Group twirling and noise tailoring for multiqubit controlled phase gates

Guoding Liu, Ziyi Xie, Zitai Xu, and Xiongfeng Ma*

Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China

(Received 19 October 2023; accepted 4 November 2024; published 2 December 2024)

Group twirling is crucial in quantum information processing, particularly in randomized benchmarking and randomized compiling. While protocols based on Pauli twirling have been effectively crafted to transform arbitrary noise channels into Pauli channels for Clifford gates—thereby facilitating efficient benchmarking and mitigating worst-case errors—the lack of practical twirling groups for multiqubit non-Clifford gates remains a challenge. To address this gap, we study the issue of finding twirling groups for generic quantum gates, focusing on a widely used circuit structure in randomized benchmarking or randomized compiling. Specifically, for multiqubit controlled phase gates, which are essential in quantum algorithms and directly implementable in practice, we determine optimal twirling groups within the realm of classically replaceable unitary operations. Contrasting with the local Pauli twirling group for Clifford gates, the optimal groups for such gates contain nonlocal operations, highlighting the overhead of tailoring noise in global non-Clifford contexts. We design new benchmarking procedures for multiqubit controlled phase gates based on the optimal twirling groups. Our simulation results show that our scheme can improve the precision and accuracy of benchmarking in small-scale systems.

DOI: 10.1103/PhysRevResearch.6.043221

I. INTRODUCTION

There has been an increased interest in quantum information processing due to its potential revolution in both science and technology. While quantum computing holds the promise of quantum advantages, its practical implementation faces significant challenges, primarily due to the inherent noise of quantum systems. When dealing with quantum noise, group-twirling-based noise tailoring is an essential step. Group twirling symmetrizes the noise channel [1,2], allowing accurate and efficient extraction of noise channel parameters. Then, it enables efficient noise characterization and subsequent quantum control optimization [3,4]. This principle underlies the randomized benchmarking (RB) methodology [5–8], which stands out as a major quantum benchmarking technique due to its low sample complexity and resilience against state preparation and measurement (SPAM) errors. Moreover, group twirling is essential in randomized compiling [9], which turns generic noise into a Pauli channel, reducing the worst-case error of quantum gates and facilitating Pauli channel learning protocols [10–13].

While randomized benchmarking and randomized compiling have seen considerable advancements, efficient noise tailoring protocols for multiqubit gates mainly focus on the Clifford case, primarily achieved through Pauli group twirling [9,14,15]. For multiqubit non-Clifford gates, researchers also made some progress and proposed benchmarking protocols for two kinds of quantum gate sets: CNOT dihedral group [16] and matchgate group [17,18]. Nonetheless, these groups are both large and highly nonlocal, posing challenges to practical implementation in expansive quantum systems. Experimental undertakings have, so far, been limited to the two-qubit domain for the CNOT dihedral group [19] and none for the matchgate group. This raises the question of whether more compact and local twirling groups might suffice for twirling and benchmarking tasks. Similar issues exist for randomized compiling. The lack of methodologies for tailoring multiqubit non-Clifford gates hinders their applications, though many of which are directly implementable or native in quantum processors [20,21].

In this work, we investigate noise tailoring strategies for generic quantum gates within RB and randomized compiling. Central to noise tailoring is the selection of appropriate twirling gates for the twirled gate, namely the gate undergoing twirling. We study this question within a frequently used circuit structure—the twirled and twirling gates are intertwined. Within this structure, we summarize the constraints between the twirled and twirling gates, and we find that any quantum gate tailoring demands a twirling gate set comparable in magnitude to the Pauli group, implying the optimality of existing noise tailoring schemes designed for Clifford gates.

In addition to the well-studied Clifford gates, our study emphasizes multiqubit controlled phase gates, given by

$$C^nZ_m = \begin{pmatrix} \mathbb{I}_{2^{n+1}-1} & \mathbf{0} \\ \mathbf{0} & e^{i\frac{2\pi}{m}} \end{pmatrix}, \tag{1}$$

where n and m are both positive integers, representing the number of control qubits and the phase, respectively. The matrix $\mathbb{I}_{2^{n+1}-1}$ denotes the identity of dimension $2^{n+1}-1$ and

*Contact author: xma@tsinghua.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

0 denotes the zero matrix. Our study can be further generalized by replacing the phase from $\frac{2\pi}{m}$ to any angle $\theta \in [0, 2\pi]$. These gates are the key components in quantum information processing, featuring in seminal algorithms such as Shor's algorithm [22], Grover's algorithm [23], linear combinations of unitary operations [24], and the preparation of hypergraph states for universal quantum computing [25]. Moreover, the C^nZ_m gate is native to superconducting [26,27], ion trap [28], and Rydberg [29] quantum systems. It is also a key ingredient to construct fault-tolerant non-Clifford gates in quantum error correction [30–32]. Thus, devising efficient noise tailoring schemes for such gates is crucial in both the theory and experiment.

Within our framework, we introduce an optimal noise tailoring scheme for the C^nZ_m gate if the twirling group falls within the class of classically replaceable unitary operations (CRU) [33] or incoherent unitary operations [34,35]. This optimality is characterized by the twirling group being the most local choice in a semiproduct structure and being the smallest one. We shall elaborate on this following Theorem 1. CRU comprises gates that can be moved after computational basis measurements, typically the Z-basis measurements, and replaced by classical postprocessing. Unlike the relatively straightforward tailoring process for Clifford gates, the optimal twirling gate set for C^nZ_m contains nonlocal quantum gates.

We further conducted numerical simulations to assess the benchmarking performance of the CS gate and C^nZ gates using various noisy twirling groups. Our findings indicate that the optimal twirling group delivers superior performance compared to the CNOT dihedral group and the Pauli group in a small-scale quantum system. We believe these results will contribute to the broader use of native non-Clifford gates in quantum computing and facilitate the practical implementation of a wide range of quantum algorithms.

The structure of this paper is organized as follows. In Sec. II, we rigorously describe the problem of finding the twirling group for a target gate and show the main results. In Sec. III, we extend the results to randomized compiling. In Sec. IV, we show simulation results of RB protocols with noisy twirling groups. Finally, we conclude in Sec. V.

II. TWIRLING GROUPS IN RANDOMIZED BENCHMARKING

We start with clarifying the algebraic relations between the twirled gate and the twirling gates in RB, following the ideas of Refs. [14,36–38]. More details about RB and the technical derivations of the results in this work are available in Appendices A and B.

The task of RB is estimating the fidelity of an individual gate, U, robust to SPAM errors. Consider a noisy quantum gate, $\widetilde{U} = U\Lambda$, where $\mathcal{U}$ is the Pauli-Liouville representation of the noiseless gate U, and $\widetilde{\cdot}$ represents the noisy version. The noise channel is denoted by Λ. The key to enabling RB is obtaining the powers of the G-twirled noise channel, Λ_G^m, where $m \in \mathbb{Z}_+$ and $\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger$. Here, G represents a twirling group that renders Λ_G diagonal in the Pauli-Liouville representation up to a unitary transformation independent of Λ. The diagonalizability ensures reliable and SPAM-error-free extraction of the diagonal elements and hence the fidelity of Λ using single-exponential fitting via character RB techniques [37]. In cases where Λ_G is not diagonal, a matrix exponential fitting challenge occurs, and accurate fidelity estimation becomes problematic.

