes the strategy to exactly optimize the single rotation gates one by one. For example, consider the step where we optimize the $R_y(\theta)$ gate contained in the PQC shown in Fig. 1 by minimizing the cost $\langle H\rangle$ as a function of θ. The point is that, in this case, $\langle H\rangle$ must be of the form of a sinusoidal function with respect to θ, and thus, the optimal $\theta_{\rm opt}$ can be exactly determined once we identify the sinusoidal function shown by the black curve in the figure. As the nature of

*katsuhiro.endo@aist.go.jp

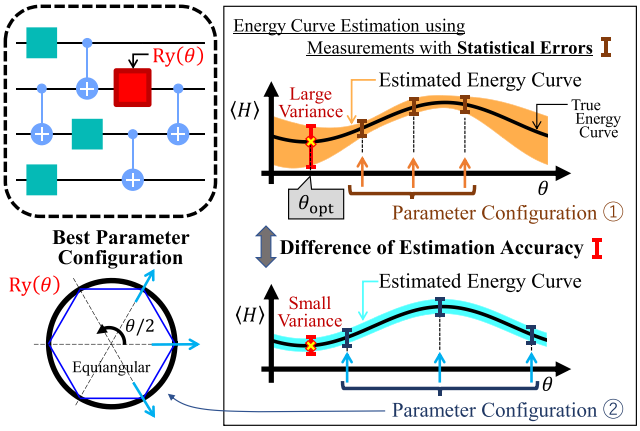

FIG. 1. A general view to explain how the estimated optimal solution varies depending on the parameter configuration when there are statistical errors in determining the cost function.

the sinusoidal function, specifying the mean values of three observables corresponding to the three points of θ allows us to exactly identify $\langle H\rangle$; we call the alignment of these three points of θ the *parameter configuration*. In the case of *free-axis selection* (FRAXIS) [21], where the freedom of a single-qubit rotation gate is served by the rotation axis with a fixed rotation angle in the Bloch sphere, $\langle H\rangle$ takes the form of a quadratic function of a real normalized vector $\mathbf{n} = (x, y, z)^T$, which can also be exactly minimized. This setup was further generalized to *free quaternion selection (FQS)* [22,23] so that the rotation angle can also be tuned; then $\langle H\rangle$ takes the form of a quadratic function of a real normalized vector $\mathbf{q} = (w, x, y, z)^T$. In this case, as shown later, the mean values of 10 observables corresponding to 10 points of $\mathbf{q}$ identify $\langle H\rangle$; we also call this $\{\mathbf{q}_1, \ldots, \mathbf{q}_{10}\}$ the parameter configuration. Note that a sequential optimization method for two-qubit gates has also been proposed, named the unitary block optimization scheme (UBOS) [24], where $\langle H\rangle$ takes a quadratic form of a unit weighting vector of two qubit Pauli strings. In UBOS, unlike ROTOSOLVE, FRAXIS, and FQS, the unit weighting vector cannot be arbitrary due to the unitary constraint and is parameterized by SU(4) gate parameters based on the Cartan decomposition. Thus, UBOS employs a numerical optimization of the quadratic form with respect to the parameters, which does not guarantee optimality.

The above sequential optimization strategy relies on the critical assumption that the mean values of observables, and accordingly $\langle H\rangle$, are exactly identified. In reality, however, those mean values can only be approximately obtained as the average of a finite number of measurement results; that is, practically, there is always a statistical error in $\langle H\rangle$. In the above one-dimensional (1D) case, as illustrated in Fig. 1, the energy curve θ_{opt}, and consequently the minimum value of $\langle H\rangle$, may largely fluctuate depending on the parameter configuration. Hence, the question is, what is the best parameter configuration for achieving a small fluctuation of $\min\langle H\rangle$? In the above 1D case, we have the intuition that the best configuration might be such that the three parameters are equally spaced (i.e., equidistant), as shown in the left bottom of Fig. 1, which is indeed true as proven later. However, the general case is, of course, nontrivial; will we have such an equidistant configuration in some sense, or would some biased configuration be the best?

In this paper, we develop the theory for determining the optimal parameter configuration. As a preliminary result, in Sec. II, we prove that, if the exact expectation values are available without any statistical error, then we have an analytical solution of the best parameters achieving $\min\langle H\rangle$ (almost) without respect to the parameter configuration for every method of Refs. [19–23]. Then in Sec. III, we give the most essential result providing the basis of the theory; that is, we derive the explicit form of the fluctuation of $\min\langle H\rangle$ under statistical errors with respect to the parameter configuration. This enables us to introduce the *C-cost* (configuration cost/overhead), a useful metric for determining $\min\langle H\rangle$ and thereby providing us with the optimal parameter configuration. Section IV gives numerical experiments to demonstrate that the optimal parameter configurations obtained using C-cost yield the best result in the sense of the statistical error of estimating $\langle H\rangle$.

Notably, beyond such utilization for numerically determining the configuration, the C-cost satisfies several interesting mathematical properties, suggesting the relevance of this metric. The first is that the lower bound of C-cost is 1; moreover, we prove that, for the minimum size of the parameter set, this bound is achievable if and only if the parameter configuration satisfies a geometric condition called the *equiangular line condition*, an important and beautiful mathematical concept in algebraic graph theory. Here, each parameter $\mathbf{q}$ corresponds to a line that passes the origin and $\mathbf{q}$. This condition rigorously supports our above-described intuition that it would be desirable for the parameters to be equally spaced for the ROTOSOLVE case shown in Figs. 1 or 2(a); this intuition holds for the case of FRAXIS, showing that there is a unique parameter configuration (up to the global rotation) satisfying the equiangular line condition, as displayed in Fig. 2(b). However, interestingly, this intuition does not apply to the most general FQS case due to the nonexistence of 10 equiangular lines in $\mathbb{R}^4$. That is, the so-called Gerzon bounds [25], Neumann theorem [26], and Haantjes bound [27] prove that a set of 10 lines satisfying the equiangular line condition in $\mathbb{R}^4$ does not exist; the maximum number of such lines is 6. Nevertheless, the C-cost is still useful in this case since it gives us a means to numerically obtain the optimal parameter configuration, which is displayed in Fig. 2(c). Furthermore, if redundant measurements are allowed, there exist parameter configurations that achieve the theoretical lower bound of the C-cost, one of which is illustrated in Fig. 2(d).

Finally, we note that equiangular lines in complex spaces are equivalent to symmetric informationally complete (SIC) positive operator-valued measures (POVMs) [28] whose properties have been much studied, e.g., it is conjectured that there is always a set of d^2 equiangular lines in $\mathbb{C}^d$ [29] (it has been proven up to some large d theoretically and numerically). The SIC POVMs defined from such lines are informationally complete because the results of other measurements can be computed from those of the SIC POVMs. In this paper, we obtain similar results connecting equiangular lines in real spaces with the variational quantum circuits using parameterized single-qubit gates.

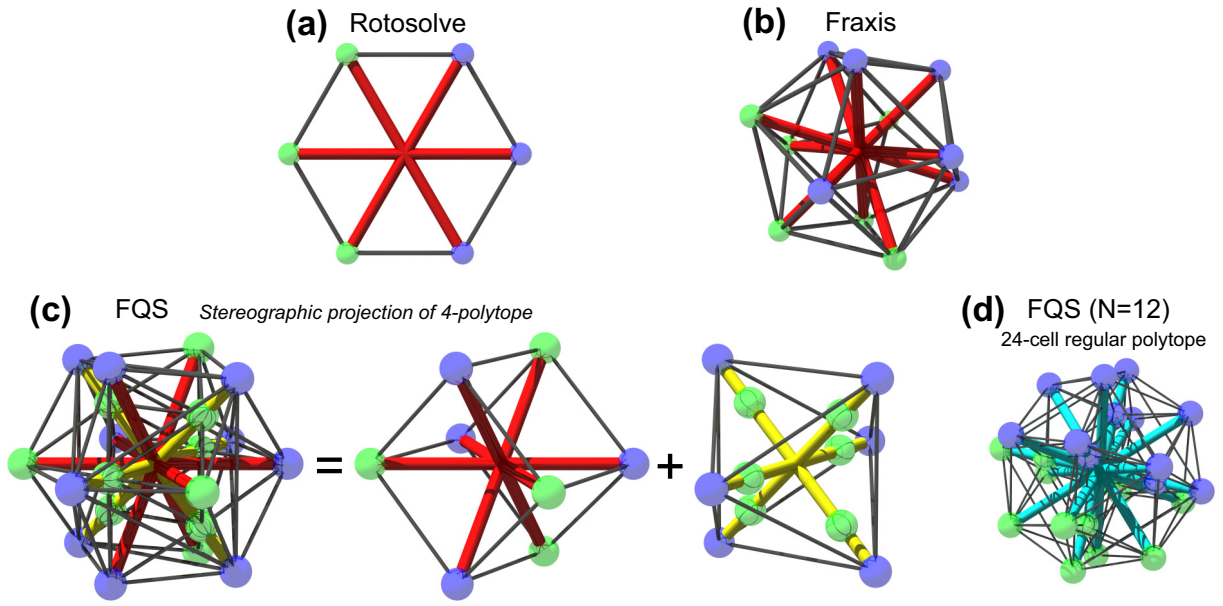

FIG. 2. Optimal parameter configurations for ROTOSOLVE, FRAXIS, and FQS models. Blue spheres represent the optimal configuration of $\{\mathbf{q}_i\}_{i=1}^N$, and green spheres represent its opposite position $\{-\mathbf{q}_i\}_{i=1}^N$. (a) ROTOSOLVE: The diagonal lines between $\mathbf{q}_i$ and $-\mathbf{q}_i$ constitute three equiangular lines in two-dimensional (2D) space (red lines). (b) FRAXIS: The diagonal lines constitute six equiangular lines in three-dimensional (3D) space. (c) FQS: The four positions and their opposite in the first term of the right-hand side constitute a regular cube in a hyperplane; the six positions in the second term (outer blues) constitute a regular octahedron in a hyperplane. The opposite of the last six positions (inner greens) also constitutes a regular octahedron. The yellow diagonals are apparently doubly overlapped due to the stereographic projection, but they are not actually overlapped. (d) FQS ($N = 12$): The 24-cell polytope in four-dimensional space, which achieves $C(A) = 1$ as a redundant parameter configuration.

II. ENERGY MINIMIZATION WITH MATRIX FACTORIZATION

A. Brief review of Rotosolve, Fraxis, and FQS

The FQS method [23] describes the procedure to completely characterize the energy landscape with respect to a single-qubit gate in a PQC. The parameterized single-qubit gate, which we call *the FQS gate*, is none other than the general single qubit gate $U^{(4)}$ expressed as [23,30]

$$U^{(4)}(\mathbf{q}) = wI - xiX - yiY - ziZ = \mathbf{q}\cdot\vec{\varsigma}, \tag{1}$$

where the superscript indicates the number of parameters: $\mathbf{q} = (w, x, y, z)^T \in \mathbb{R}^4$ satisfying $\|\mathbf{q}\|^2 = 1$. Here, i is the imaginary unit, I is the 2×2 identity matrix, and X, Y, Z are the Pauli matrices. Also, $\vec{\varsigma} = (\varsigma_I, \varsigma_X, \varsigma_Y, \varsigma_Z)^T$ denotes an extension of the Pauli matrices defined as

$$\vec{\varsigma} = (I, -iX, -iY, -iZ)^T. \tag{2}$$

The dimension of the parameter $\mathbf{q}$ is four, but because the parameter $\mathbf{q}$ is constrained on the unit hypersphere, the degree of freedom of the parameter is three.

In FRAXIS, the rotation angle is constrained to π, which corresponds to the case $w = 0$ of Eq. (1) as

$$U^{(3)}(\mathbf{n}) = -xiX - yiY - ziZ, \tag{3}$$

where the parameter of the gate is $\mathbf{n} = (x, y, z)^T$ such that $\|\mathbf{n}\|^2 = 1$. We term this $U^{(3)}$ as the *Fraxis gate*. Thus, the Fraxis gate has two degrees of freedom.

In ROTOSOLVE, the rotation axis is fixed, and the rotation angle serves as the parameter. The *Rx gate* fixes the rotation axis to the x axis; in the form of Eq. (1), this corresponds to $y = z = 0$, and thus,

$$U^{(2)}(\mathbf{r}) = wI - xiX, \tag{4}$$

where the parameter of the gate is $\mathbf{r} = (w, x)^T$ such that $\|\mathbf{r}\|^2 = 1$. Thus, the degree of freedom of the Rx gate is one. Similarly, Ry and Rz gates are obtained by replacing X in Eq. (4) with Y and Z, respectively.

In what follows, we use the most general FQS gate to describe the optimization algorithm. The sequential optimization method takes the strategy to update respective FQS gates in a coordinate-wise manner, where all parameters are fixed except for the focused FQS gate $U^{(4)}(\mathbf{q})$. The entire quantum circuit containing FQS gates is supposed to be the PQC $V = \prod_i U_i^{(4)}(\mathbf{q}_i)W_i$ on the n-qubit system, where $U_i^{(4)}$ is the ith FQS gate and W_i is a fixed multiqubit gate.

Now, let V_1 and V_2 be the gates placed before and after the focused FQS gate $U^{(4)}(\mathbf{q})$. Then a density matrix ρ prepared by the PQC is expressed as

$$\rho = V_2 U^{(4)}(\mathbf{q}) V_1 \rho_{\text{in}} V_1^\dagger [U^{(4)}(\mathbf{q})]^\dagger V_2^\dagger, \tag{5}$$

where ρ_{in} is an input density matrix. Thus, the expectation value $\langle H\rangle$ of a given Hamiltonian H with respect to ρ is then

$$\begin{aligned}\langle H\rangle &= \text{Tr}(HV_2U^{(4)}(\mathbf{q})V_1\rho_{\text{in}}V_1^\dagger[U^{(4)}(\mathbf{q})]^\dagger V_2^\dagger)\\ &= \text{Tr}(H'U^{(4)}(\mathbf{q})\rho'_{\text{in}}[U^{(4)}(\mathbf{q})]^\dagger),\end{aligned} \tag{6}$$

where $H' = V_2^\dagger H V_2$ and $\rho'_{\text{in}} = V_1\rho_{\text{in}}V_1^\dagger$. Substituting Eq. (1) into Eq. (6) yields

$$\langle H\rangle = \mathbf{q}^T G^{(4)}\mathbf{q}, \tag{7}$$

where $G^{(4)}$ is a 4×4 real-symmetric matrix:

$$G^{(4)} = \begin{bmatrix} G_{II} & G_{IX} & G_{IY} & G_{IZ} \\ G_{IX} & G_{XX} & G_{XY} & G_{XZ} \\ G_{IY} & G_{XY} & G_{YY} & G_{YZ} \\ G_{IZ} & G_{XZ} & G_{YZ} & G_{ZZ} \end{bmatrix}, \tag{8}$$

and each element $G_{\mu\nu}$ ($\mu, \nu = I, X, Y, Z$) is defined by

$$G_{\mu\nu} = \tfrac{1}{2}\mathrm{Tr}[\rho'_{\mathrm{in}}(\varsigma_\mu^\dagger H' \varsigma_\nu + \varsigma_\nu^\dagger H' \varsigma_\mu)]. \tag{9}$$

Thus, the energy landscape with respect to the FQS gate is completely characterized by the matrix $G^{(4)}$. Because Eq. (7) is a quadratic form with respect to the parameter $\mathbf{q}$ with the constraint $\|\mathbf{q}\|^2 = 1$, the eigenvector $\mathbf{p}_1$ associated with the lowest eigenvalue λ_1 of the matrix $G^{(4)}$ minimizes the energy in Eq. (7); see Appendix A for the details.