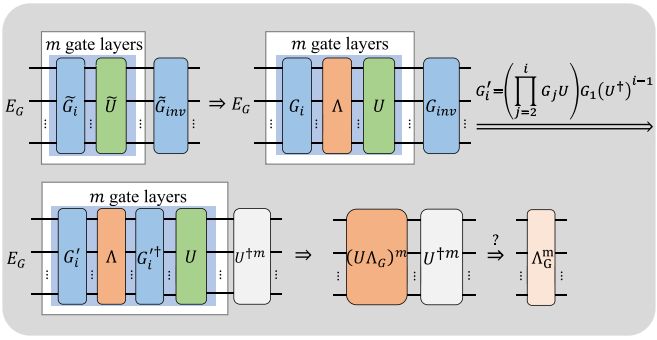


FIG. 1. Random twirling gates $G_1, G_2, \cdots, G_m$ interleaved with U in RB. The inverse gate $G_{\text{inv}} = (\prod_{i=1}^m UG_i)^\dagger$; $\widetilde{\cdot}$ represents the noisy version of a quantum gate; Λ denotes the composite noise channel of U and the twirling group, which reduces to the noise channel of U in the case of noiseless twirling group; Λ_G is the G-twirled noise channel.

To obtain Λ_G^m, we consider the circuit in Fig. 1. One independently and randomly samples m twirling gates G_i from group G and implements them interleaved with target gate U. The circuit ends with the inverse gate $G_{\text{inv}} = (\prod_{i=1}^m UG_i)^\dagger$. This circuit structure is first devised in interleaved RB [38], but here U is not necessarily in G. The target gate U is the one whose noise channel is twirled, so we call U the twirled gate. The gates from G are named the twirling gates.

Note that in practice, the interleaved circuit measures noise from both the target gate U and the twirling group G. One can set the target gate as the identity to measure the average noise of the twirling group solely. While this requires a gate-independent noise for the twirling group, a higher-order fitting [7,38] can alleviate this constraint. By comparing the fidelities measured with and without the twirled gate, one can derive an interval estimation of the fidelity of the target gate. The accuracy of this estimation improves as the fidelity of the twirling group increases [38]. Later in the simulation part, we will demonstrate the influence of the gate-dependent noise of the twirling group on the fidelity estimation of the target gate. It turns out that, in this case, the interleaved technique still works in a small-scale system and enables reliable fidelity estimation. In the following, we omit the noise of the twirling group for brevity and use the noise channel of U, Λ, to represent the composite noise channel of U and the twirling group.

Focusing on the interleaved circuit, we mathematically describe it as follows:

$$\begin{aligned}\widetilde{\mathcal{S}}_m &= \mathcal{G}_{inv}\prod_{i=1}^m \mathcal{U}\Lambda\mathcal{G}_i \\ &= \mathcal{U}^{\dagger m}\prod_{i=1}^m \mathcal{U}\mathcal{G}_i'^\dagger\Lambda\mathcal{G}_i',\end{aligned} \tag{2}$$

where $G'_i = (\prod_{j=2}^{i} G_j U) G_1 U^{\dagger i-1}$ for $1 \leqslant i \leqslant m$. To get Λ_G^m from $\widetilde{\mathcal{S}}_m$, we just need $U\mathsf{G}U^\dagger = \mathsf{G}$ as shown below.

If $U\mathsf{G}U^\dagger = \mathsf{G}$, then $\{G'_i, 1 \leqslant i \leqslant m\}$ are independent and random elements from G. After taking expectation, Eq. (2) would become $\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m$. For multiqubit controlled phase gates, we found that $U\mathsf{G}U^\dagger = \mathsf{G}$ ensures $\Lambda_G\mathcal{U} = \mathcal{U}\Lambda_G$. Then, Eq. (2) would further reduce to Λ_G^m as we want. For other quantum gates, one can at least obtain the fidelity of $(\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}$, which we proved to be a lower bound of the fidelity of Λ in Lemma 11 in the Appendix. The lower bound of the fidelity is known to be fidelity witness [39] and is also useful in quantum benchmarking. When $U\mathsf{G}U^\dagger = \mathsf{G}$, one can also choose m such that $U^m = \mathbb{I}$ to ensure the inverse gate belonging to G and implement the inverse gate without additional difficulty. Below, we summarize the requirements of the twirling group to tailor U in RB.

Question 1. Given a gate, U, find a twirling group, G such that,

$$\begin{aligned} &\text{for any quantum channel } \Lambda, \\ &\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \text{ is diagonal up to a} \\ &\text{unitary transformation independent of } \Lambda, \end{aligned} \tag{3}$$

and

$$U\mathsf{G}U^\dagger = \mathsf{G}. \tag{4}$$

Equation (3) implies that the twirling group G must make the twirled noise channel Λ_G sufficiently symmetric, irrespective of the characteristics of the original noise channel. Additionally, Eq. (4) ensures that the operation of U does not destroy this symmetry. Identifying the smallest and most local groups within Question 1 is not only pivotal for devising practical benchmarking schemes but also crucial for understanding the varying levels of difficulty in benchmarking different quantum gates. To find the optimal group for a gate, we examine the diagonalizability of Λ_G, which imposes constraints on G as follows.

Lemma 1. If a finite n-qubit unitary subgroup, G, satisfies Eq. (3), then the Pauli-Liouville representation of G is multiplicity-free. As a corollary, the cardinality of the twirling group $|\mathsf{G}| \geqslant 4^n$.

In the proof of Lemma 1, we leverage the arbitrariness of Λ to demonstrate that any trace-preserving map can be twirled into a diagonal form. From this, we prove G multiplicity-free, which means that any irreducible representation of G appears at most once in the decomposition of the Liouville representation of G. Then, with Burnside's theorem [40], which relates the group size to irreducible representations, we obtain that $|\mathsf{G}| \geqslant 4^n$. Interestingly, Lemma 1 is a converse proposition of a result in character RB [37], which asserts that a multiplicity-free group G can twirl any channel into a diagonal form.

The above lemma reveals that the twirling group for noise tailoring cannot be smaller than the (projective) Pauli group, whose cardinality achieves 4^n. Pauli gates are also local. Thus, we show the superiority of Clifford gates in noise tailoring, for which the Pauli group is a solution to Question 1. For a non-Clifford gate, this solution fails, for which we present a simple but may not be an optimal solution in Appendix B 3.

Lemma 1 shows requirements for generic unitary twirling groups. Below, in Lemma 2, we consider twirling groups belonging to CRU and show much stronger requirements other than the cardinality constraint for twirling groups. CRU comprises all gates in the product of a permutation matrix, like Pauli X and Toffoli gates, and a diagonal matrix on the computational basis, like Pauli Z gate. This set is large and becomes universal after adding Hadamard gates. Using CRU subgroups for twirling lets us replace the inverse gate with classical postprocessing when measurements are under the computational basis [33], bringing additional advantages in practical implementation. We provide more discussions about CRU twirling groups in Appendix B 4.

Lemma 2 (Informal Version). If a finite n-qubit CRU subgroup, G, satisfies Eq. (3), then G can interchange any two computational basis states. That is, for any two computational basis states $|\mathbf{i}\rangle$ and $|\mathbf{j}\rangle$ where $\mathbf{i}, \mathbf{j} \in \{0, 1\}^n$, there exists a gate $G \in \mathsf{G}$ such that $|\mathbf{j}\rangle = G|\mathbf{i}\rangle$.

To prove Lemma 2, we leverage the decomposable property of CRU into a permutation matrix and a diagonal matrix, facilitating the analysis of twirling. We also introduce Burnside's lemma [41] to transform the issue of calculating multiplicities into a problem of orbit counting. More specifically, we translate the multiplicity of the trivial representation into the orbit of G acting on computational basis states. Then, Lemma 2 follows from Lemma 1.

Note that diagonal gates, CNOT gates, and Toffoli gates do not affect $|\mathbf{0}\rangle$. Lemma 2 implies that the CRU subgroup G must include a gate set like $\mathsf{X} = \langle X\rangle^{\otimes n}$. Combining this with Eq. (4), we can deduce the necessary quantum gates that G must contain for tailoring a gate, U. Furthermore, when U is a multiqubit controlled phase gate, we derive its optimal twirling group as detailed in Theorem 1.

Theorem 1. The optimal twirling group G in CRU for the multiqubit controlled phase gate, $U = C^nZ_m$, with $n \geqslant 1$, $m \geqslant 2$, is given by

$$\begin{aligned} \mathsf{G} = \Big\{ \Pi \Big(\prod_{i=1}^{t} (\Pi_i^\dagger U \Pi_i U^\dagger)^{l_i} \Big) | \Pi \in \mathsf{X}, t \in \mathbb{N}, \forall i, \\ l_i \in \pm 1, \Pi_i \in \mathsf{X} \Big\}. \end{aligned} \tag{5}$$

Group G is the smallest group containing X and normalized by U, which we prove to satisfy Question 1. Note that any CRU subgroup is decomposable into the semiproduct of a permutation group and a diagonal group,

$$\mathsf{G}_{\mathrm{CRU}} = \Pi \ltimes \mathsf{W} = \{\Pi W | \Pi \in \Pi, W \in \mathsf{W}\}. \tag{6}$$

The optimal group can be expressed as $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ where $\mathsf{W}_X = \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \mathsf{X}\}$ only comprises diagonal gates. The optimality here means that any $\mathsf{G}_{\mathrm{CRU}}$ satisfying Question 1 implies $\mathsf{W} \supseteq \mathsf{W}_X$ and $|\Pi| \geqslant |\mathsf{X}|$. In this semiproduct structure, X and W_X are the smallest and the most local choices in the permutation part and the diagonal part, respectively. A direct corollary is that any $\mathsf{G}_{\mathrm{CRU}}$ satisfying Question 1 implies $|\mathsf{G}_{\mathrm{CRU}}| \geqslant |\mathsf{X} \ltimes \mathsf{W}_X|$. In practice, one may determine the permutation part and the diagonal part separately to sample a gate in CRU. The gate is realized by sequentially implementing these two parts. In this case, the most local

TABLE I. Scaling of the group size and computational complexity of the twirling group for tailoring C^nZ and CZ_m in this work and Ref. [16]; CXD represents the CNOT dihedral group. The term "Complexity" refers to the complexity of computing the multiplication and the inverse of group elements. The size and the complexity are expressed with the qubit number N, controlled qubit number n, and phase angle index m. Unlike Ref. [16], which specifically addresses computational complexity for $m = 2^k$, our results apply to generic positive integer m.

C^nZ			
	Group	Size	Complexity
This work	$\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle$	$O(N^n)$	$O(N^n)$
CXD [16]	$\langle CX, Z_{2^{n+1}}, X\rangle$	$O(N^{n+1})$	$O(N^{3n+1})$
CZ_m			
	Group	Size	Complexity
This work	$\langle CZ_m, Z_m, X\rangle$	$O(N^2 \log m)$	$O(N^2 \log m)$
CXD	$\langle CX, Z_{2m}, X\rangle$	$O(N^{\log m})$	$O(N^{3\log m+1})$

property in the semiproduct structure of the optimal twirling group helps its implementation.

Specific forms of twirling groups for C^nZ and CZ_m are summarized in Table I. Group G reduces to $\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle$ and $\langle CZ_m, Z_m, X\rangle$ for $U = C^nZ$ and $U = CZ_m$ with odd m, respectively. For $U = CZ_m$ with even m, $\mathsf{G} \leqslant \langle CZ_{m/2}, Z_m, X\rangle$. Within group generator $\langle\cdot\rangle$, we use X to denote $\{X_1, X_2, \cdots, X_n\}$, the Pauli X gates on all qubits, and similarly omit subscripts associated with the acting qubits when referring to $C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z$ and CZ_m, Z_m.

Note that enabling noise tailoring requires sampling from the twirling group and performing group element multiplication. These tasks introduce considerations for sample complexity, directly related to group size and computational complexity, contingent on the algorithm used for group multiplication. Table I provides complexity results for C^nZ and CZ_m and shows that our method surpasses previous results in Ref. [16]. Nonetheless, even with the optimal approach, the twirling group becomes large and highly nonlocal as the controlled qubit number increases. This leads to unfavorable sample and computational complexity for large quantum gates, highlighting an inherent challenge in benchmarking multiqubit non-Clifford gates.

III. RANDOMIZED COMPILING FOR MULTIQUBIT NON-CLIFFORD GATES

In this part, we discuss the application of previous results in randomized compiling. The task of randomized compiling is turning the noisy quantum gate $\widetilde{\mathcal{U}} = \mathcal{U}\Lambda$ to $\mathcal{U}\Lambda_G$, where the twirled noise channel Λ_G is a Pauli channel. More specifically, suppose that we implement a quantum circuit, $\widetilde{C} = \cdots\widetilde{\mathcal{U}_2}\widetilde{\mathcal{U}_1} = \cdots\mathcal{U}_2\Lambda_2\mathcal{U}_1\Lambda_1$, as shown in Fig. 2(a). The aim is tailoring $\widetilde{C}$ into $\cdots\mathcal{U}_2\Lambda_{2G}\mathcal{U}_1\Lambda_{1G}$ as shown in Fig. 2(c). The quantum gates $U_i, i \in \mathbb{Z}_+$ are tailored gates. To tailor them, we add twirling gates G_i and $G_i' = U_iG_i^\dagger U_i^\dagger$ beside the tailored gates as shown in Fig. 2(b), where G_i is randomly sampled from a twirling group, G. In reality, the gates G_i and G_{i-1}' are merged and implemented together to reduce quantum resource consumption. Thus, the twirling gates in general belong to $\mathsf{V} = \bigcup_{U\in\mathsf{U}} \mathsf{G}U\mathsf{G}U^\dagger$.