In the following, we call the matrix $G^{(4)}$ the *FQS matrix*. Note that the above result can be directly extended to the case of FRAXIS and ROTOSOLVE, in which case Eq. (8) is replaced by

$$G^{(3)} = \begin{bmatrix} G_{XX} & G_{XY} & G_{XZ} \\ G_{XY} & G_{YY} & G_{YZ} \\ G_{XZ} & G_{YZ} & G_{ZZ} \end{bmatrix}, \tag{10}$$

and

$$G^{(2)} = \begin{bmatrix} G_{II} & G_{IX} \\ G_{IX} & G_{XX} \end{bmatrix}, \tag{11}$$

respectively.

B. FQS with arbitrary parameter configurations

Because $G^{(4)}$ is a real-symmetric matrix, we can expand Eq. (7) in the following form:

$$\begin{aligned} \langle H \rangle = {} & G_{II}w^2 + G_{XX}x^2 + G_{YY}y^2 + G_{ZZ}z^2 \\ & + 2G_{IX}wx + 2G_{IY}wy + 2G_{IZ}wz \\ & + 2G_{XY}xy + 2G_{XZ}xz + 2G_{YZ}yz. \end{aligned} \tag{12}$$

Equation (12) indicates that, if we know all 10 coefficients $(G_{II}, \ldots, G_{YZ})$, we can exactly estimate the expectation $\langle H \rangle$ for any parameter $\mathbf{q}$. In other words, only algebraic calculations on classical computers are required to find the parameters achieving the minimum expectation value for the target gate.

Therefore, it is crucial to obtain the coefficients with as few measurements as possible. To consider this problem, we define the function $h^{(4)}(\mathbf{q})$ that outputs the normalized vector ($\|\mathbf{h}^{(4)}(\mathbf{q})\| = 1$):

$$\begin{aligned} \mathbf{h}^{(4)}(\mathbf{q}) = {} & (w^2, x^2, y^2, z^2, \sqrt{2}wx, \sqrt{2}wy, \sqrt{2}wz, \\ & \sqrt{2}xy, \sqrt{2}xz, \sqrt{2}yz)^T, \end{aligned} \tag{13}$$

and the vector $\mathbf{g}^{(4)}$:

$$\begin{aligned} \mathbf{g}^{(4)} = {} & (G_{II}, G_{XX}, G_{YY}, G_{ZZ}, \sqrt{2}G_{IX}, \sqrt{2}G_{IY}, \\ & \sqrt{2}G_{IZ}, \sqrt{2}G_{XY}, \sqrt{2}G_{XZ}, \sqrt{2}G_{YZ})^T. \end{aligned} \tag{14}$$

Then the relation between the parameter $\mathbf{q}$ and the expectation $\langle H \rangle$ is expressed as

$$\langle H \rangle = \mathbf{h}^{(4)}(\mathbf{q})^T \mathbf{g}^{(4)}. \tag{15}$$

Suppose measurements with different parameters $\{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$ and the N expectation values of the measurement results $\mathbf{b} = (b_1, \ldots, b_N)^T$ were obtained; we can also write the relations between the expectation values $\mathbf{b}$ and the coefficient vector $\mathbf{g}^{(4)}$ as

$$\mathbf{b} = A^{(4)}\mathbf{g}^{(4)}, \tag{16}$$

where the matrix $A^{(4)} \in \mathbb{R}^{N\times 10}$ is

$$A^{(4)} = [\mathbf{h}^{(4)}(\mathbf{q}_1), \ldots, \mathbf{h}^{(4)}(\mathbf{q}_N)]^T, \tag{17}$$

which encodes the information of the parameter configurations $\{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$.

It is obvious that, if $N < 10$, $\mathbf{g}^{(4)}$ is not uniquely determined. Hence, we suppose $N \geqslant 10$ throughout this paper. If $\mathrm{rank}(A) = 10$, A^TA is invertible, and there exists the generalized inverse $A^+ := (A^TA)^{-1}A^T$ [31]. Accordingly, we can obtain the vector $\mathbf{g}^{(4)}$ by solving linear equations as

$$\mathbf{g}^{(4)} = A^+\mathbf{b}. \tag{18}$$

In other words, a single execution of FQS requires at least 10 sets of parameters and the corresponding observables. However, it may not necessarily be the case when input states and/or the Hamiltonian has symmetry, which reduces the number of required measurements to construct $G^{(4)}$ in Eq. (8). We also note that it is possible that $\mathrm{rank}(A) < 10$ if the rows of A are dependent on each other. However, it is plausible to exclude such a situation because the input parameters are controllable. Hereafter, we suppose that all columns of A are independent of each other, equivalently, $\mathrm{rank}(A) = 10$.

The same argument is applicable to the Fraxis gate as

$$\begin{aligned} \langle H \rangle = {} & G_{XX}x^2 + G_{YY}y^2 + G_{ZZ}z^2 \\ & + 2G_{XY}xy + 2G_{XZ}xz + 2G_{YZ}yz, \end{aligned} \tag{19}$$

$$\mathbf{h}^{(3)}(\mathbf{n}) = (x^2, y^2, z^2, \sqrt{2}xy, \sqrt{2}xz, \sqrt{2}yz)^T, \tag{20}$$

$$\mathbf{g}^{(3)} = (G_{XX}, G_{YY}, G_{ZZ}, \sqrt{2}G_{XY}, \sqrt{2}G_{XZ}, \sqrt{2}G_{YZ})^T. \tag{21}$$

Likewise, for Rx gates

$$\langle H \rangle = G_{II}w^2 + G_{XX}x^2 + 2G_{IX}wx, \tag{22}$$

$$\mathbf{h}^{(2)}(\mathbf{r}) = (w^2, x^2, \sqrt{2}wx)^T, \tag{23}$$

$$\mathbf{g}^{(2)} = (G_{II}, G_{XX}, \sqrt{2}G_{IX})^T. \tag{24}$$

The minimum sizes of the parameter configuration required to construct $G^{(d)}$ are $d(d+1)/2$, i.e., 6 in FRAXIS ($d = 3$) and 3 in ROTOSOLVE ($d = 2$). For simplicity, we omit superscript d from $G^{(d)}$, $h^{(d)}$, and $\mathbf{g}^{(d)}$ for $d = 2, 3, 4$ in the following sections and formulate them based on the FQS framework unless otherwise noted.

III. C-COST WITH FINITE RUNS OF QUANTUM CIRCUITS

A. Evaluation of the parameter configurations

If the infinite number of measurements were allowed, there would be no estimation errors in the expectation values $\mathbf{b}$ and the resulting vector $\mathbf{g}$ is exactly obtained if the matrix A is invertible. This allows for the exact evaluation of the

optimal solution of the FQS matrix. In this section, we quantitatively evaluate the error propagation from the shot noise in the expectation values $\mathbf{b}$ to the estimation of the minimum expectation value for a target gate. Although we focus on the FQS for generality, it can be easily applied to other sequential quantum optimizers, ROTOSOLVE and FRAXIS. Suppose an FQS matrix is estimated using N parameters $\{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$ by Eq. (18). Due to the finite number of shots, the expectation values are no longer deterministic and randomly distributed around the true values $\mathbf{b}^*$ obtained with infinite shots as

$$\mathbf{b} = \mathbf{b}^* + \boldsymbol{\epsilon}, \tag{25}$$

where $\boldsymbol{\epsilon}$ is the random variable reflecting the errors in the measurements.

Note that the relation between $\mathbf{b}$ and $\mathbf{g}$ is no longer valid under the finite number of measurements. Alternatively, we employed the least-square solution $\mathbf{g}$:

$$\mathbf{g} = \arg\min_{\tilde{\mathbf{g}}} \|\mathbf{b} - A\tilde{\mathbf{g}}\|^2 = (A^T A)^{-1} A^T \mathbf{b} = A^+ \mathbf{b}, \tag{26}$$

as a plausible estimate of $\mathbf{g}^*$. Apparently, Eq. (26) has the same form as Eq. (18), but the resulting vector $\mathbf{g}$ is an estimate of the true vector $\mathbf{g}^*$ in the context of maximum likelihood [32] and deviates from the ideal vector $\mathbf{g}^*$ due to errors for finite measurement. Substituting Eq. (25) into Eq. (26), we get

$$\mathbf{g} = A^+ \mathbf{b} = A^+ (\mathbf{b}^* + \boldsymbol{\epsilon}) = \mathbf{g}^* + A^+ \boldsymbol{\epsilon}, \tag{27}$$

where the third equality follows $\mathbf{g}^* = A^+ \mathbf{b}^*$. Equation (27) implies the errors of the estimated coefficient vector $\mathbf{g} - \mathbf{g}^* = A^+ \boldsymbol{\epsilon}$ are amplified by the linear transformation A^+ from the shot errors $\boldsymbol{\epsilon}$.

Let G be an FQS matrix generated from the estimated vector $\mathbf{g}$ with a finite number of measurements. Below, we focus on the FQS procedure to estimate the minimum eigenvalue of G. Here, for convenience, we define the half-vectorization function vech : $\mathbb{R}^{4\times4} \to \mathbb{R}^{10}$ such that

$$\begin{aligned}\mathrm{vech}(G) = (&G_{II}, G_{XX}, G_{YY}, G_{ZZ}, G_{IX}, G_{IY}, G_{IZ},\\ &G_{XY}, G_{XZ}, G_{YZ})^T,\end{aligned} \tag{28}$$

where the order of elements corresponds to $\mathbf{g}$. In addition, the scaling matrix D is defined as

$$D = \mathrm{diag}(1, 1, 1, 1, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}, \sqrt{2}). \tag{29}$$

Using these notations, we have the following relations:

$$\mathbf{g} = D\,\mathrm{vech}(G) \Leftrightarrow G = \mathrm{vech}^{-1}(D^{-1}\mathbf{g}), \tag{30}$$

where the function vech^{-1} is a linear mapping as $\mathrm{vech}^{-1}(\mathbf{s} + \mathbf{t}) = \mathrm{vech}^{-1}(\mathbf{s}) + \mathrm{vech}^{-1}(\mathbf{t})$ for $\mathbf{s}, \mathbf{t} \in \mathbb{R}^{10}$. Accordingly, G is expressed as

$$G = \mathrm{vech}^{-1}(D^{-1}\mathbf{g}) = G^* + \mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon}), \tag{31}$$

which implies that the ideal FQS matrix $G^* = \mathrm{vech}^{-1}(D^{-1}A^+\mathbf{b}^*)$ is perturbed by $\mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})$.

In the following part, we quantitatively evaluate the matrix perturbation effect on the optimization result. Let λ_i^* and $\mathbf{p}_i^*$ be the ith lowest eigenvalue and the corresponding eigenvector of G^*. Likewise, $\lambda_i(\boldsymbol{\epsilon})$ and $\mathbf{p}_i(\boldsymbol{\epsilon})$ are the ith lowest eigenvalue and its corresponding eigenvector of the estimated matrix G. For quantitative evaluation of the perturbation, we suppose two metrics: (1) $\mathrm{Var}[\lambda_1(\boldsymbol{\epsilon})]$, the variance of the lowest eigenvalue which corresponds to the estimated minimum value of $\langle H \rangle$, and (2) $\mathbb{E}[\Delta E]$, the mean error in the expectation values after an FQS procedure with and without the shot noise. That is, ΔE is the deviation of the expectation value with the estimated parameter set $\mathbf{p}_1$ from the true minimum expectation value, defined as

$$\Delta E = \mathbf{p}_1^T G^* \mathbf{p}_1 - \mathbf{p}_1^{*T} G^* \mathbf{p}_1^* \geqslant 0, \tag{32}$$

where the positivity of ΔE comes from the fact that $\mathbf{p}_1^*$ gives the minimum value of the quadratic form. We suppose that $\mathrm{Var}[\lambda_1(\boldsymbol{\epsilon})]$ is a measure for verifying the estimated energy λ_1 by one-time execution of FQS, while $\mathbb{E}[\Delta E]$ is a measure to qualify the estimated parameter $\mathbf{p}_1$. Throughout the following parts, for simplicity, we employed $\mathrm{Var}[\lambda_1]$ as the indicator of shot errors (see Appendix B 3 for $\mathbb{E}[\Delta E]$).

Because G is a 4×4 symmetric matrix, it is represented by eigendecomposition as

$$G = P \Lambda P^T, \tag{33}$$

where $P = (\mathbf{p}_1, \ldots, \mathbf{p}_4)^T$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_4)$. From the first-order perturbation theory [33], the minimum eigenvalue λ_1 of G is approximated as

$$\lambda_1 = \lambda_1^* + \mathbf{p}_1^{*T}\,\mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})\mathbf{p}_1^*. \tag{34}$$

Then $\mathrm{Var}[\lambda_1]$ is evaluated as

$$\mathrm{Var}[\lambda_1] = \mathrm{Var}[\mathbf{p}_1^{*T}\,\mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})\mathbf{p}_1^*]. \tag{35}$$

To deal with Eq. (35), we apply a simple model to the measurement errors $\boldsymbol{\epsilon}$ satisfying as

$$\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}, \tag{36}$$

$$\mathbb{E}[\epsilon_i \epsilon_j] = \begin{cases} 0 & \text{for } i \neq j \\ \sigma^2/s & \text{for } i = j \end{cases}, \tag{37}$$

where s denotes the number of measurement shots to evaluate an expectation value of observables and σ^2 is a part to specific to observables.

In addition, we assume the first eigenvector $\mathbf{p}_1$ follows a uniform distribution on the unit sphere. Based on the models, Eq. (35) can be further calculated as

$$\mathrm{Var}[\lambda_1] = \frac{\sigma^2}{sd(d+2)}\mathrm{Tr}\left[(A^TA)^{-1}\left(\mathbf{1}_d\mathbf{1}_d^T + 2I\right)\right], \tag{38}$$

where $d = \dim(\mathbf{q})$ (4 for FQS, 3 for FRAXIS, and 2 for Rx) and $\mathbf{1}_d \in \mathbb{R}^{d(d+1)/2}$ is the vector that the first d elements are unity, and the others are zero (e.g., $\mathbf{1}_d = (1, 1, 1, 1, 0, 0, 0, 0, 0, 0)^T$ for FQS). Derivation of Eq. (38) is detailed in Appendix B 2.