Similar to the discussion in RB, we investigate the requirements of the twirling gates for tailoring a gate set $\mathsf{U} = \{U_1, U_2, \cdots\}$. More specifically, we should optimize the choice of G to make the twirling gate set $\mathsf{V} = \bigcup_{U\in\mathsf{U}} \mathsf{G}U\mathsf{G}U^\dagger$ smaller and more local in practice. We summarize the question below.

Question 2. Given a gate set, $\mathsf{U} = \{U_1, U_2, \cdots\}$, find a twirling gate set, V, such that $\mathsf{V} = \bigcup_{U\in\mathsf{U}} \mathsf{G}U\mathsf{G}U^\dagger$ and the twirling group G satisfies

$$\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \text{ is a Pauli channel.} \tag{7}$$

When the tailored gate set U only comprises Clifford gates and single-qubit T gates, the twirling gate set suffices to be chosen as the dihedral group $\langle X, S\rangle$ [9]. Here, we consider that U is composed of a multiqubit controlled phase gate $U = C^nZ_m$. Question 2 reduces to finding gate set $\mathsf{G}U\mathsf{G}U^\dagger$ while Λ_G is a Pauli channel. When considering G as a CRU subgroup, $\mathsf{G}U\mathsf{G}U^\dagger$ can be written in a form,

$$\mathsf{V} = \mathsf{G}\mathsf{W} = \{GW | G \in \mathsf{G}, W \in \mathsf{W}\}, \tag{8}$$

where $\mathsf{W} = \{G^\dagger UGU^\dagger, G \in \mathsf{G}\}$ is a set of diagonal gates. Similar to the results in RB, for $U = C^nZ_m$, $\mathsf{W} \supseteq \mathsf{W}_X = \{\Pi^\dagger U\Pi U^\dagger, \Pi \in \mathsf{X}\}$ based on Lemma 2, proved in the proof of Theorem 1. Meanwhile, based on Lemma 1, G cannot be

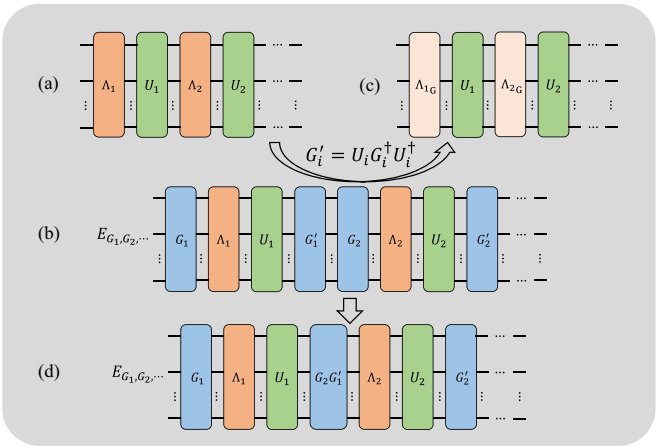


FIG. 2. Randomized compiling for U_1, U_2, $\cdots$, by inserting twirling gates G_i and $G_i' = U_iG_i^\dagger U_i^\dagger$ beside U_i with $i \in \mathbb{Z}_+$; Λ_{iG} is the G-twirled noise channel required to be a Pauli channel. In reality, G_i and G_{i-1}' are merged and implemented together as quantum gate G_iG_{i-1}'.

smaller and more local than the Pauli group P_n. However, by choosing G as Pauli group P_n, we would obtain a solution $\mathsf{V} = \mathsf{P}_n U \mathsf{P}_n U^\dagger = \mathsf{P}_n \mathsf{W}_X$ to Question 2. This is the optimal solution for tailoring a multiqubit controlled phase gate with G restricted in CRU if one separately implements the permutation matrix and the diagonal matrix for a CRU gate.

Note that the twirling gate set $\mathsf{P}_n \mathsf{W}_X$ is simpler than the twirling group $\mathsf{X} \ltimes \mathsf{W}_X$ for $C^n Z_m$. Generally speaking, the requirements for the twirling gates in randomized compiling are much simpler than those in RB. A direct observation to Question 2 is that as long as $U \mathsf{P}_n U^\dagger$ only comprises local gates, U can be tailored with only local twirling gates in randomized compiling. We leave the problem of tailoring more kinds of quantum gates in the setting of randomized compiling to the future.

IV. SIMULATION OF BENCHMARKING MULTIQUBIT CONTROLLED PHASE GATES

Below, we present the numerical benchmarking results for CS and $C^n Z$ gates with noisy twirling groups. The CS gate involves 2 qubits, and the $C^n Z$ gate involves $n+1$ qubits in the simulation. We simulate practical benchmarking experiments by considering gate-dependent noise in the twirling group. Here, both twirling gates and the target gate are CRU, which enables efficient gate sampling and simulation by separately determining their permutation and diagonal parts:

(1) The permutation part is generated by Pauli X and CNOT gates. We consider unitary errors $e^{i\delta X}$ and controlled-$e^{i\delta X}$ for the two gates, respectively, where δ is small.

(2) The diagonal part is generated by multiqubit controlled phase gates and single-qubit Z rotation. The unitary error of the former is considered to be a diagonal gate where the diagonal elements depend on the control qubit number. The unitary error of the latter is a small Z rotation.

We also consider local dephasing and local amplitude-damping noise, where the noise strength depends on the time to implement the gate. The SPAM is also simulated with flip errors. More details about the noise model are shown in Appendix C 1.

Notice that one needs first to obtain the fidelity of the composite noise channel, denoted as F_1, via the interleaved circuit and the fidelity of the twirling group, denoted as F_2, via the reference circuit. The fidelity of the target gate is evaluated by

$$F = \frac{d^2 F_1 - 1}{d^2 F_2 - 1}\left(1 - \frac{1}{d^2}\right) + \frac{1}{d^2}. \tag{9}$$

The derivation is given in Appendix A 5. In this section, we only present the fidelity of the target gate, while the other two kinds of fidelities are available in Appendix C 3. We also simulate the case of the noiseless twirling group, and the results are shown in Appendix C 5.

Depending on the SPAM settings, we simulate two different benchmarking procedures and show the results in Secs. IV A and IV B, respectively. We compare our scheme with the Pauli group and the CNOT dihedral group. Note that the Pauli group does not satisfy the requirements in Question 1 to benchmark multiqubit controlled phase gates. In this case, the inverse gate in Fig. 1 has to be outside the Pauli group to make the ideal circuit equal to the identity. In general, this inverse gate can take any element within the optimal group given by Eq. (5), which may be nonlocal.

A. Benchmarking with random SPAM settings

Recall that the gate fidelity is estimated by evaluating different diagonal terms of twirled noise channels in the Pauli-Liouville representation. The number of different diagonal terms of the twirled noise channel depends on the twirling group. This number increases exponentially with respect to the qubit number for the Pauli group and the optimal group and is 2 for the CNOT dihedral group. In the simulation, we sample and evaluate 20 different diagonal terms instead of all of them for the former two groups. For the CNOT dihedral group, we only evaluate 2 different diagonal terms.

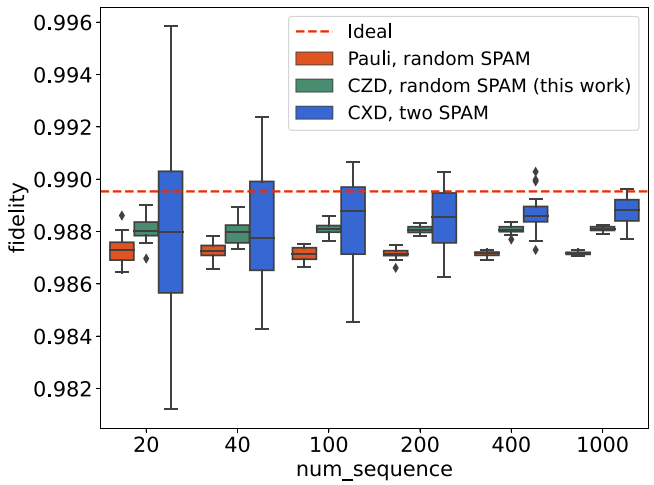
To evaluate a diagonal term, $\mathrm{tr}(P\Lambda(P))/2^N$, where Λ is the noise channel, N is the qubit number, and P is a nonidentity Pauli observable, we simulate the circuit in Fig. 1 with the eigenstate of P as input and P as measurement observable. By simulating randomly sampled gate sequences and averaging the results over different sequences, we estimate $\mathrm{tr}(\widetilde{P}\Lambda_{\mathsf{G}}^m(\widetilde{|\psi\rangle\langle\psi|}))$ where Λ_{G} is the twirled noise channel and $P|\psi\rangle = |\psi\rangle$. The circuit depth m is taken from $\{2, 4, 6, 8, 10\}$, and the number of different gate sequences for each circuit depth varies. The total number of gate sequences for different groups is the same to make a fair comparison. By fitting $\mathrm{tr}(\widetilde{M}\Lambda_{\mathsf{G}}^m(\widetilde{\rho}))$ to $A\lambda^m$ one can get λ and estimate $\mathrm{tr}(P\Lambda(P))/2^N$.

In Fig. 3, we present the simulation results for the CS and CCZ gates. The results demonstrate that the fluctuation of the benchmarking results for the optimal group is as small as the Pauli group and is much smaller than the CNOT dihedral group for benchmarking two gates. The large fluctuation of the CNOT dihedral group mainly results from a more severe gate-dependent noise.

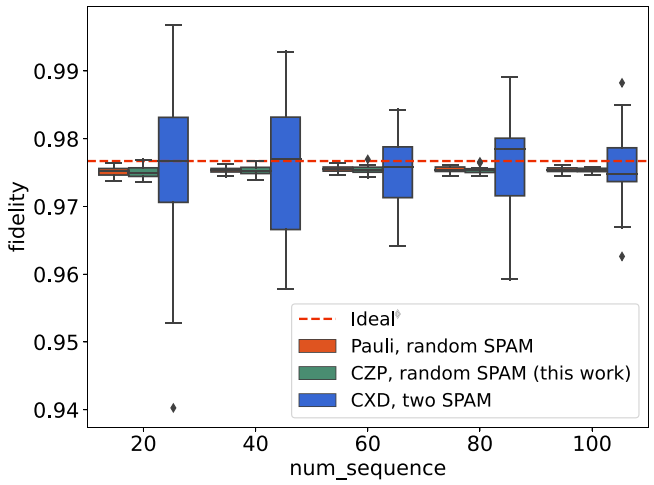
Concerning bias in benchmarking results, the performance of the optimal group lies between that of the CNOT dihedral group and Pauli group when benchmarking CS gate. In contrast, the performance is similar across all three groups when benchmarking CCZ gates. The bias observed in the Pauli group arises from a less effective twirling action compared to the other two groups, leading to matrix decay rather than a single exponential decay. This introduces bias in fidelity estimation. Furthermore, the sampling of diagonal terms introduces bias to the results of the Pauli group and the optimal group. Additionally, gate-dependent noise affects the accuracy of these results. The CNOT dihedral group is the most affected, followed by the optimal group, and least impacted is the Pauli group.

B. Benchmarking with the ZX-SPAM setting

We notice that when the twirling group contains the CZ dihedral group, $\langle CZ, Z, S\rangle$, another SPAM setting can be employed to enhance the performance, which we name ZX-SPAM. Instead of extracting the diagonal term $\mathrm{tr}(P\Lambda(P))/2^N$ once at a time, by preparing $|0\rangle^{\otimes N}$ and measuring in $Z^{\otimes N}$, one can extracting $\mathrm{tr}(P_Z\Lambda(P_Z))/2^N$ for all $P_Z \in \{\mathbb{I}, Z\}^{\otimes N}$. Similarly, by preparing $|+\rangle^{\otimes N}$ and measuring in $X^{\otimes N}$, one