For convenience, let us define the total number of shots S, which is required for a one-time optimization. Since the number of shots s is used to evaluate $\langle H \rangle$ with a single parameter $\mathbf{q}_i$, $S = sN$ holds for the parameter configuration $\{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$ of the size N. As a special case of s for $N = N_{\min} := d(d+1)/2$, we also define $\tilde{s}$, and thus, $S = \tilde{s}N_{\min}$. To focus on the optimization performance, suppose that N varies, while S is kept to constant.

Because $sN = \tilde{s}N_{\min}$ for a redundant parameter configuration with $N > N_{\min}$, $s = \tilde{s}N_{\min}/N$. As a result,

$$\mathrm{Var}[\lambda_1] = \frac{\sigma^2}{\tilde{s}} C(A), \tag{39}$$

where we define the C-cost $C(A)$, as

$$C(A) := \frac{N}{N_{\min} d(d+2)} \mathrm{Tr}\big[(A^T A)^{-1}\big(\mathbf{1}_d \mathbf{1}_d^T + 2I\big)\big]. \tag{40}$$

Equation (39) indicates that $\mathrm{Var}[\lambda_1]$ is separable into the number of shots ($\tilde{s}$) dependent and the parameter configuration dependent, i.e., a 50% reduction of $C(A)$ is equivalent to doubling the number of shots. The C-cost is a metric to estimate $\mathrm{Var}[\lambda_1]$ under the condition that the number of shots to optimize a single-qubit gate is constant.

Now, the conditions for the minimum $C(A)$ are of interest to minimize the estimation error. We rigorously give the lower bound of the C-cost as the following theorem (see Appendix C for the proof of this theorem):

Theorem 1. For the C-cost $C(A)$ in Eq. (40), $C(A) \geqslant 1$ holds with equality if and only if the parameter configurations $\{\mathbf{q}_i\}_{i=1}^N$ satisfy

$$A^T A = \frac{N}{d(d+2)}\big(\mathbf{1}_d \mathbf{1}_d^T + 2I\big). \tag{41}$$

In other words, the parameter configuration that satisfies Eq. (41) is optimal for minimizing the effect of shot noise. Although it may not be straightforward to find the optimal parameter sets that satisfy Eq. (41), in the case of the minimum parameter set ($N = N_{\min}$), a useful formula is available as the following corollary of Theorem 1 (see Appendix C for the proof):

Corollary 1. For the minimum number of parameters ($N = N_{\min}$), the C-cost $C(A)$ in Eq. (40) is always $C(A) \geqslant 1$ with equality if and only if the parameter configurations $\{\mathbf{q}_i\}_{i=1}^N$ satisfy

$$|\mathbf{q}_i \cdot \mathbf{q}_j| = \frac{1}{\sqrt{d+2}} \quad (\text{for all } i \neq j). \tag{42}$$

The equality condition in Corollary 1 tells us that the parameters must be equiangular unit vectors. This equiangular property is known as *equiangular lines* in real spaces [25,34,35], equivalent to the algebraic graph theory of *two-graphs* [36]. The existence of $N_{\min} = d(d+1)/2$ equiangular lines in $\mathbb{R}^d$ is known as the *Gerzon bounds* and has only been shown to hold for $d = 2, 3, 7,$ and 23. For our optimal parameter configurations, only in the cases of Rx and Fraxis gates ($d = 2, 3$), there exists a unique set of $N_{\min}$ equiangular unit vectors (up to rotation), and such a parameter configuration uniquely achieves the minimum value of C-cost $C(A)$. The nonexistence of such an optimal parameter configuration for an FQS gate ($d = 4$) is due to the nonexistence of equiangular lines satisfying the condition of Corollary 1, which is attributed to Haantjes [27] and Neumann in Ref. [25] (see also Ref. [26]).

B. The rotation invariance of C-cost

The C-cost $C(A)$ in Eq. (40) is invariant to the rotation of a parameter configuration. In other words, a parameter configuration $(\mathbf{q}_1, \dots, \mathbf{q}_K)$ and its rotated configuration $(R\mathbf{q}_1, \dots, R\mathbf{q}_K)$ have the same value of the C-cost, where $R \in \mathbb{R}^{d\times d}$ is a rotation matrix ($R^T R = I$). See Appendix D for the proof of rotation invariance. This implies that, for any parameter $\mathbf{q}$ of a single-qubit gate of interest, there exists an optimal parameter configuration $\{\mathbf{q}_i\}$ such that $\mathbf{q} \in \{\mathbf{q}_i\}$. This property allows us to skip evaluating an expectation value for one parameter in the matrix construction, i.e., to reduce the number of required expectation values to 2 for ROTOSOLVE, 5 for FRAXIS, and 9 for FQS by reusing the previous results to the subsequent gate update. The reduction for ROTOSOLVE has been known [20] but not for FRAXIS and FQS. In each step of the sequential optimizations, the resulting cost value after the parameter update can be estimated without additional measurement. As all parameters are fixed except for that of the target gate, this estimated cost can be regarded as one of the observable expectation values b_1 in the subsequent application, where the parameter $\mathbf{q}_1$ of the next gate of interest is diverted from the previous application.

The detailed procedure is as follows: (1) Prepare an optimal parameter configuration $\{\mathbf{q}^*\}$, the gate parameter set $\{\mathbf{q}^{(m)}\}$ for $m = 1, \dots, M$, and the temporal cost value $\langle H\rangle(\{\mathbf{q}^{(m)}\})$, where m and M denote the gate index and the total number of parameterized gates, respectively. (2) Find a rotation matrix R such that $\mathbf{q}_1^* = R^T\mathbf{q}^{(m)}$, where the mth gate is of interest and sets $b_1 = \langle H\rangle$. (3) Measure the cost values with the parameter $\{R\mathbf{q}_i^*\}$ for $i = 2, \dots, N_{\min}$, setting $b_i = \langle H\rangle(R\mathbf{q}_i^*)$. (4) Construct the matrix G from $\mathbf{b}$ and $\{R\mathbf{q}_i^*\}$. (5) Diagonalize the matrix to estimate the new parameter $\mathbf{q}^{(m)}$ and the new cost $\langle H\rangle$, which can be reused in the next iteration and go back to (2) until convergence. The pseudocode of this procedure is given in Algorithm 1.

Algorithm 1. Algorithm to reuse optimization results of the previous gate.

Input: The parameter $\mathbf{q}^{\mathrm{pre}}$ of the target gate, the estimated minimum eigenvalue λ^{pre} in the previous FQS, and the optimal parameter configurations $\{\mathbf{q}_1^*, \dots, \mathbf{q}_N^*\}$.
Output: The optimized parameter of the target gate $\mathbf{q}^{\mathrm{opt}}$ and the updated cost λ.
Find a rotation matrix R such that $\mathbf{q}^{\mathrm{pre}} = R\mathbf{q}_1^*$.
Set $b_1 = \lambda^{\mathrm{pre}}$ [instead of measuring $b_1 = \langle H\rangle(\mathbf{q}_1^*)$].
for $i = 2$ to N **do**
    Measure $b_i = \langle H\rangle(R\mathbf{q}_i^*)$.
Set $G = \mathrm{vech}^{-1}(DA^+\mathbf{b})$.
Diagonalize G and obtain the minimum eigenvalue λ and the corresponding eigenvector $\mathbf{q}^{\mathrm{opt}}$.
Return $\mathbf{q}^{\mathrm{opt}}$ and λ.

C. Optimal configurations

The minimum sizes of parameter configuration ($N_{\min}$) for Rx, FRAXIS, and FQS are 3, 6, and 10, respectively. According to Corollary 1 in the case of the Rx gate, the three equiangular vectors on a unit circle are trivially represented by $\mathbf{q} = [\cos\frac{2}{3}\pi n\theta, \sin\frac{2}{3}\pi n\theta]^T$ for $n = 0, \pm 1$, that is, the vector angle $\Delta\theta = 2\pi/3$ (equivalently $\pi/3$), as shown in Fig. 2(a). Conversely, the original parameter configuration proposed in

ROTOSOLVE [20] was $\Delta\theta = \pi/2$, which resulted in $C(A) = \frac{3}{2}$. As our optimal parameter configuration ($\Delta\theta = 2\pi/3$) gives $C(A) = 1$, it requires two-thirds as many shots as the original parameter configuration ($\Delta\theta = \pi/2$) to achieve the same estimation accuracy. It is worth noting that Ref. [19] argued that arbitrary parameter configurations can be used due to the sinusoidal property of the expectation value but did not discuss the estimation accuracy dependent on the parameter configurations under the finite measurements. The use of $\Delta\theta = 2\pi/3$ was also proposed in Ref. [37], but no quantitative arguments about its advantages were given.

Corollary 1 is also instrumental for FRAXIS with $d = 3$ because it is also possible to find the equiangular formation of six unit vectors in three-dimensional (3D) space. Figure 2(b) shows the unique optimal parameter configuration except for the rotational degrees of freedom, where they form a regular icosahedron. The original parameter configuration of FRAXIS has $C(A) = 1.8$ [21] (see Appendix E). Thus, the optimal configuration improves the estimation accuracy 1.8 times with a consistent number of shots.

Conversely, it was proved that $N_{\min}$ (=10) equiangular unit vectors cannot be placed in d (= 4)-dimensional Euclidean space. Namely, Corollary 1 tells us that no parameter configuration satisfies $C(A) = 1$ for $N = 10$. In addition, Corollary 1 also implies that the minimum size of the parameter configuration ($N = 10$) may not be the most efficient if the total number of shots is limited for a single FQS execution, even though it is not straightforward to know the analytical minimum value and the corresponding parameter configurations. Instead, we searched the numerical solution by classical optimization, where $C(A)$ is minimized based on the gradient descent method. As the algorithm may lead to a local minimum solution, we repeated the algorithm independently 10^5 times, starting from random initial configurations.

For $N = 10$, we have obtained the same optimized C-cost value [$C(A) \approx 1.033172$] from all the initial configurations as far as our experimental trials, which implies that all simulations presumably reached the global minimum. Although the obtained configurations were not numerically identical, we found that they were attributed to a unique configuration just by reversal and rotational operations. Since the reversal of each parameter does not affect the expectation value [i.e., $h(\mathbf{q}) = h(-\mathbf{q})$] and the uniform rotation of the parameter configuration gives the identical value of the C-cost (see Sec. III B), all the configurations were equivalent, which seem to be optimal.

Figure 2(c) shows the unique optimal parameter configurations for the FQS case. In this figure, the parameter configurations are projected into 3D space by a stereographic projection. It means that extra 1D components that cannot be displayed are projected in the radial direction. See Appendix E for the parameter values of the optimal and other parameter configurations. From the parameter values of the (numerically obtained) optimal parameter configuration [Eq. (E7)], we can see the optimal parameter configuration has highly symmetrical structure; the first four parameters $\{\mathbf{q}_1, \ldots, \mathbf{q}_4\}$ and their opposite $\{-\mathbf{q}_1, \ldots, -\mathbf{q}_4\}$ constitute a regular cube in a hyperplane, and the last six parameters $\{\mathbf{q}_5, \ldots, \mathbf{q}_{10}\}$ constitute a regular octahedron in a hyperplane (their opposite also constitute another regular octahedron), as shown in Fig. 2(c).

TABLE I. C-cost values for the different sizes of parameter configurations of FQS. (A) Comparison of the C-cost $C(A)$ with a constant number of shots for evaluating an expectation value. (B) Comparison of scaled C-cost $(N-1)C(A)/N$ with a constant number of shots per single-gate optimization. Boldface indicates best performance.

N	10	11	12
(A)	1.03317	1.00539	**1.00000**
(B)	0.92985	**0.91399**	0.91667

For FQS, the original parameter configuration has $C(A) = 3.0$, and the optimal parameter configuration estimated with numerical experiments is approximately $C(A) \approx 1.033172$. Thus, to achieve a certain accuracy, the optimal parameter configuration reduces the number of required shots 3 times that of the original.

Likewise, we also conducted numerical optimization to find the optimal parameter configuration for redundant measurements with $N = 11, 12$. As a result, all the optimizations converged to a consistent value of $C(A)$ within computational precision, which is consistent with the case of $N = N_{\min}$. However, the optimal configurations are not necessarily unique, which is in contrast to $N = N_{\min}$. While the obtained $C(A) \approx 1.005390$ for $N = 11$, $C(A)$ was exactly converged to unity for $N = 12$. It is also notable that the optimal configurations of $C(A) = 1$ for $N = 12$ include the regular 24-cell polytope in four-dimensional (4D) space, as shown in Fig. 2(d).

Therefore, if the total number of shots for A matrix construction is constant, the optimal sizes of N are 3 for ROTOSOLVE, 6 for FRAXIS, and 12 for FQS.

Next, we focus on the optimal N, allowing the reduction of measurements exploiting the rotation invariance as mentioned in Sec. III B. Assuming a constant number of shots per gate, the measurement reduction modifies the relation between $C(A)$ and $\tilde{s}$ as

$$\mathrm{Var}[\lambda_1^*] = \frac{\sigma^2}{\tilde{s}} \frac{N-1}{N} C(A), \tag{43}$$

where the C-cost is apparently scaled by $(N-1)/N$. Note that this factor does not change the optimal N for ROTOSOLVE and FRAXIS. Thus, it is most efficient to revert the estimated value in the previous optimization to construct A and additionally execute 2 and 5 measurements for ROTOSOLVE and FRAXIS, respectively. It is worth noting that Table I shows that the optimal N for FQS is shifted from 12 to 11 by measurement reduction, although the difference is <1%. Altogether, under the limitation of the total number of shots, it is most efficient to construct the matrix A by 3-, 6-, and 12-type measurements for the expectation values at the beginning of ROTOSOLVE, FRAXIS, and FQS optimizations, respectively. Conversely, during the sequential optimization, matrix A should be made by one estimation value from the previous step and 2, 5, and 10 values from subsequent measurements of ROTOSOLVE, FRAXIS, and FQS, respectively.

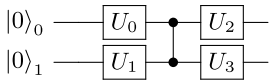

FIG. 3. Two-qubit ansatz.

It should be also noted that this optimal condition may differ depending on the supposed condition of real devices. For instance, if parallel computation is allowed, where a constant number of shots is available for evaluating an expectation value even though when N varies, $C(A)$ would not be an appropriate metric because the assumption about the number of shots is not valid, and thus, one should trivially employ as large N as possible.

IV. EXPERIMENTS

In the following, we provide several experiments to numerically verify our proposed method on the condition of $N = N_{\min}$; it is 10 for FQS, 6 for FRAXIS, and 3 for ROTOSOLVE.

A. Estimation accuracy of one-time optimization with different parameter configurations

We focused on the one-time optimization rather than an entire VQE process. To this end, we examined the averaged error of FQS between the exact minimum and the estimated minimum energies with a limited number of shots for several parameter configurations. We used the two-qubit hydrogen moleculelike Hamiltonian [38], defined as

$$H = I \otimes Z + Z \otimes I + X \otimes X \tag{44}$$

in this experiment. We use the two-qubit ansatz in Fig. 3, where we applied the corresponding single-qubit gate representation of ROTOSOLVE (=RzRy), FRAXIS, and FQS methods to U_i. Here, the target gate to be optimized is U_2 for FQS and FRAXIS, and the Ry gate in U_2 for the ROTOSOLVE case. The experiments were performed with the following procedure: (1) Prepare 100 distinct parameter configurations (100 sets of $\{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$) by sampling each $\mathbf{q}_i$ in the state random manner. (2) Optimize the respective parameter configurations by the steepest decent with 50 iterations using $C(A)$ as a cost function. This process is required to put the $C(A)$ distribution in a modest range. Otherwise, the initial random parameter configurations often have extremely large values of $C(A)$ beyond 10^4. (3) Evaluate A^+ and $C(A)$ of the 100 parameter configurations. (4) Randomly initialize the PQC in the state-random manner for the respective single-qubit gates [21]. (5) Obtain $\mathbf{b}$ (and $\mathbf{b}^*$) by the observable measurements based on the 100 parameter configurations, and evaluate FQS matrices G (and G^*) using the respective sets of A^+, $\mathbf{b}$ (and $\mathbf{b}^*$). (6) Execute FQS (FRAXIS/ROTOSOLVE) for G (and G^*) to obtain $\mathbf{p}$ (and $\mathbf{p}^*$). We repeat the procedure (4)–(6) 10^4 times and evaluate the averaged error $\langle \Delta E \rangle$ in Eq. (32) for each parameter configuration. In Fig. 4, we plotted 100 independent parameter configurations in the $C(A)$ vs $\langle \Delta E \rangle$ graph. By definition, $C(A)$ and $\langle \Delta E \rangle$ are metrics to qualify the estimated energy and the estimated parameter, respectively. Although both metrics are linked through the following equation:

$$\frac{N_{\min} d}{N} C(A) + \frac{sd(d-1)}{k\sigma^2} \mathbb{E}[\Delta E] = \mathrm{Tr}[(A^T A)^{-1}], \tag{45}$$

the concrete behaviors are not necessarily trivial because of dependency on A and the observable. Here, we confirmed that the energy errors $\langle \Delta E \rangle$ are roughly proportional to $C(A)$ for all the cases, and $\langle \Delta E \rangle$ is inversely proportional to the number of shots approximately. We also found that the optimal parameter configuration (red) achieves the lowest error $\langle \Delta E \rangle$, indicating that the optimal parameter configurations are effective in minimizing the estimation error. Although the magnitude of $\langle \Delta E \rangle$ in FQS is seemingly larger than that of ROTOSOLVE, we note that it does not necessarily indicate the advantage of ROTOSOLVE with respect to error suppression because the single-gate expressibility is not comparable among the respective methods. For instance, sequential ROTOSOLVE applications of a series of three single-qubit gates are comparable to one-time FQS applications. In this case, however, it is not straightforward to compare them because of error propagation, which is beyond the present framework. In the next section, instead, we examine the effect of the parameter configuration on the entire performance in comparison with the optimization methods.

B. The influence of the C-cost on VQE performances

We investigate the effect of different parameter configurations on the results of VQE when we sequentially optimize single-qubit gates in quantum circuits by the framework of FQS [23]. We employed the five-qubit quantum Heisenberg model [39], defined as

$$H = J \sum_{i=1}^{5} \sum_{\sigma=X,Y,Z} \sigma_i \sigma_{i+1} + h \sum_{i=1}^{5} Z_i, \tag{46}$$

where $\sigma_i = I^{\otimes i-1} \otimes \sigma \otimes I^{\otimes 5-i} (1 \leqslant i \leqslant 5)$, $\sigma_6 = \sigma_1$. We herein set $J = h = 1$. We used a Cascading-block ansatz, shown in Fig. 5, where the gates within the dashed lines are repeated L times. We set $L = 1, 3, 5$ in this experiment. According to the optimization method, we applied the respective single-qubit representations to U_i in the PQC. We begin VQE by randomly initializing PQC in a state-random manner for respective single-qubit gates in the PQC. In VQE, we sequentially applied ROTOSOLVE/FRAXIS/FQS to U_i in the order of subscripts in Fig. 5, i.e., from the top left to right bottom. We term this procedure to update all gates in the PQC once as a *sweep*. In a single VQE run, we carried out 100 sweeps to obtain the estimated minimum eigenvalue E of the Hamiltonian. We performed independent 100 VQE runs and plotted the error distribution $\Delta E := E - E^*$ for respective 100 trials in Fig. 6, where E^* is the exact minimum eigenvalue of the Hamiltonian. We evaluated the resulting distributions using the number of shots to 100, 1000, 10 000, and ∞ for two or three different parameter configurations (see Appendix E for the specific parameter values). Note that we used a statevector for VQE with an infinite number of shots. Figure 6 suggests that parameter configurations strongly affect the entire VQE performance and shows that the optimized parameter configuration [$C(A) \simeq 1$] achieves

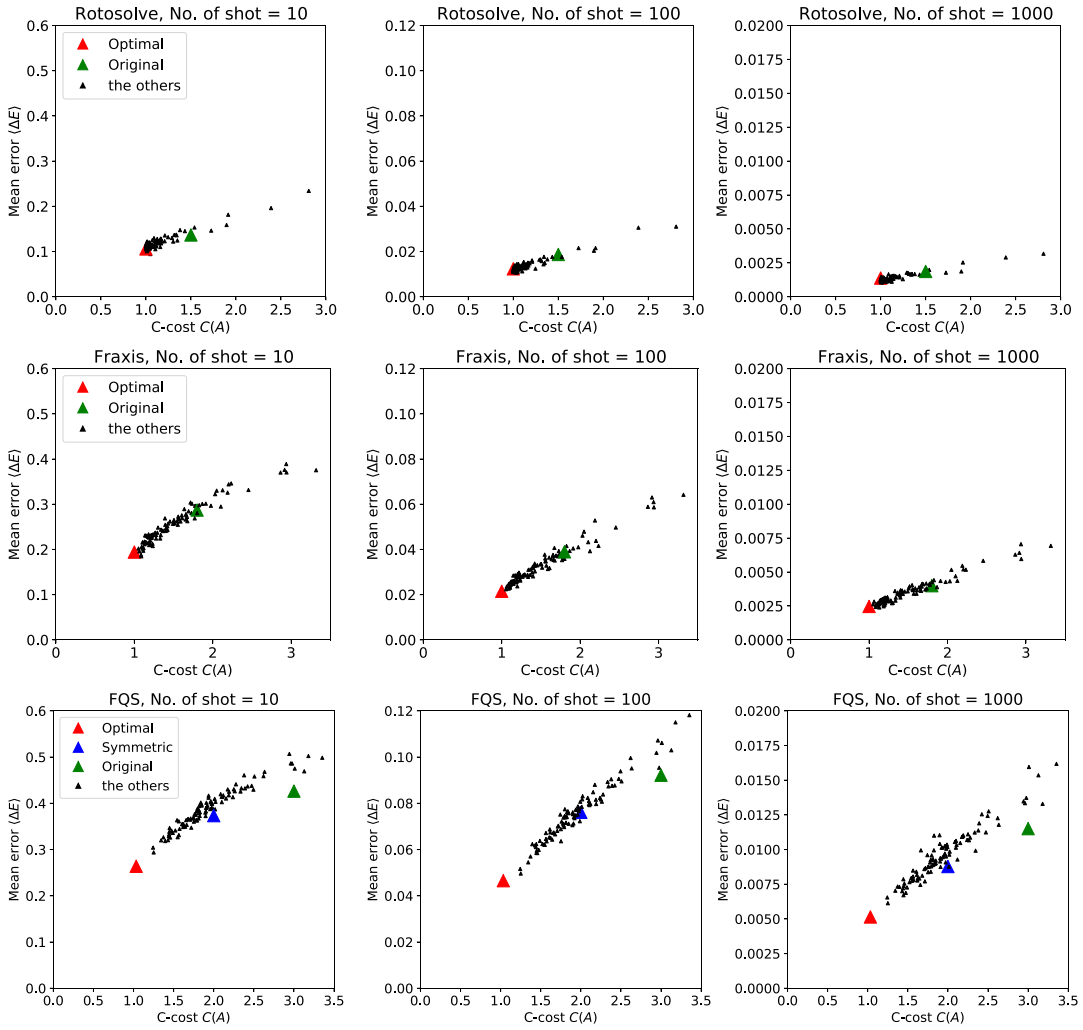

FIG. 4. The average energy error for one-time optimization with different parameter configurations. Each subplot shows the averaged deviations of the estimated minimum from the true minimum energy (vertical axis), where the former energy was evaluated from G made with randomly generated parameter configurations under the limited total number of shots, while the latter energy was obtained via STATEVECTOR SIMULATOR. The left, center, and right columns show the results with 10, 100, and 1000 shots per circuit, respectively. The top, center, and bottom rows show the results for ROTOSOLVE, FRAXIS, and FQS, respectively. The results of the original and optimal parameter configurations are highlighted in the figure. The description of the number of shots above each subplot represents the number of shots used for a single mean value of the Hamiltonian based on a parameter configuration.

the smallest errors on all the conditions with finite numbers of shots. The optimal parameter configuration works more effectively as the number of shots is smaller, which is in line with the analysis of the one-time application to a single-qubit gate in Fig. 4. In addition, the impact of the parameter configuration on the VQE performance is not visible on the shallow circuit and more distinct as the number of the layer increases. In general, more expressive ansatz can potentially approximate the state of interest with higher precision. Correspondingly, one has to increase the number of shots because, for accuracy ϵ, the number of required shots scales in $O(1/\epsilon^2)$. Otherwise, the enhanced expressibility by the circuit extension may not be highlighted. As the gain of $C(A)$ is equivalent to the increase of measurements, the optimal parameter configuration will be more beneficial as the desired accuracy in VQE becomes higher. In fact, FQS is superior to ROTOSOLVE and FRAXIS, and the statevector simulation implies that FQS with the ansatz of $L = 5$ can potentially achieve the accuracy $\epsilon < 10^{-2}$. However, this energy level was not achieved by the original and symmetric parameter configurations with the 10 000 shots, which is a practical standard for the present quantum devices, e.g., IBM-Q device.

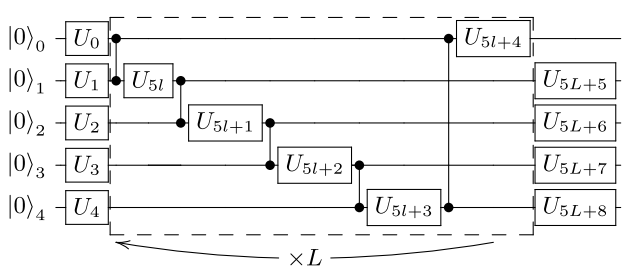

FIG. 5. Cascading-block ansatz.

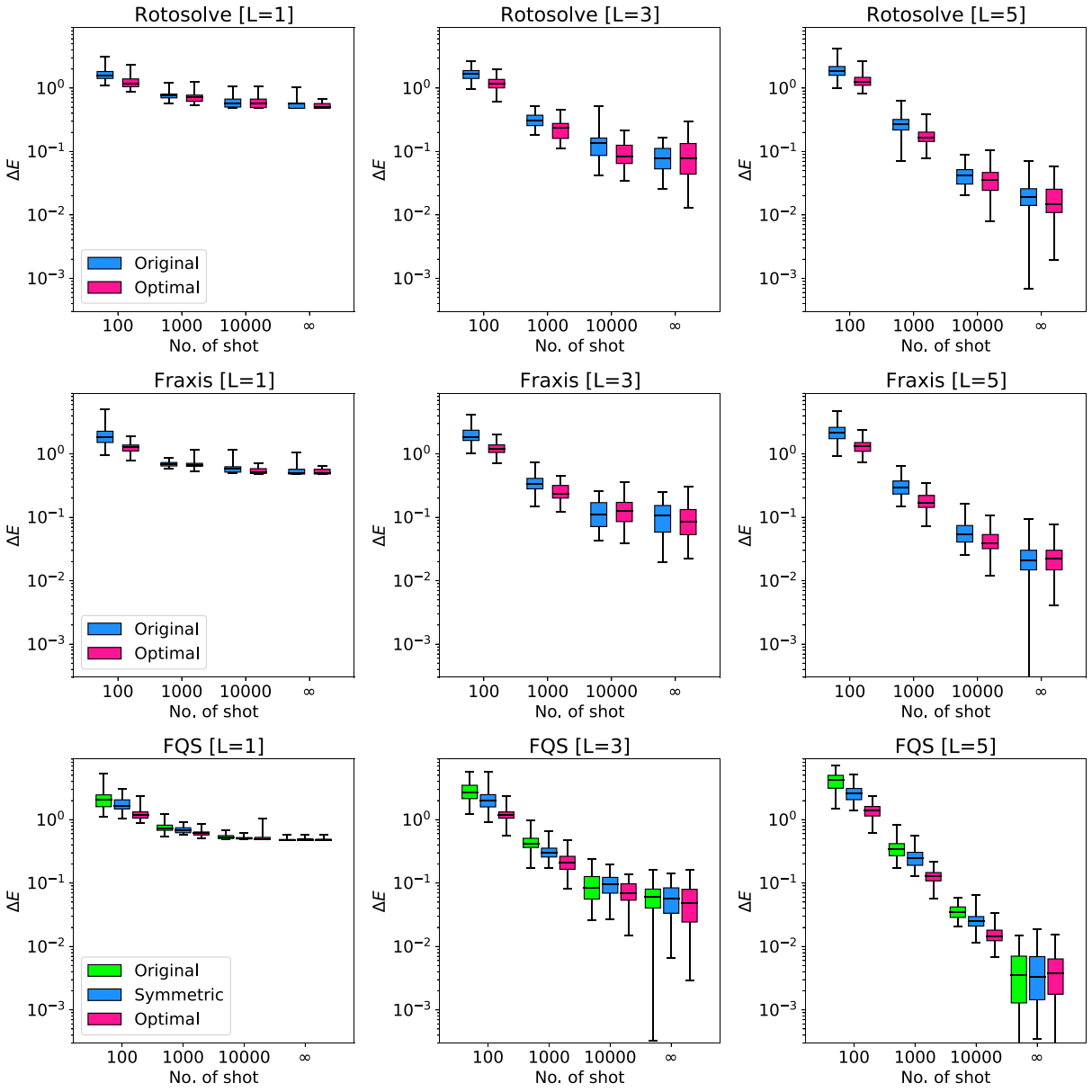

FIG. 6. Comparison of the variational quantum eigensolver (VQE) performance with different parameter configurations. The vertical axis represents the deviation of the resulting VQE energy from the exact ground-state energy. The respective energies were evaluated after 100 sweeps, where one sweep stands for sequential updates of all the single-qubit once. The box-and-whisker plots show the statistics of the energy deviations (ΔE) obtained by independent 100 VQE runs. We carried out the VQEs with the circuit layers (L) from 1, 3, and 5 and showed the results in respective subplots.