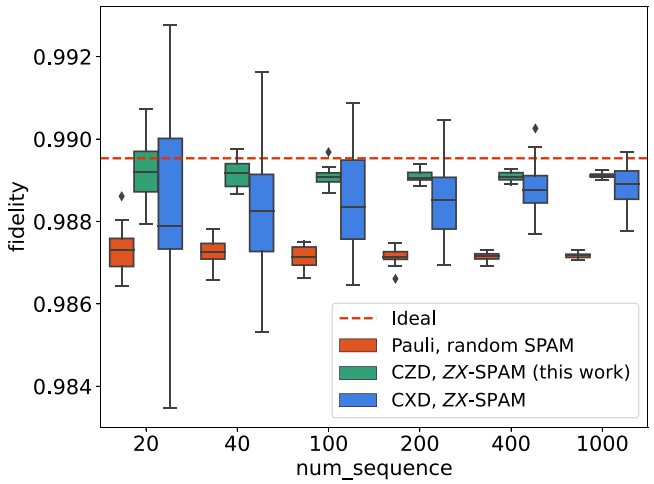


(a) CS

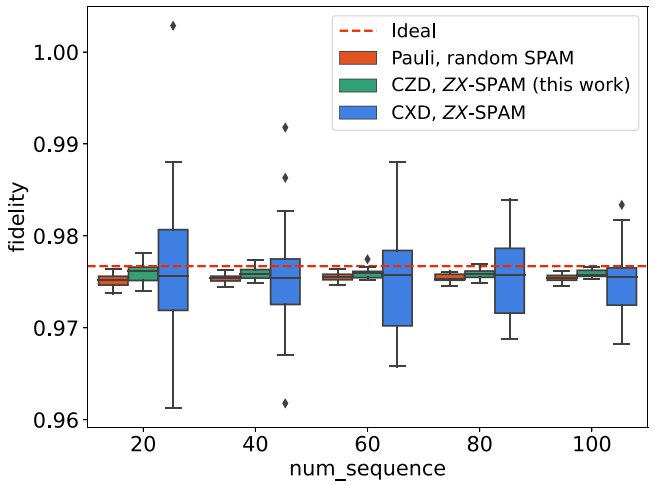


(b) CCZ

FIG. 3. Benchmarking results for the CS and CCZ gates in panels (a) and (b), respectively, with the optimal twirling group, the Pauli group, and the CNOT dihedral group. The optimal twirling group is the CZ dihedral group $\langle CZ, Z, S\rangle$ for CS, and the CZ Pauli group $\langle CZ, Z, P\rangle$ for CCZ. The red dashed line is the theoretical value of the noise channel fidelity. Each box plot contains 20 fidelities, and each fidelity is estimated with circuit depths $\{2, 4, 6, 8, 10\}$, and the total number of different gate sequences for each depth is specified by the horizontal axis. Here, for the Pauli group and the optimal group, we randomly sample and estimate 20 different diagonal terms of the twirled noise channel. We mark this setting with "random SPAM" on the label. For the CNOT dihedral group, we estimate only two different diagonal terms of the twirled noise channel. We mark this setting with "two SPAM" on the label. Each SPAM setting prepares an eigenstate of a Pauli observable with eigenvalue 1 and measures this Pauli observable. In "two SPAM," the two Pauli observables are chosen as $Z^{\otimes N}$ and $X^{\otimes N}$.



(a) CS



(b) CCZ

FIG. 4. Benchmarking results for the CS and CCZ gates in panels (a) and (b), respectively, with the Pauli group, the CZ dihedral group $\mathsf{G}^Z = \langle CZ, X, S\rangle$, and the CNOT dihedral group. The group G^Z is denoted as CZD in the label. The red dashed line is the theoretical fidelity value. Each box plot contains 20 fidelities. The setting of circuit depths and sampling is the same as in Fig. 3. Nonetheless, for G^Z and the CNOT dihedral group, the SPAM setting and the postprocessing differ from that in Fig. 3, which is preparing $|0\rangle^{\otimes N}$ and measuring in Z basis and preparing $|+\rangle^{\otimes N}$ and measuring in X basis to get all different diagonal terms of the twirled noise channel. We mark this setting with "ZX-SPAM" on the label.

can extract $\mathrm{tr}(P_X \Lambda(P_X))/2^N$ for all $P_X \in \{\mathbb{I}, X\}^{\otimes N}$. These two SPAM settings are sufficient to obtain all different diagonal terms as long as the twirling group includes the CZ dihedral group. Compared to extracting the diagonal term once at a time, the procedure with ZX-SPAM estimates the fidelity more precisely and accurately, thanks to more efficient extraction of diagonal terms and the absence of sampling errors. More details about this simulation procedure are available in Appendix C 2.

We simulate this enhanced benchmarking procedure and present the results in Fig. 4. For C^nZ gates, we use group $\mathsf{G}^Z = \langle C^{n-1}Z, \cdots, CZ, X, S\rangle$ instead of the optimal one $\langle C^{n-1}Z, \cdots, CZ, X, Z\rangle$ for twirling to make the twirling group contain the CZ dihedral group. A little overhead of adding phase gate S simplifies the noise channel more and helps to improve the benchmarking performance. For the CS

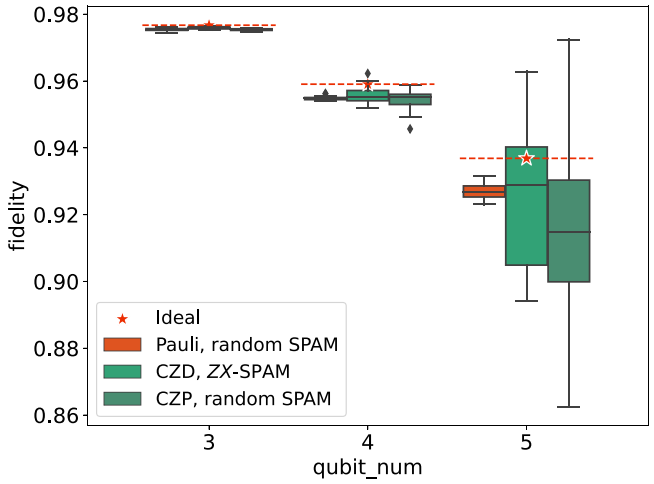


FIG. 5. Benchmarking results for C^nZ gates with the Pauli group, $\mathsf{G}^Z = \langle C^{n-1}Z, \cdots, CZ, X, S\rangle$, and the optimal group $\langle C^{n-1}Z, \cdots, CZ, X, Z\rangle$. The three methods are labeled with "Pauli, random SPAM," "CZD, ZX-SPAM," and "CZP, random SPAM," respectively. The benchmarking setting is the same in Figs. 3 and 4. The red five-pointed star with a line denotes the ideal fidelity value. Each box plot contains 20 fidelities.

gate, the optimal group is the CZ dihedral group. To make a fair comparison, in this simulation, we adopt the ZX-SPAM setting for the CNOT dihedral group as well.

Figure 4 shows improved benchmarking results for G^Z and the CNOT dihedral group with the enhanced benchmarking procedure. Among all benchmarking methods, utilizing G^Z with the ZX-SPAM enjoys the best precision and accuracy for benchmarking CS and CCZ gates, providing a practical scheme for tailoring multiqubit controlled phase gates in small-scale systems.

When the system size increases, the CNOT dihedral group becomes ineffective and fails to provide accurate fidelity estimates due to severe gate-dependent noise within the twirling group. In Fig. 5, we present benchmarking results for C^nZ gates with up to 5 qubits, comparing the Pauli group, G^Z, and the optimal group. With larger gate-dependent noise, results for the latter two exhibit significant fluctuations, especially for systems with 5 qubits. In practice, this fluctuation can be mitigated by increasing the number of sampled sequences, as illustrated in Figs. 3 and 4, with further simulation results provided in Appendix C. Notably, the results from G^Z using the ZX-SPAM remain closest to the ideal values, a consistency observed for the 6-qubit case, detailed in Appendix C 4. These simulations highlight the crucial role of identifying small and local twirling groups that possess sufficient twirling capabilities. Moreover, all results are applicable to Toffoli gates, controlled-$\sqrt{Y}$ gates, and various other nondiagonal gates through the use of local gauge transformations [15].

V. OUTLOOK

In this work, we study noise tailoring from a distinct perspective of identifying suitable twirling groups and show nearly optimal results for multiqubit controlled phase gates C^nZ_m. Besides benchmarking gate fidelity, the tailoring scheme can help extract more information, such as Pauli eigenvalues, and study the learnability of Pauli noise [42]. The results also benefit fault-tolerant protocols and quantum algorithms utilizing multiqubit controlled phase gates.