Therefore, the parameter configuration optimization assists VQE in lowering the reachable energy level distinctively, although it is not the case for ROTOSOLVE and FRAXIS because the number of shots available is sufficient relative to their expressibility.

V. CONCLUSIONS

In this paper, we showed that the parameter configuration affects the performance of analytical optimization of a single-qubit gate. This estimation error was quantified by the C-cost $C(A)$, the variance of the estimated value of the cost function. We theoretically proved that the lower bound of $C(A)$ is unity. We also showed that, when the size of the parameter configuration is minimal, the C-cost becomes unity if and only if the parameter configuration satisfies the equiangular condition. Exploiting this property, we found the optimal parameter configuration for ROTOSOLVE and FRAXIS. Although we revealed no parameter configuration of minimum size for FQS achieves $C(A) = 1$, it turned out the parameter configuration of $N = 12$ corresponding to the regular 24-cell polytope in 4D space satisfies $C(A) = 1$. In addition, we also demonstrated how to reduce the number of measurements for matrix construction by making use of the rotation invariance of $C(A)$. Then the optimal parameter configurations exhibited the best results, improving efficiency by 1.5 times for ROTOSOLVE, 1.8 times

for FRAXIS, and 3.0 times for FQS, compared with the original parameters. Additional numerical experiments showed that the parameter configuration affects the performance of not only the one-time optimization but also the entire VQE. We also found that the parameter configuration is more instrumental in eliciting the VQE performance as the ansatz becomes more expressive.

ACKNOWLEDGMENTS

R.R. would like to thank Prof. David Avis of Kyoto University for discussion on equiangular lines. H.C.W. was supported by JSPS Grants No. 20K03885 and 23K03266. N.Y. and H.C.W. were supported by the MEXT Quantum Leap Flagship Program Grants No. JPMXS0118067285 and No. JPMXS0120319794.

APPENDIX A: FREE QUATERNION SELECTION

We show the minimum value of Eq. (7) is the minimum eigenvalue λ_1 of G achieved when $\mathbf{q} = \mathbf{p}_1$ for the corresponding eigenvector $\mathbf{p}_1$ of G.

For the Lagrange multiplier method, we first define a function $l(\mathbf{q}, \lambda)$ corresponding to the above optimization problem as

$$l(\mathbf{q}, \lambda) = \mathbf{q}^T G\mathbf{q} - \lambda(\|\mathbf{q}\|^2 - 1), \tag{A1}$$

where λ is a Lagrange multiplier. Taking the partial derivatives for $l(\mathbf{q}, \lambda)$ and setting them to zero, we can obtain

$$G\mathbf{q} = \lambda\mathbf{q}. \tag{A2}$$

Thus, the candidates for the local minimum/maximum value of $l(\mathbf{q}, \lambda)$ and the solutions are the eigenvalues λ_i and its normalized eigenvectors $\mathbf{p}_i$, respectively.

Substituting the normalized eigenvectors $\mathbf{p}_i$ into Eq. (7), we get

$$\langle H\rangle = \mathbf{p}_i^T G\mathbf{p}_i = \mathbf{p}_i^T(\lambda_i\mathbf{p}_i) = \lambda_i. \tag{A3}$$

This means the global minimum value of Eq. (7) and its solution are given by the minimum eigenvalue λ_1 and the corresponding normalized eigenvector $\mathbf{p}_1$.

APPENDIX B: DERIVATION OF ANALYTICAL FORM OF THE MEASURES

1. Expectation value over an orthogonal basis

We show several equations that are useful for the derivation of the analytical form of the measures.

Let $Z \in \mathbb{R}^{d\times d}$ be a random symmetric matrix that satisfies $\mathbb{E}[Z_{ij}] = 0$ for all i, j. Independently, let $P = (\mathbf{p}_1, \ldots, \mathbf{p}_d)^T \in \mathbb{R}^{d\times d}$ be a random orthogonal matrix [i.e., the matrix is uniformly sampled from the orthogonal group $O(4)$]. Then the following equations hold:

$$\begin{aligned}\mathbb{E}\big[\mathbf{p}_i^{*T} Z\mathbf{p}_j^*\big] &= \sum_{k,l}\mathbb{E}[(\mathbf{p}_i^*)_k(\mathbf{p}_j^*)_l Z_{kl}]\\ &= \sum_{k,l}\mathbb{E}[(\mathbf{p}_i^*)_k(\mathbf{p}_j^*)_l]\mathbb{E}[Z_{kl}]\\ &= 0,\end{aligned} \tag{B1}$$

and so,

$$\begin{aligned}\mathrm{Var}\big[\mathbf{p}_i^{*T} Z\mathbf{p}_j^*\big] &= \mathbb{E}\big[\big(\mathbf{p}_i^{*T} Z\mathbf{p}_j^*\big)^2\big] - \mathbb{E}\big[\mathbf{p}_i^{*T} Z\mathbf{p}_j^*\big]^2\\ &= \mathbb{E}\big[\big(\mathbf{p}_i^{*T} Z\mathbf{p}_j^*\big)^2\big].\end{aligned} \tag{B2}$$

For $i = j$,

$$\begin{aligned}\mathrm{Var}\big[\mathbf{p}_i^{*T} Z\mathbf{p}_i^*\big] &= \mathbb{E}\left\{\left[\sum_{k,l}(\mathbf{p}_i)_k Z_{kl}(\mathbf{p}_j)_l\right]^2\right\}\\ &= \sum_{k,l,m,n}\mathbb{E}[(\mathbf{p}_i)_k(\mathbf{p}_i)_l(\mathbf{p}_i)_m(\mathbf{p}_i)_n]\mathbb{E}[Z_{kl}Z_{mn}]\\ &= \sum_{k(=l)}\sum_{m(=n\neq k)}\mathbb{E}\big[(\mathbf{p}_i)_k^2(\mathbf{p}_i)_m^2\big]\mathbb{E}[Z_{kk}Z_{mm}]\\ &\quad+\sum_{k(=m)}\sum_{l(=n\neq k)}\mathbb{E}\big[(\mathbf{p}_i)_k^2(\mathbf{p}_i)_l^2\big]\mathbb{E}\big[Z_{kl}^2\big]\\ &\quad+\sum_{k(=n)}\sum_{l(=m\neq k)}\mathbb{E}\big[(\mathbf{p}_i)_k^2(\mathbf{p}_i)_l^2\big]\mathbb{E}[Z_{kl}Z_{lk}]\\ &\quad+\sum_{k}\mathbb{E}\big[(\mathbf{p}_i)_k^4\big]\mathbb{E}\big[Z_{kk}^2\big]\\ &= \sum_{k,l}\frac{\mathbb{E}[Z_{kk}Z_{ll}] + \mathbb{E}\big[Z_{kl}^2\big] + \mathbb{E}[Z_{kl}Z_{lk}]}{d(d+2)}\\ &= \sum_{k,l}\frac{\mathbb{E}[Z_{kk}Z_{ll}] + 2\mathbb{E}\big[Z_{kl}^2\big]}{d(d+2)}.\end{aligned} \tag{B3}$$

For the fourth equality, we employed the following relation:

$$\mathbb{E}[(\mathbf{x}^T Z\mathbf{x})^2] = d(d+2)\mathbb{E}\left[\left(\frac{\mathbf{x}^T}{\|\mathbf{x}\|}Z\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)^2\right], \tag{B4}$$

where $\mathbf{x}$ is a random vector in $\mathbb{R}^d$, which follows the d-dimensional multivariate standard normal distributions $\mathcal{N}(\mathbf{0}, I)$, and $\mathbf{x}, Z$ are independent of each other.

To evaluate Eq. (B2) for $i \neq j$, we suppose another random vector $\mathbf{y}$ as $\mathbf{x}$, but independent of $\mathbf{x}$ and Z:

$$\begin{aligned}\mathbb{E}[(\mathbf{x}^T Z\mathbf{y})^2] &= \mathbb{E}\left\{\left[\sum_{i,j}(\mathbf{x})_i Z_{ij}(\mathbf{y})_j\right]^2\right\}\\ &= \mathbb{E}\left[\sum_{i,j,s,t}(\mathbf{x})_i(\mathbf{x})_s Z_{ij}Z_{st}(\mathbf{y})_j(\mathbf{y})_t\right]\\ &= \sum_{i,j,s,t}\mathbb{E}[(\mathbf{x})_i(\mathbf{x})_s]\mathbb{E}[(\mathbf{y})_j(\mathbf{y})_t]\mathbb{E}[Z_{ij}Z_{st}]\\ &= \sum_{i,j,s,t}\delta_{j,s}\delta_{i,t}\mathbb{E}[Z_{ij}Z_{st}]\\ &= \sum_{i,j}\mathbb{E}\big[Z_{ij}^2\big].\end{aligned} \tag{B5}$$

Here, we introduce two vectors as

$$\mathbf{y}_\parallel = \frac{(\mathbf{y}\cdot\mathbf{x})}{\|\mathbf{x}\|^2}\mathbf{x}, \quad \mathbf{y}_\perp = \mathbf{y} - \mathbf{y}_\parallel. \tag{B6}$$

Using these vectors, we obtained the following relation:

$$\begin{aligned}\mathbb{E}[(\mathbf{x}^T Z\mathbf{y})^2] &= \mathbb{E}\{[\mathbf{x}^T Z(\mathbf{y}_\parallel + \mathbf{y}_\perp)]^2\} \\ &= \mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\parallel)^2] + \mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\perp)^2] \\ &\quad + 2\mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\parallel)(\mathbf{x}^T Z\mathbf{y}_\perp)] \\ &= \mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\parallel)^2] + \mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\perp)^2]. \end{aligned} \tag{B7}$$

For the third equality, we use the probability distribution f that satisfies $f(\mathbf{x}, \mathbf{y}_\parallel, \mathbf{y}_\perp) = f(\mathbf{x}, \mathbf{y}_\parallel, -\mathbf{y}_\perp)$, and thus,

$$\mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\parallel)(\mathbf{x}^T Z\mathbf{y}_\perp)] = \mathbb{E}\{(\mathbf{x}^T Z\mathbf{y}_\parallel)[\mathbf{x}^T Z(-\mathbf{y}_\perp)]\}, \tag{B8}$$

equivalently,

$$\mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\parallel)(\mathbf{x}^T Z\mathbf{y}_\perp)] = 0. \tag{B9}$$

On the other hand,

$$\begin{aligned}\mathbb{E}\left\{\left[\mathbf{p}_i^{*T} Z\mathbf{p}^*_{j(\neq i)}\right]^2\right\} &= \mathbb{E}\left[\left(\frac{\mathbf{x}^T}{\|\mathbf{x}\|} Z \frac{\mathbf{y}_\perp}{\|\mathbf{y}_\perp\|}\right)^2\right] \\ &= \frac{1}{\mathbb{E}[\|\mathbf{x}\|^2]\mathbb{E}[\|\mathbf{y}_\perp\|^2]}\mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\perp)^2] \\ &= \frac{1}{d(d-1)}\mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\perp)^2], \end{aligned} \tag{B10}$$

where we suppose that the probability distribution f satisfies $f(\mathbf{x}/\|\mathbf{x}\|, \mathbf{y}_\perp/\|\mathbf{y}_\perp\|) = f(\mathbf{p}_i, \mathbf{p}_{j(\neq i)})$, a.e., $\mathbb{E}[\|\mathbf{x}\|^2] = d$ and $\mathbb{E}[\|\mathbf{y}_\perp\|^2] = d - 1$.

In addition, the first term in Eq. (B7):

$$\begin{aligned}\mathbb{E}[(\mathbf{x}^T Z\mathbf{y}_\parallel)^2] &= \mathbb{E}\left[\|\mathbf{x}\|^2\|\mathbf{y}_\parallel\|^2\left(\frac{\mathbf{x}^T}{\|\mathbf{x}\|} Z \frac{\mathbf{y}_\parallel}{\|\mathbf{y}_\parallel\|}\right)^2\right] \\ &= \mathbb{E}[\|\mathbf{x}\|^2]\mathbb{E}[\|\mathbf{y}_\parallel\|^2]\mathbb{E}\left[\left(\frac{\mathbf{x}^T}{\|\mathbf{x}\|} Z \frac{\mathbf{y}_\parallel}{\|\mathbf{y}_\parallel\|}\right)^2\right] \\ &= d\,\mathbb{E}\left[\left(\mathbf{p}_i^{*T} Z\mathbf{p}_i^*\right)^2\right] \\ &= \sum_{k,l}\frac{\mathbb{E}[Z_{kk}Z_{ll}] + 2\mathbb{E}\left[Z_{kl}^2\right]}{d+2}. \end{aligned} \tag{B11}$$

where the second equality arises from the independence of the random variables, and the third equality is based on $\mathbf{x}/\|\mathbf{x}\| = \mathbf{y}_\parallel/\|\mathbf{y}_\parallel\|$, $f(\mathbf{x}/\|\mathbf{x}\|) = f(\mathbf{p}_i)$, a.e., $\mathbb{E}[\|\mathbf{x}\|^2] = d$ and $\mathbb{E}[\|\mathbf{y}_\parallel\|^2] = 1$.