In the future, it is crucial to explore optimal twirling groups for practical quantum gates and assess the feasibility of using only local twirling gates for efficient noise tailoring. The gate-dependent noise and the unfavorable size and computational complexity of the optimal twirling groups for large-scale non-Clifford gate benchmarking pose a significant challenge. Potential solutions include exploring circuit structures beyond those shown in Fig. 1 and seeking benchmarking methodologies that do not rely on a group structure [36,43–45]. Based on our simulation results, a twirling gate set between the Pauli group and the optimal group may be an intermediate option to twirl multiqubit controlled phase gates, like the set of gates implemented within a short depth in the optimal group.

Beyond noise tailoring, the insights gained on CRU subgroups in this study can further enhance random matrix protocols involving them, such as classical shadow [46] and gate-set shadow [47,48], and contribute to the exploration of the classical simulation capabilities of various groups.

ACKNOWLEDGMENTS

We thank Boyang Chen, Zhuo Chen, Daojin Fan, and Shaowei Li for the helpful discussions. This work was supported by the National Natural Science Foundation of China Grants No. 12174216 and the Innovation Program for Quantum Science and Technology Grants No. 2021ZD0300804 and No. 2021ZD0300702.

APPENDIX

Here, we briefly introduce the content of the Appendices. In Appendix A, we introduce the notations used in this work and preliminaries about quantum channels, group theory, randomized benchmarking, and classically replaceable unitary operations. In Appendix B, we show the proof of the main results in this work, including restating the requirements of the twirling groups in randomized benchmarking, proof of main lemmas and theorems in the main text, and the structure analysis of the optimal twirling groups for multiqubit controlled phase gates. In Appendix C, we show the simulation details, including the noise model, the benchmarking protocol, and additional benchmarking results.

APPENDIX A: PRELIMINARY

In this part, we introduce some related works and basic mathematical tools. Below, we first introduce basic notations. The Hilbert space for n qubits is denoted as $\mathcal{H}$ and the set of linear operators on $\mathcal{H}$ is denoted as $\mathcal{L}(\mathcal{H})$. Same in the main text, we denote a quantum gate, or a unitary transformation acting on $\mathcal{H}$, in the standard representation with a capital letter, like U and G. A set composed of quantum gates is denoted with sans serif fonts like S. A quantum gate set satisfying group condition under the gate composition is always denoted as G. Note that, in this work, we distinguish two concepts: the twirling gate set and the twirling group. The

TABLE II. Notation.

Quantity	Notation
n	Number of controlled qubit in C^nZ_m or qubit number
m	Phase of controlled qubit in C^nZ_m or circuit depth
N	Qubit number
$\mathcal{H}$	Hilbert space
$\mathcal{L}(\mathcal{H})$	Set of linear operators on $\mathcal{H}$
U	Target gate
V	Twirling gate
G	Twirling gate in group G
$\mathcal{U}$	Pauli Liouville representation of U
Λ	Noise channel
Λ_G	G-twirled noise channel
$\widetilde{U}$	Noisy version of U
V	Twirling gate set
G	Twirling gate group
$C^nZ_m = \begin{pmatrix} \mathbb{I}_{2^{n+1}-1} & \mathbf{0} \\ \mathbf{0} & e^{i\frac{2\pi}{m}} \end{pmatrix}$	Multiqubit controlled phase gate
X_j	Pauli X on qubit j
Y_j	Pauli Y on qubit j
Z_j	Pauli Z on qubit j
$\mathbb{I}$	Identity operator
I	Identity channel
$\langle G_1, G_2, ...\rangle$	Group generated by $G_1, G_2, ...$
P_n	n-qubit Pauli group $\langle X, Z\rangle^{\otimes n}$
D_n^m	n-qubit local dihedral group $\langle X, Z_m\rangle^{\otimes n}$
χ	Character function of a group
Π	Projector or permutation matrix

twirling group is a twirling gate set with a group structure. We normally use V and G to represent the twirling gate set and the twirling group, respectively. The Liouville representation of U is denoted as its calligraphic letter, $\mathcal{U}$. In this paper, the Liouville representation refers to the Pauli-Liouville representation, which will be reviewed in detail in the following subsection. Quantum channels are defined as completely positive and trace-preserving (CPTP) linear maps on $\mathcal{L}(\mathcal{H})$. The noise channel of U is normally denoted as Λ, and we use the same expression for the map representation and the Liouville representation for Λ. We use a wavy line to represent the noisy version of U like $\widetilde{U}$ and $\widetilde{\mathcal{U}}$. Note that a quantum gate is always a quantum channel, but a quantum channel is generally not a quantum gate. We summarize the frequently used notations in Table II. Note that we sometimes reuse part of notations, and the meanings of these notations are relevant to the context.

1. Pauli-Liouville representation

Here, we introduce the Pauli-Liouville representation of quantum channels. Normally, a quantum channel, Λ, is defined within its Kraus representation. For any $O \in \mathcal{L}(\mathcal{H})$, the action of channel Λ is defined with

$$\Lambda(O) = \sum_{l=1}^{k} K_l O K_l^\dagger, \tag{A1}$$

where $k \in \mathbb{Z}_+$ and $\{K_l, 1 \leqslant l \leqslant k\}$ is the set of Kraus operators satisfying the condition

$$\sum_{l=1}^{k} K_l^\dagger K_l = \mathbb{I}, \tag{A2}$$

where $\mathbb{I}$ represents the identity operator.

The Kraus representation is not a matrix representation and, hence, not convenient for our work. For further elaboration, we introduce the Liouville representation, which is defined on a set of trace-orthonormal basis elements in $\mathcal{L}(\mathcal{H})$. In this work, we choose the basis to be the normalized and projective Pauli group. In this case, the representation is named Pauli-Liouville representation. The n-qubit Pauli group, denoted as P_n, is given by

$$\begin{aligned} \mathsf{P}_n &= \bigotimes_{j=1}^{n} \langle X, Z\rangle \\ &= \langle X_1, Z_1, X_2, Z_2, \cdots, X_n, Z_n\rangle \\ &= \{\pm 1, \pm i\} \times \left\{ \bigotimes_{j=1}^{n} P_j | P_j \in \{\mathbb{I}, X, Y, Z\} \right\}, \end{aligned} \tag{A3}$$

where $\mathbb{I} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is the identity operator with dimension 2. X, Y, and Z are the single-qubit Pauli matrices given by

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A4}$$

X_j is the single-qubit Pauli X matrix acting on jth qubit given by $(\bigotimes_{i=1}^{j-1}\mathbb{I}_i)\otimes X(\bigotimes_{i=j+1}^{n}\mathbb{I}_i)$. The notation $\langle\cdot\rangle$ denotes the group generated by $\cdot$. The elements in $\langle\rangle$ are called group generators. For instance, $\langle X_j, Z_j\rangle = \{X_j^{m_1}Z_j^{n_1}X_j^{m_2}Z_j^{n_2}\cdots|\forall i, m_i, n_i\in\mathbb{Z}\} = \{\pm1, \pm i\}\times\{\mathbb{I}_j, X_j, Y_j, Z_j\}$. In the main text and below, for brevity we sometimes use notation $\langle X, Z\rangle^{\otimes n}$ or even $\langle X, Z\rangle$ to represent $\langle X_1, Z_1, X_2, Z_2, \cdots, X_n, Z_n\rangle$ where in this case, X and Z denote Pauli X and Pauli Z gates on all qubits, respectively. We would always omit subscripts, representing which qubits the gates act on, in $\langle\rangle$ when the group generators contain the same gates acting on all qubits.