From Eqs. (B3), (B7), (B10), and (B11), we finally obtain

$$\begin{aligned}\mathrm{Var}\left[\mathbf{p}_i^{*T} Z\mathbf{p}^*_{j(\neq i)}\right] &= \mathbb{E}\left\{\left[\mathbf{p}_i^{*T} Z\mathbf{p}^*_{j(\neq i)}\right]^2\right\} \\ &= \sum_{k,l}\frac{d\mathbb{E}\left[Z_{kl}^2\right] - \mathbb{E}[Z_{kk}Z_{ll}]}{d(d-1)(d+2)}. \end{aligned} \tag{B12}$$

2. Derivation of analytical form of Var[$\lambda_1(\epsilon)$]

Using the noise model $\langle\epsilon_i\rangle = \mathbf{0}$, $\langle\epsilon_i\epsilon_j\rangle = \sigma^2\delta_{i,j}/s$ and the uniformly distributed model of the first eigenvector $\mathbf{p}_1^*$. We show the analytical form of Var[$\lambda_1(\boldsymbol{\epsilon})$] in Eq. (35):

$$\mathrm{Var}[\lambda_1(\boldsymbol{\epsilon})] = \mathrm{Var}\left[\mathbf{p}_1^{*T}\mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})\mathbf{p}_1^*\right].$$

For simplicity, we write $Z = \mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})$. Note that Z is a symmetric matrix and satisfies $\mathbb{E}[Z] = O$ because $\mathbb{E}[D^{-1}A^+\boldsymbol{\epsilon}] = D^{-1}A^+\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}$. Thus, using Appendix B 1,

$$\mathrm{Var}\left[\mathbf{p}_1^{*T} Z\mathbf{p}_1^*\right] = \frac{\sum_{i,j}\mathbb{E}[Z_{ii}Z_{jj}] + 2\mathbb{E}\left[Z_{ij}^2\right]}{d(d+2)}. \tag{B13}$$

Then we deal with the first term $\sum_{i,j}\mathbb{E}[Z_{ii}Z_{jj}]$ and the second term $\sum_{i,j}\mathbb{E}[Z_{ij}^2]$ separately. To this end, we introduce some useful representations. We note Eq. (15) can be rewritten as

$$\langle H\rangle = h(\mathbf{q})^T\,\mathbf{g} = (\mathbf{q}^T \otimes \mathbf{q}^T)\mathrm{vec}(G). \tag{B14}$$

Here, $\mathbf{q}$ is the parameter of the target single-qubit gate, G is the FQS matrix, and $\mathrm{vec} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d^2}$ is the vectorization operator for matrices.

Next, we introduce a linear transformation $L \in \mathbb{R}^{d^2\times d(d+1)/2}$ between the vector $\mathbf{g}$ and vec(G) in the Rx, FRAXIS, and FQS gates as

$$L = \begin{bmatrix}1&0&0\\0&0&c\\0&0&c\\0&1&0\end{bmatrix}, \quad \begin{bmatrix}1&0&0&0&0&0\\0&0&0&c&0&0\\0&0&0&0&c&0\\0&0&0&c&0&0\\0&1&0&0&0&0\\0&0&0&0&0&c\\0&0&0&0&c&0\\0&0&1&0&0&0\end{bmatrix}, \quad \begin{bmatrix}1&0&0&0&0&0&0&0&0&0\\0&0&0&0&c&0&0&0&0&0\\0&0&0&0&0&c&0&0&0&0\\0&0&0&0&0&0&c&0&0&0\\0&0&0&0&c&0&0&0&0&0\\0&1&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&c&0&0\\0&0&0&0&0&0&0&0&c&0\\0&0&0&0&0&c&0&0&0&0\\0&0&0&0&0&0&c&0&0&0\\0&0&1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&c\\0&0&0&0&0&0&c&0&0&0\\0&0&0&0&0&0&0&0&c&0\\0&0&0&0&0&0&0&0&0&c\\0&0&0&1&0&0&0&0&0&0\end{bmatrix}, \tag{B15}$$

respectively, where $c = 1/\sqrt{2}$. Note that the transformation satisfies

$$L^T L = I, \tag{B16}$$

$$\mathrm{vec}(G) = L\mathbf{g}, \tag{B17}$$

and so,

$$(\mathbf{q}^T \otimes \mathbf{q}^T)L = h(\mathbf{q})^T. \tag{B18}$$

We may also consider the inverse transformation of vectorization $\mathrm{vec}^{-1} : \mathbb{R}^{d^2} \to \mathbb{R}^{d\times d}$ as

$$\mathrm{vec}^{-1}(L\mathbf{g}) = G \quad \forall \mathbf{g} \in \mathbb{R}^{d(d+1)/2}, \tag{B19}$$

and $\mathrm{vech}^{-1} : \mathbb{R}^{d^2} \to \mathbb{R}^{d\times d}$,

$$\mathrm{vech}^{-1}(D^{-1}\mathbf{g}) = G \quad \forall \mathbf{g} \in \mathbb{R}^{d(d+1)/2}. \tag{B20}$$

This leads to

$$Z = \mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon}) = \mathrm{vec}^{-1}(LA^+\boldsymbol{\epsilon}). \tag{B21}$$

Then the first term in Eq. (B13) is rewritten as

$$\begin{aligned}
&\sum_{i,j}\mathbb{E}[Z_{ii}Z_{jj}] \\
&= \mathbb{E}\left[\sum_{i,j}\mathrm{vec}^{-1}(LA^+\boldsymbol{\epsilon})_{ii}\mathrm{vec}^{-1}(LA^+\boldsymbol{\epsilon})_{jj}\right] \\
&= \mathbb{E}\{[\mathrm{vec}(I)^T LA^+\boldsymbol{\epsilon}][\mathrm{vec}(I)^T LA^+\boldsymbol{\epsilon}]\} \\
&= \frac{\sigma^2}{s}[\mathrm{vec}(I)^T LA^+ \cdot \mathrm{vec}(I)^T LA^+] \\
&= \frac{\sigma^2}{s}\mathrm{vec}(I)^T LA^+(A^+)^T L^T \mathrm{vec}(I) \\
&= \frac{\sigma^2}{s}\mathbf{1}_d^T A^+(A^+)^T \mathbf{1}_d \\
&= \frac{\sigma^2}{s}\mathbf{1}_d^T (A^T A)^{-1}\mathbf{1}_d,
\end{aligned} \tag{B22}$$

where I is the identity matrix, and $\mathbf{1}_d := L^T\mathrm{vec}(I)$ is the vector in $\mathbb{R}^{d(d+1)/2}$ whose first d element is unity, and the rest are zero. For the sixth equality, we used the following relation:

$$\begin{aligned}
A^+(A^+)^T &= [(A^TA)^{-1}A^T][(A^TA)^{-1}A^T]^T \\
&= (A^TA)^{-1}A^TA(A^TA)^{-1} \\
&= (A^TA)^{-1}.
\end{aligned} \tag{B23}$$

The second term in Eq. (B13) is also rewritten as

$$\begin{aligned}
\sum_{i,j}\mathbb{E}\big[Z_{ij}^2\big] &= \mathbb{E}\left[\sum_{i,j}\mathrm{vec}^{-1}(LA^+\boldsymbol{\epsilon})_{ij}\mathrm{vec}^{-1}(LA^+\boldsymbol{\epsilon})_{ij}\right] \\
&= \mathbb{E}[(LA^+\boldsymbol{\epsilon})^T(LA^+\boldsymbol{\epsilon})] \\
&= \frac{\sigma^2}{s}\mathrm{Tr}[(LA^+)^T(LA^+)] \\
&= \frac{\sigma^2}{s}\mathrm{Tr}[A^+(A^+)^T L^T L] \\
&= \frac{\sigma^2}{s}\mathrm{Tr}[(A^TA)^{-1}],
\end{aligned} \tag{B24}$$

Summarizing Eqs. (B13), (B22), and (B24), $\mathrm{Var}[\lambda_1(\boldsymbol{\epsilon})]$ is expressed as

$$\begin{aligned}
\mathrm{Var}[\lambda_1] &= \frac{\sigma^2}{sd(d+2)}\left\{\mathbf{1}_d^T(A^TA)^{-1}\mathbf{1}_d + 2\mathrm{Tr}[(A^TA)^{-1}]\right\} \\
&= \frac{\sigma^2}{sd(d+2)}\mathrm{Tr}\big[(A^TA)^{-1}\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big)\big],
\end{aligned} \tag{B25}$$

where the following identity:

$$\mathbf{1}_d^T(A^TA)^{-1}\mathbf{1}_d = \mathrm{Tr}\big[(A^TA)^{-1}\mathbf{1}_d\mathbf{1}_d^T\big], \tag{B26}$$

is employed for the last equality.

3. Discussion of $\mathbb{E}[\Delta E]$ for the perturbation effect

Using the second-order perturbation theory of the matrix in Ref. [40], the energy error ΔE is approximated as

$$\Delta E = \sum_{i>1}^{d}\frac{\big[\mathbf{p}_i^{*T}\mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})\mathbf{p}_1^*\big]^2}{\lambda_i^* - \lambda_1^*}. \tag{B27}$$

Note that Eq. (B27) is not applicable when the lowest-energy eigenstate is degenerated. However, the following argument has been found to hold well experimentally. This equation leads to

$$\mathbb{E}[\Delta E] = \mathbb{E}\left\{\sum_{i>1}^{d}\frac{\big[\mathbf{p}_i^{*T}\mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})\mathbf{p}_1^*\big]^2}{\lambda_i^* - \lambda_1^*}\right\}. \tag{B28}$$

However, unlike $\mathrm{Var}[\mathbf{p}_1(\boldsymbol{\epsilon})]$, this measure also depends on the probability distribution $f(\lambda_1^*, \ldots, \lambda_d^*)$ of the eigenvalues of the matrix G. Assuming these eigenvalues are independent of each other, that is,

$$f(\lambda_1^*, \ldots, \lambda_M^*) = \prod_{i=1}^{d} f(\lambda_i^*), \tag{B29}$$

and the matrix of the eigenvectors $P = (\mathbf{p}_1, \ldots, \mathbf{p}_d)^T$ is a random orthogonal, Eq. (B28) can be written as

$$\begin{aligned}
\mathbb{E}[\Delta E] &= \sum_{i>1}^{d}\mathbb{E}\left[\frac{1}{\lambda_i^* - \lambda_1^*}\right] \\
&\quad\times \mathbb{E}\left\{\big[\mathbf{p}_i^{*T}\mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})\mathbf{p}_1^*\big]^2\right\} \\
&= k\mathbb{E}\left\{\big[\mathbf{p}_2^{*T}\mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})\mathbf{p}_1^*\big]^2\right\},
\end{aligned} \tag{B30}$$

where $k := \sum_{i>1}^{d}\mathbb{E}[(\lambda_i^* - \lambda_1^*)^{-1}]$. This means the measure $\mathbb{E}[\Delta E]$ can be evaluated with some modeling of the true FQS matrix G and the measurement errors $\boldsymbol{\epsilon}$.

For simplicity, we now write $Z = \mathrm{vech}^{-1}(D^{-1}A^+\boldsymbol{\epsilon})$. From Eq. (B12):

$$\begin{aligned}
\mathbb{E}[\Delta E] &= k\sum_{i,j}\frac{d\mathbb{E}\big[Z_{ij}^2\big] - \mathbb{E}[Z_{ii}Z_{jj}]}{d(d-1)(d+2)} \\
&= \frac{k\sigma^2}{sd(d-1)(d+2)}\left\{d\,\mathrm{Tr}[(A^TA)^{-1}] - \mathbf{1}_d^T(A^TA)^{-1}\mathbf{1}_d\right\} \\
&= \frac{k\sigma^2}{sd(d-1)(d+2)}\mathrm{Tr}\big[(A^TA)^{-1}\big(dI - \mathbf{1}_d\mathbf{1}_d^T\big)\big].
\end{aligned} \tag{B31}$$

In addition, if $C(A) = 1$, i.e., the case of a theoretical lower bound, $A^T A = \frac{N}{d(d+2)}(2I + \mathbf{1}_d\mathbf{1}_d^T)$ holds from Theorem 1. As a result, we obtain

$$\mathbb{E}[\Delta E] = \frac{k\sigma^2}{s}\frac{d(d+2)}{4N}, \tag{B32}$$

where we used the following relation:

$$\begin{aligned}
&\mathrm{Tr}\big[(A^TA)^{-1}\big(dI - \mathbf{1}_d\mathbf{1}_d^T\big)\big] \\
&= \frac{d(d+2)}{N}\mathrm{Tr}\big[\big(2I + \mathbf{1}_d\mathbf{1}_d^T\big)^{-1}\big(dI - \mathbf{1}_d\mathbf{1}_d^T\big)\big] \\
&= \frac{d(d+2)}{N}\mathrm{Tr}\left\{\left[\frac{1}{2}I - \frac{1}{2(d+2)}\mathbf{1}_d\mathbf{1}_d^T\right]\big(dI - \mathbf{1}_d\mathbf{1}_d^T\big)\right\} \\
&= \frac{d(d+2)}{N}\mathrm{Tr}\left[\frac{d}{2}I - \frac{1}{2}\mathbf{1}_d\mathbf{1}_d^T\right] \\
&= \frac{d(d+2)}{N}\frac{d}{2}(N_{\min} - 1) \\
&= \frac{d^2(d+2)^2(d-1)}{4N}.
\end{aligned} \tag{B33}$$

APPENDIX C: PROOF OF THEOREM 1 AND COROLLARY 1

In this section, we first present useful lemmas to prove Theorem 1 and its Corollary 1 that allow for analytical calculation of the optimal bound of the C-cost.

The first lemma is trivial from the singular-value decomposition of a matrix $A = U\Sigma V^T$, where U and V are orthogonal matrices, and Σ is the diagonal matrix that contains the singular values of A.

Lemma 1. Let A be a real matrix. The multiset of nonzero eigenvalues of AA^T is the same as the multiset of nonzero eigenvalues of A^TA.

Lemma 2. Let A be a real symmetric matrix such that one of its eigenvalues is a and the rest are b's. Then it holds that $A = (a-b)\mathbf{u}\mathbf{u}^T + bI$, where $\mathbf{u}$ is the (normalized) eigenvector corresponding to the eigenvalue a.

Proof. By seeing that $A\mathbf{u} = a\mathbf{u}$, and $A\mathbf{v} = b\mathbf{v}$ hold for every $\mathbf{v}$ which is orthogonal to $\mathbf{u}$, i.e., $\mathbf{v}^T\mathbf{u} = 0$ ■.

Let A be an $n \times n$ positive definite matrix with the largest eigenvalue $\lambda_{\max}$ and the smallest eigenvalue $\lambda_{\min}$ such that $\kappa = \lambda_{\max}/\lambda_{\min}$. It is known that $n^2/\kappa \leqslant \mathrm{Tr}(A)\,\mathrm{Tr}(A^{-1}) \leqslant n^2\kappa$ holds with equality if and only if $\kappa = 1$, i.e., $A = \lambda I$ for some $\lambda > 0$. We formalize this in the following lemma.

Lemma 3. Any positive-definite real symmetric matrix $A \in \mathbb{R}^{n\times n}$ satisfies $\mathrm{Tr}(A^{-1}) \geqslant n^2\mathrm{Tr}(A)^{-1}$ with equality if and only if $A = \lambda I$ for $\lambda > 0$.

We now prove Theorem 1 and its Corollary 1 concerning lower bounds and its equality conditions for C-cost. Here, we revisit Theorem 1 for convenience.

Theorem 1. Suppose a single-qubit gate is expressed by a parameter $\mathbf{q}$ in $\mathbb{R}^d$, where $|\mathbf{q}| = 1$. Let $\{\mathbf{q}_1, \ldots, \mathbf{q}_N\}$ be a parameter configuration, and let A be the corresponding matrix $A = [h(q_1), \ldots, h(q_N)]^T$ in $\mathbb{R}^{N\times N_{\min}}$, where $N \geqslant N_{\min} \equiv d(d+1)/2$. The C-cost $C(A)$ is defined as

$$C(A) = \frac{N}{N_{\min}d(d+2)}\mathrm{Tr}\big[(A^TA)^{-1}\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big)\big] \tag{C1}$$

and satisfies $C(A) \geqslant 1$ with equality if and only if the parameter configuration $\{\mathbf{q}_i\}$ and A satisfies

$$A^TA = \frac{N}{d(d+2)}\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big). \tag{C2}$$

Proof. Using the Woodbury matrix identity giving

$$\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big)^{-1} = \left[\frac{1}{2}I - \frac{1}{2(d+2)}\mathbf{1}_d\mathbf{1}_d^T\right], \tag{C3}$$

we obtain the lower bound of Eq. (C3) as

$$\begin{aligned}
&\mathrm{Tr}\big[(A^TA)^{-1}\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big)\big] \\
&= \mathrm{Tr}\big\{\big[\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big)^{-1}(A^TA)\big]^{-1}\big\} \\
&= \mathrm{Tr}\left\{\left[\frac{1}{2}A^TA - \frac{1}{2(d+2)}\mathbf{1}_d\mathbf{1}_d^TA^TA\right]^{-1}\right\} \\
&\geqslant N_{\min}^2\mathrm{Tr}\left[\frac{1}{2}A^TA - \frac{1}{2(d+2)}\mathbf{1}_d\mathbf{1}_d^TA^TA\right]^{-1} \\
&= \frac{N_{\min}d(d+2)}{N},
\end{aligned} \tag{C4}$$

where the inequality in the fourth line is derived by Lemma 3. To obtain the last line, we use $\mathrm{Tr}(A^TA) = \mathrm{Tr}(AA^T) = N$ and $\mathrm{Tr}(\mathbf{1}_d\mathbf{1}_d^TA^TA) = N$ as well as $N_{\min} = d(d+1)/2$. Therefore, $C(A) \geqslant 1$.

According to Lemma 3, the equality in the fourth line in Eq. (C4) is given as

$$\frac{1}{2}A^TA - \frac{1}{2(d+2)}\mathbf{1}_d\mathbf{1}_d^TA^TA = \lambda I, \tag{C5}$$

where λ is a constant. Tracing over both sides of Eq. (C5), we have

$$\lambda = \frac{N}{d(d+2)}. \tag{C6}$$

Therefore, $C(A) = 1$ if and only if

$$A^TA = \frac{N}{d(d+2)}\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big). \tag{C7}$$

■

Corollary 1. For the minimum number of parameters ($N = N_{\min}$), it holds that $C(A) \geqslant 1$ with equality if and only if the parameter configurations $\{\mathbf{q}_i\}_{i=1}^N$ satisfy

$$|\mathbf{q}_i \cdot \mathbf{q}_j| = \frac{1}{\sqrt{d+2}} \quad (\text{for all } i \neq j). \tag{C8}$$

Proof. We show Eq. (C7) is equivalent to Eq. (C8) if $N = N_{\min}$. We first show

$$\begin{aligned}
A^TA &= \frac{N_{\min}}{d(d+2)}\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big) \\
\implies \quad |\mathbf{q}_i \cdot \mathbf{q}_j| &= \frac{1}{\sqrt{(d+2)}} \quad (\text{for all } i \neq j).
\end{aligned}$$

Recall that $A = [\mathbf{h}(\mathbf{q}_1), \ldots, \mathbf{h}(\mathbf{q}_N)]^T$. If $N = N_{\min}$, both AA^T and A^TA lie in $\mathbb{R}^{N_{\min}\times N_{\min}}$, $A^TA = \sum_i^{N_{\min}}\mathbf{h}(\mathbf{q}_i)\mathbf{h}(\mathbf{q}_i)^T$, and so

$$\sum_{i=1}^{N_{\min}}\mathbf{h}(\mathbf{q}_i)\mathbf{h}(\mathbf{q}_i)^T = \frac{N_{\min}}{d(d+2)}\big(\mathbf{1}_d\mathbf{1}_d^T + 2I\big). \tag{C9}$$

Multiplying both sides by $\mathbf{1}_d$ from the right, we obtain

$$\sum_{i=1}^{N_{\min}} \mathbf{h}(\mathbf{q}_i) = \frac{N_{\min}}{d}\mathbf{1}_d. \tag{C10}$$

According to Lemma 1, A^TA and AA^T have an identical set of nonzero eigenvalues, i.e., one of the eigenvalues is $d+2$, and the rest are 2. Then using Lemma 2, AA^T can be expressed as

$$AA^T = \frac{N_{\min}}{d(d+2)}(d\mathbf{v}\mathbf{v}^T + 2I), \tag{C11}$$

where $\mathbf{v} \in \mathbb{R}^{N_{\min}}$ is a unit vector. On the other hand, the (i, j) component of AA^T has a relation:

$$(AA^T)_{ij} = \mathbf{h}(\mathbf{q}_i)^T\mathbf{h}(\mathbf{q}_j) = \frac{N_{\min}}{d(d+2)}(dv_iv_j + 2\delta_{ij}). \tag{C12}$$

Summing Eq. (C12) over j from 1 to $N_{\min}$ and using $\mathbf{h}(\mathbf{q})^T\mathbf{1}_d = 1$ and Eq. (C10), we obtain

$$v_j = \frac{1}{\sum_i^{N_{\min}} v_i}. \tag{C13}$$

Because $\mathbf{v}$ is a unit vector, $\mathbf{v} = \pm\mathbf{1}_{N_{\min}}/\sqrt{N_{\min}}$, where $\mathbf{1}_{N_{\min}} \in \mathbb{R}^{N_{\min}}$ is a vector whose elements are all 1. Therefore,

$$\mathbf{h}(\mathbf{q}_i)^T\mathbf{h}(\mathbf{q}_j) = \frac{1}{d+2}, \quad \text{for } i \neq j. \tag{C14}$$

Using the relation $\mathbf{h}(\mathbf{q}_i)^T\mathbf{h}(\mathbf{q}_j) = (\mathbf{q}_i \cdot \mathbf{q}_j)^2$, we obtain

$$|\mathbf{q}_i \cdot \mathbf{q}_{j(\neq i)}| = \frac{1}{\sqrt{d+2}}. \tag{C15}$$

Next, we prove that

$$|\mathbf{q}_i \cdot \mathbf{q}_j| = \frac{1}{\sqrt{(d+2)}} \quad (\text{for all } i \neq j)$$
$$\implies \quad A^TA = \frac{N}{d(d+2)}\left(\mathbf{1}_d\mathbf{1}_d^T + 2I\right).$$

Using the relation $\mathbf{h}(\mathbf{q}_i)^T\mathbf{h}(\mathbf{q}_j) = (\mathbf{q}_i \cdot \mathbf{q}_j)^2$ and $|\mathbf{q}_i|^2 = 1$ again, we obtain

$$\mathbf{h}(\mathbf{q}_i)^T\mathbf{h}(\mathbf{q}_j) = \frac{1}{d+2} + \frac{d+1}{d+2}\delta_{ij}. \tag{C16}$$

As $\mathbf{h}(\mathbf{q}_i)^T\mathbf{h}(\mathbf{q}_j)$ is the (i, j) component of AA^T, we can write

$$AA^T = \frac{N_{\min}}{d(d+2)}\left(\frac{d}{N_{\min}}\mathbf{1}_{N_{\min}}\mathbf{1}_{N_{\min}}^T + 2I\right), \tag{C17}$$

because $N_{\min} = d(d+1)/2$. Using Eq. (C17), Lemmas 1 and 2, we can write A^TA as

$$A^TA = \frac{N_{\min}}{d(d+2)}(d\mathbf{v}'\mathbf{v}'^T + 2I), \tag{C18}$$

where $v' = [v'_1, \ldots, v'_{N_{\min}}]^T$ is a unit vector. As $A^TA = \sum_i \mathbf{h}(\mathbf{q}_i)\mathbf{h}(\mathbf{q}_i)^T$ and $\forall i, \mathbf{h}(\mathbf{q}_i)^T\mathbf{1}_d = 1$, multiplying Eq. (C18) by $\mathbf{1}_d$ from the right side yields

$$\sum_i \mathbf{h}(\mathbf{q}_i) = \frac{N_{\min}}{d(d+2)}(d\mathbf{v}'\mathbf{v}'^T\mathbf{1}_d + 2\mathbf{1}_d). \tag{C19}$$

Summing Eq. (C16) over i and j from 1 to $N_{\min}$, we obtain

$$\sum_{i,j} \mathbf{h}(\mathbf{q}_i)^T\mathbf{h}(\mathbf{q}_j) = \frac{N_{\min}^2}{d}, \tag{C20}$$

which further yields

$$\left(\mathbf{1}_d^T\mathbf{v}'\right)^2 = d, \tag{C21}$$

by substituting Eq. (C19) into Eq. (C20) and rearranging the resultant equation. As $\|\mathbf{1}_d\| = \sqrt{d}$, Eq. (C21) means $\mathbf{v}' = \pm\mathbf{1}_d/\sqrt{d}$. Therefore, Eq. (C18) becomes

$$A^TA = \frac{N_{\min}}{d(d+2)}\left(\mathbf{1}_d\mathbf{1}_d^T + 2I\right), \tag{C22}$$

which is just the equality condition of $C(A) = 1$. ■

APPENDIX D: PROOF OF ROTATION INVARIANCE OF C-COST

Here, we prove the C-cost $C(A)$ is invariant with respect to the parameter rotations as

$$C(A) = C(A_R), \tag{D1}$$

where the subscript R stands for the rotated parameter set. Let $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N\}$ be the original parameter configuration. Then a rotation matrix $R \in SO(d)$ gives another parameter configuration $\{R\mathbf{q}_1, R\mathbf{q}_2, \ldots, R\mathbf{q}_N\}$. For convenience, we define a matrix Q as

$$Q = (\mathbf{q}_1 \otimes \mathbf{q}_1, \mathbf{q}_2 \otimes \mathbf{q}_2, \ldots, \mathbf{q}_N \otimes \mathbf{q}_N)^T. \tag{D2}$$

Likewise,

$$\begin{aligned} Q_R &= [(R\mathbf{q}_1) \otimes (R\mathbf{q}_1), \ldots, (R\mathbf{q}_N) \otimes (R\mathbf{q}_N)]^T \\ &= Q(R^T \otimes R^T). \end{aligned} \tag{D3}$$

Using Eq. (B18), Q is linked to A as

$$AL^T = Q \quad \text{and} \quad A = QL. \tag{D4}$$

which implies Q encodes the parameter configurations as well as A. Thus, the matrix A for the rotated parameter set is given as

$$A_RL^T = Q(R^T \otimes R^T) \quad \text{and} \quad A_R = Q(R^T \otimes R^T)L. \tag{D5}$$

From Eqs. (B22) and (B24), the C-cost contains the Gram matrix A^TA. For the rotated parameter set, the corresponding Gram matrix is given as

$$\begin{aligned} A_R^TA_R &= L^T(R \otimes R)Q^TQ(R^T \otimes R^T)L \\ &= L^T(R \otimes R)LA^TAL^T(R^T \otimes R^T)L \\ &= R_LA^TAR_L^T, \end{aligned} \tag{D6}$$

where we denote $R_L := L^T(R \otimes R)L$.

In fact, the first and second terms of Eq. (B13) are independently invariant for parameter rotations as follows.

For the first term $\mathbf{1}_d(A^TA)^{-1}\mathbf{1}_d^T$ [Eq. (B22)], the rotated version of the first term is expanded as

$$\begin{aligned} \mathbf{1}_d^T\left(A_R^TA_R\right)^{-1}\mathbf{1}_d &= \mathbf{1}_d^T\left(R_LA^TAR_L^T\right)^{-1}\mathbf{1}_d \\ &= \left(R_L^{-1}\mathbf{1}_d\right)^T(A^TA)^{-1}R_L^{-1}\mathbf{1}_d \\ &= \mathbf{1}_d^T(A^TA)^{-1}\mathbf{1}_d, \end{aligned} \tag{D7}$$

where we use the fact $R_L^{-1}\mathbf{1}_d = \mathbf{1}_d$, which is easily derived as

$$\begin{aligned} R_L\mathbf{1}_d &= L^T(R\otimes R)LL^T\mathrm{vec}(I) \\ &= L^T LL^T(R\otimes R)\mathrm{vec}(I) \\ &= L^T(R\otimes R)\mathrm{vec}(I) \\ &= L^T\mathrm{vec}(RIR^T) \\ &= L^T\mathrm{vec}(I) \\ &= \mathbf{1}_d, \end{aligned} \tag{D8}$$

where we employed $LL^T(R\otimes R) = (R\otimes R)LL^T$ and $L^TL = I$. This equation implies the first term is rotation invariant.

For the second term $\mathrm{Tr}[(A^TA)^{-1}]$ [Eq. (B24)], the rotated version of the second term is expanded as

$$\begin{aligned} \mathrm{Tr}\left[\left(A_R^TA_R\right)^{-1}\right] &= \mathrm{Tr}\left[\left(R_LA^TAR_L^T\right)^{-1}\right] \\ &= \mathrm{Tr}\left[\left(A^TAR_L^TR_L\right)^{-1}\right] \\ &= \mathrm{Tr}[(A^TA)^{-1}], \end{aligned} \tag{D9}$$

where for the second equality $I = R_L^TR_L$ is employed, which is derived as

$$\begin{aligned} R_L^TR_L &= L(R^T\otimes R^T)LL^T(R\otimes R)L \\ &= L^TLL^T(R\otimes R)(R^T\otimes R^T)L \\ &= L^TLL^T[(RR^T)\otimes(RR^T)]L \\ &= L^TLL^TL \\ &= I, \end{aligned} \tag{D10}$$

where we employed $LL^T(R\otimes R) = (R\otimes R)LL^T$ and $L^TL = I$. This equation implies the second term is rotation invariant. Consequently, the C-cost is rotation invariant because both terms in the C-cost are rotation invariant ($\mathbb{E}[\Delta E]$ is also rotation invariant because it is the weighted sum of these two terms). ■

APPENDIX E: COMPARISON OF OUR PARAMETERS WITH THE ORIGINAL METHODS.

We show the parameter values used as a sequential optimization in the main text as follows. The parameters are in no particular order.

1. Rx gate

The original parameter configuration for Rx gate $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ proposed in Ref. [20] is represented as

$$\begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \cos(\pi/4) & \sin(\pi/4) \\ \cos(\pi/4) & -\sin(\pi/4) \end{bmatrix}. \tag{E1}$$

The unique optimal parameter configuration for the Rx gate with minimum number parameter set $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ is analytically derived as

$$\begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \cos(\pi/3) & \sin(\pi/3) \\ \cos(\pi/3) & -\sin(\pi/3) \end{bmatrix} \tag{E2}$$

and its arbitrary rotation and (individual) reversal.

2. Fraxis gate

The original parameter configuration for the Fraxis gate $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_6$ proposed in Ref. [21] is represented as

$$\begin{bmatrix} \mathbf{n}_1^T \\ \mathbf{n}_2^T \\ \mathbf{n}_3^T \\ \mathbf{n}_4^T \\ \mathbf{n}_5^T \\ \mathbf{n}_6^T \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \tag{E3}$$

The unique (up to arbitrary rotation and individual reversal) optimal parameter configuration for the Fraxis gate with minimum number parameter set $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_6$ is analytically derived as the vertices of the icosahedron:

$$\begin{bmatrix} \mathbf{n}_1^T \\ \mathbf{n}_2^T \\ \mathbf{n}_3^T \\ \mathbf{n}_4^T \\ \mathbf{n}_5^T \\ \mathbf{n}_6^T \end{bmatrix} = \frac{1}{\sqrt{1+\varphi^2}}\begin{bmatrix} 0 & 1 & \varphi \\ 0 & 1 & -\varphi \\ 1 & \varphi & 0 \\ 1 & -\varphi & 0 \\ \varphi & 0 & 1 \\ -\varphi & 0 & 1 \end{bmatrix}, \tag{E4}$$

where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio.

3. FQS gate

The original parameter configuration for FQS $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{10}$ proposed in Ref. [23] is represented as

$$\begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \mathbf{q}_3^T \\ \mathbf{q}_4^T \\ \mathbf{q}_5^T \\ \mathbf{q}_6^T \\ \mathbf{q}_7^T \\ \mathbf{q}_8^T \\ \mathbf{q}_9^T \\ \mathbf{q}_{10}^T \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{2} & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{E5}$$

The symmetric parameter configuration for FQS gate $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{10}$ which is only used for the experimental results in the main text is represented as

$$\begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \mathbf{q}_3^T \\ \mathbf{q}_4^T \\ \mathbf{q}_5^T \\ \mathbf{q}_6^T \\ \mathbf{q}_7^T \\ \mathbf{q}_8^T \\ \mathbf{q}_9^T \\ \mathbf{q}_{10}^T \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 & \sqrt{2} \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{E6}$$

The unique optimal parameter configuration for the FQS gate with minimum number parameter set $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{10}$ is

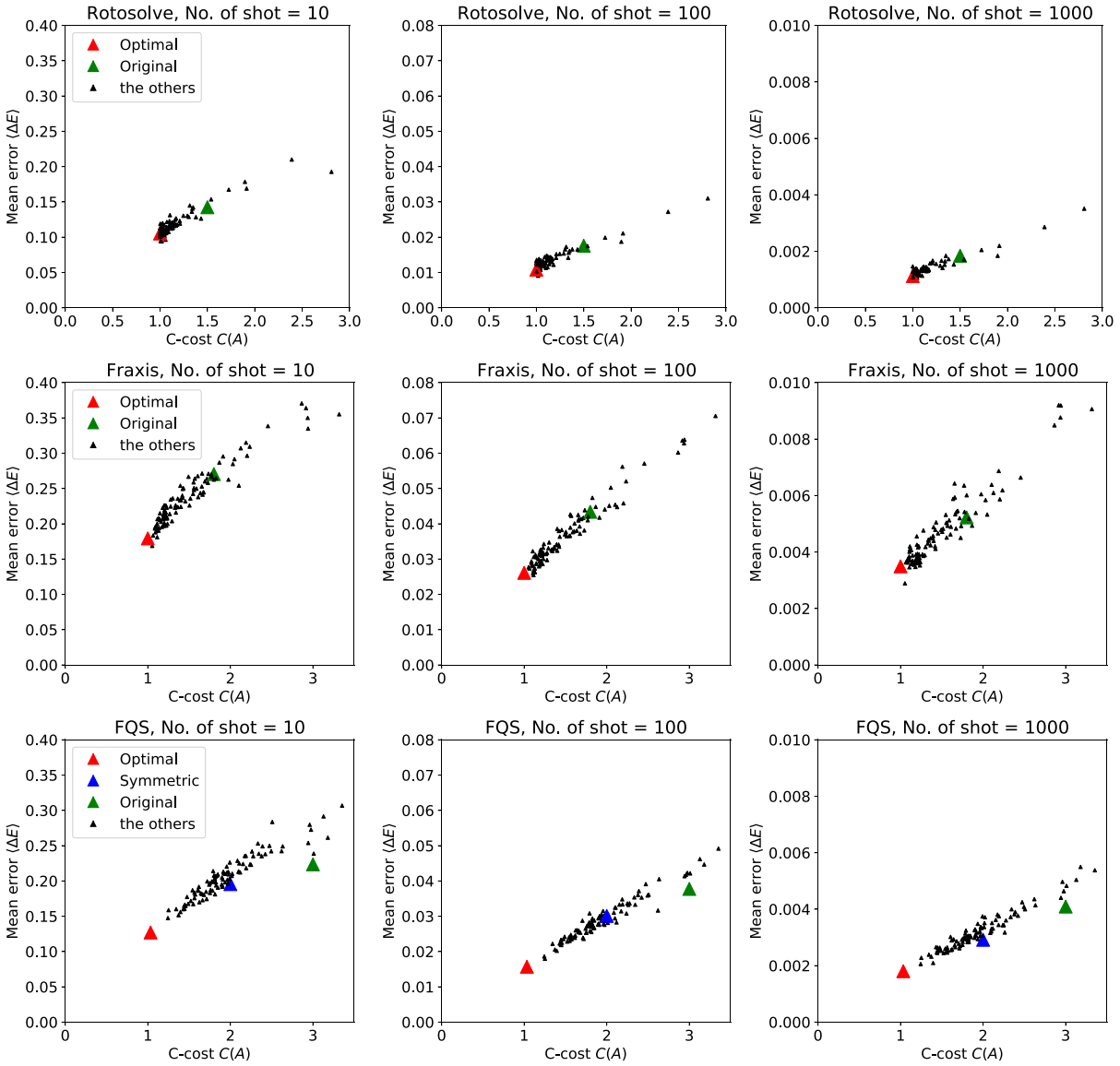

FIG. 7. Additional results of the average energy error for the one-time optimization on different target gates in the ansatz with various parameter configurations.

numerically derived as

$$\begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \mathbf{q}_3^T \\ \mathbf{q}_4^T \\ \mathbf{q}_5^T \\ \mathbf{q}_6^T \\ \mathbf{q}_7^T \\ \mathbf{q}_8^T \\ \mathbf{q}_9^T \\ \mathbf{q}_{10}^T \end{bmatrix} = \begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \\ c & c & d & d \\ c & d & c & d \\ c & d & d & c \\ d & c & c & d \\ d & c & d & c \\ d & d & c & c \end{bmatrix} \tag{E7}$$

and its arbitrary rotation and (individual) reversal, where $a = \sqrt{3}/2$, $b = -1/(2\sqrt{3})$, and $c^2 + d^2 = \frac{1}{2}$, where $c \approx 0.7049$, $d \approx -0.0561$.

APPENDIX F: ΔE DISTRIBUTIONS SAMPLED WITH VARIOUS PARAMETER CONFIGURATIONS

In the experiment in Sec. IV A, we performed optimization of only one gate to investigate the estimation error of the target gate. In the main text, we show only the case of U_2 (Ry gate in U_2 for ROTOSOLVE case) as the target gate. In this section, we show another case, that is, the case of the target gate is U_0 for the FQS and FRAXIS cases and the Ry gate of U_0 for the ROTOSOLVE case. Note that the number of shots per circuit S is set to 10, 100, and 1000. The parameters of all the gates are initialized to random values, and only the target gate is optimized. Figure 7 shows the results of all the additional experiments. The title of each subplot tells the target gate and other settings.

[1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[2] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[3] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth *et al.*, The variational quantum eigensolver: A review of methods and best practices, Phys. Rep. **986**, 1 (2022).

[4] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[5] H.-L. Huang, X.-Y. Xu, C. Guo, G. Tian, S.-J. Wei, X. Sun, W.-S. Bao, and G.-L. Long, Near-term quantum computing techniques: Variational quantum algorithms, error mitigation, circuit compilation, benchmarking and classical simulation, Sci. China-Phys. Mech. Astron. **66**, 250302 (2023).

[6] Q. Gao, G. O. Jones, M. Motta, M. Sugawara, H. C. Watanabe, T. Kobayashi, E. Watanabe, Y.-y. Ohnishi, H. Nakamura, and N. Yamamoto, Applications of quantum computing for investigations of electronic transitions in phenylsulfonyl-carbazole TADF emitters, npj Comput. Mater. **7**, 70 (2021).

[7] B. Fuller, C. Hadfield, J. R. Glick, T. Imamichi, T. Itoko, R. J. Thompson, Y. Jiao, M. M. Kagele, A. W. Blom-Schieber, R. Raymond *et al.*, Approximate solutions of combinatorial problems via quantum relaxations, arXiv:2111.03167.

[8] D. Amaro, M. Rosenkranz, N. Fitzpatrick, K. Hirano, and M. Fiorentini, A case study of variational quantum algorithms for a job shop scheduling problem, EPJ Quantum Technol. **9**, 5 (2022).

[9] C. Zoufal, R. V. Mishmash, N. Sharma, N. Kumar, A. Sheshadri, A. Deshmukh, N. Ibrahim, J. Gacon, and S. Woerner, Variational quantum algorithm for unconstrained black box binary optimization: Application to feature selection, Quantum **7**, 909 (2023).

[10] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, Variational quantum linear solver, arXiv:1909.05820.

[11] X. Xu, J. Sun, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational algorithms for linear algebra, Sci. Bull. **66**, 2181 (2021).

[12] Y. Sato, R. Kondo, S. Koide, H. Takamatsu, and N. Imoto, Variational quantum algorithm based on the minimum potential energy for solving the Poisson equation, Phys. Rev. A **104**, 052409 (2021).

[13] Y. Sato, H. C. Watanabe, R. Raymond, R. Kondo, K. Wada, K. Endo, M. Sugawara, and N. Yamamoto, Variational quantum algorithm for generalized eigenvalue problems and its application to the finite-element method, Phys. Rev. A **108**, 022429 (2023).

[14] R. Haghshenas, J. Gray, A. C. Potter, and G. K.-L. Chan, Variational power of quantum circuit tensor networks, Phys. Rev. X **12**, 011047 (2022).

[15] M. Foss-Feig, D. Hayes, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, J. M. Pino, and A. C. Potter, Holographic quantum algorithms for simulating correlated spin systems, Phys. Rev. Res. **3**, 033002 (2021).

[16] R. Haghshenas, Optimization schemes for unitary tensor-network circuit, Phys. Rev. Res. **3**, 023148 (2021).

[17] F. Barratt, J. Dborin, M. Bal, V. Stojevic, F. Pollmann, and A. G. Green, Parallel quantum simulation of large systems on small NISQ computers, npj Quantum Inf. **7**, 79 (2021).

[18] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, Variational quantum eigensolver with fewer qubits, Phys. Rev. Res. **1**, 023025 (2019).

[19] K. M. Nakanishi, K. Fujii, and S. Todo, Sequential minimal optimization for quantum-classical hybrid algorithms, Phys. Rev. Res. **2**, 043158 (2020).

[20] M. Ostaszewski, E. Grant, and M. Benedetti, Structure optimization for parameterized quantum circuits, Quantum **5**, 391 (2021).

[21] H. C. Watanabe, R. Raymond, Y.-Y. Ohnishi, E. Kaminishi, and M. Sugawara, Optimizing parameterized quantum circuits with free-axis selection, in *Proceedings of the 2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, New York, 2021), pp. 100–111.

[22] K. Wada, R. Raymond, Y.-Y. Ohnishi, E. Kaminishi, M. Sugawara, N. Yamamoto, and H. C. Watanabe, Simulating time evolution with fully optimized single-qubit gates on parametrized quantum circuits, Phys. Rev. A **105**, 062421 (2022).

[23] K. Wada, R. Raymond, Y. Sato, and H. C. Watanabe, Sequential optimal selection of a single-qubit gate and its relation to barren plateau in parameterized quantum circuits, arXiv:2209.08535.

[24] L. Slattery, B. Villalonga, and B. K. Clark, Unitary block optimization for variational quantum algorithms, Phys. Rev. Res. **4**, 023072 (2022).

[25] P. Lemmens and J. Seidel, Equiangular lines, J. Algebra **24**, 494 (1973).

[26] Y. C. R. Lin and W. H. Yu, Equiangular lines and the Lemmens-Seidel conjecture, Discrete Math. **343**, 111667 (2020).

[27] J. Haantjes, Equilateral point-sets in elliptic two-and three-dimensional spaces, Nieuw Arch. Wiskunde (2) **22**, 355 (1948).

[28] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Symmetric informationally complete quantum measurements, J. Math. Phys. **45**, 2171 (2004).

[29] A. J. Scott and M. Grassl, Symmetric informationally complete positive-operator-valued measures: A new computer study, J. Math. Phys. **51**, 042203 (2010).

[30] K. Wharton and D. Koch, Unit quaternions and the Bloch sphere, J. Phys. A: Math. Theor. **48**, 235302 (2015).

[31] R. Penrose, A generalized inverse for matrices, Math. Proc. Camb. Philos. Soc. **51**, 406 (1955).

[32] F. H. Don, The use of generalized inverses in restricted maximum likelihood, Lin. Alg. Appl. **70**, 225 (1985).

[33] T. Kato, *Perturbation Theory for Linear Operators* (Springer Science & Business Media, New York, 2013), Vol. 132.

[34] G. Greaves, J. H. Koolen, A. Munemasa, and F. Szöllősi, Equiangular lines in Euclidean spaces, J. Comb. Theory, Ser. A **138**, 208 (2016).

[35] Z. Jiang, J. Tidor, Y. Yao, S. Zhang, and Y. Zhao, Equiangular lines with a fixed angle, Ann. Math. **194**, 729 (2021).

[36] C. Godsil and G. F. Royle, *Algebraic Graph Theory*, Graduate Texts in Mathematics (Springer, New York, 2001).

[37] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, General parameter-shift rules for quantum gradients, Quantum **6**, 677 (2022).

[38] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic Hamiltonians, arXiv:1701.08213.

[39] F. Bonechi, E. Celeghini, R. Giachetti, E. Sorace, and M. Tarlini, Heisenberg XXZ model and quantum Galilei group, J. Phys. A: Math. Gen. **25**, L939 (1992).

[40] P. D. Cha and A. Shin, Perturbation methods for the eigencharacteristics of symmetric and asymmetric systems, Shock Vib. **2018**, 8609138 (2018).