In literature, people always use another definition for the Pauli group, which is, in fact, the projective Pauli group. Projective Pauli group is the quotient of the Pauli group by its center $\{\pm1, \pm i\}$, $\langle X_1, Z_1, X_2, Z_2, \cdots, X_n, Z_n\rangle/\{\pm1, \pm i\}$. Thus, we normally use $\bigotimes_{j=1}^{n}\{\mathbb{I}, X, Y, Z\}$ to represent it. As in quantum systems, the overall phase is not important. The quotient by the phase $\{\pm1, \pm i\}$ does not influence the quantum state. Below we simply use the Pauli group to denote the projective Pauli group and use the following definition:

$$\mathsf{P}_n = \bigotimes_{j=1}^{n}\{\mathbb{I}, X, Y, Z\}. \tag{A5}$$

The normalized Pauli group is obtained by normalizing the elements of the n-qubit Pauli group by a factor of $1/\sqrt{2^n}$ as follows:

$$\mathsf{P}'_n = \left\{\sigma_i = \frac{1}{\sqrt{2^n}}P_i | P_i\in\mathsf{P}_n\right\}. \tag{A6}$$

These normalized Pauli operators are complete and satisfy the orthonormality under the Hilbert-Schmidt inner product,

$$\mathrm{tr}(\sigma_i^\dagger\sigma_j) = \delta_{ij}, \tag{A7}$$

where δ_{ij} is the Kronecker δ symbol. Thus, we can represent the quantum state and the quantum channel with normalized Pauli operators as a set of bases.

Any n-qubit operator O in $\mathcal{L}(\mathcal{H})$ can be decomposed with the 4^n normalized Pauli operators,

$$O = \sum_{\sigma_i\in\mathsf{P}'_n}\mathrm{tr}(\sigma_i^\dagger O)\sigma_i. \tag{A8}$$

Thus, we can use the expansion coefficients $\mathrm{tr}(\sigma_i^\dagger O)$ to represent O and in Pauli-Liouville representation, we express Eq. (A8) as

$$|O\rangle\!\rangle = \sum_{\sigma_i\in\mathsf{P}'_n}\mathrm{tr}(\sigma_i^\dagger O)|\sigma_i\rangle\!\rangle, \tag{A9}$$

where $|\sigma_i\rangle\!\rangle$ in Pauli-Liouville representation is a length-4^n vector with only one nonzero element 1.

Furthermore, any quantum channel Λ can be represented as a matrix in the Liouville representation. The action of the channel on an operator O is given by

$$|\Lambda(O)\rangle\!\rangle = \Lambda|O\rangle\!\rangle, \tag{A10}$$

and the elements of the Liouville representation of Λ are given by

$$\Lambda_{ij} = \langle\!\langle\sigma_i|\Lambda|\sigma_j\rangle\!\rangle = \mathrm{tr}(\sigma_i\Lambda(\sigma_j)). \tag{A11}$$

Specifically, the diagonal terms of Pauli-Liouville representation are named Pauli fidelity [15], defined as follows:

$$\lambda_i = \langle\!\langle\sigma_i|\Lambda|\sigma_i\rangle\!\rangle = \frac{1}{d}\mathrm{tr}(P_i\Lambda(P_i)), \tag{A12}$$

where P_i is a Pauli operator, and d is the dimension of the Hilbert space.

In the Liouville representation, the concatenation of two channels can be represented as the product of their corresponding matrices,

$$|\Lambda_2\circ\Lambda_1(O)\rangle\!\rangle = \Lambda_2|\Lambda_1(O)\rangle\!\rangle = \Lambda_2\Lambda_1|O\rangle\!\rangle. \tag{A13}$$

The Liouville representation also allows us to vectorize the measurement operators using the Liouville bra-notation. The measurement probability of a state ρ using a positive operator-valued measure (POVM) F_i is given by

$$p_i = \langle\!\langle F_i|\rho\rangle\!\rangle = \mathrm{tr}(F_i^\dagger\rho), \tag{A14}$$

where $\langle\!\langle F_i|$ is the Liouville bra-notation of F_i and is equal to the conjugate transpose of $|F_i\rangle\!\rangle$.

We also briefly introduce another representation called the χ-matrix representation for quantum channel Λ. Given an n-qubit state, ρ, $\Lambda(\rho)$ can be expanded as

$$\Lambda(\rho) = d\sum_{i,j}\chi_{ij}\sigma_i\rho\sigma_j^\dagger, \tag{A15}$$

where the process matrix χ is uniquely determined by the orthonormal Pauli operator basis σ_j and $d = 2^n$ represents the dimension of the quantum system. Note that $\sigma_0 = \frac{\mathbb{I}_d}{\sqrt{d}}$. If a channel is diagonal in this representation, then it is called a Pauli channel. It is easy to verify that a Pauli channel is also diagonal in the Pauli-Liouville representation. The diagonal terms of Pauli-Liouville representation and that of χ-matrix representation are related by Walsh-Hadamard transformation as shown in the following equation:

$$\lambda_j = \sum_i(-1)^{\langle i,j\rangle}\chi_{ii}, \tag{A16}$$

where $\langle i, j\rangle = 0$ if P_i commutes with P_j, and $\langle i, j\rangle = 1$ otherwise. The inverse Walsh-Hadamard transformation is given by

$$\chi_{jj} = \frac{1}{d^2}\sum_i(-1)^{\langle i,j\rangle}\lambda_i. \tag{A17}$$

2. Quantum channel fidelity

To facilitate the understanding of RB, we introduce the concepts of process fidelity and average fidelity. These two fidelities are equivalent to each other by linear transformation. The task of RB is estimating the fidelity of a given quantum gate or a given gate set.

The process fidelity of a channel, Λ, can be defined with its χ-matrix representation as follows:

$$F(\Lambda) = \chi_{00}(\Lambda). \tag{A18}$$

The process fidelity can be obtained from Eq. (A17) as

$$F(\Lambda) = \chi_{00}(\Lambda) = \frac{1}{d^2}\sum_j\lambda_j = \frac{1}{d^2}\mathrm{tr}(\Lambda), \tag{A19}$$

which is equal to the trace of Λ in Liouville representation divided by d^2. Equation (A19) provides an alternative definition of process fidelity.

There exists a relation between the commonly used average fidelity F_{ave} and the process fidelity [49],

$$F_{\text{ave}} = \frac{dF + 1}{d + 1}. \tag{A20}$$

The average fidelity is defined as

$$F_{\text{ave}} = \int d\psi \, \text{tr}(|\psi\rangle\langle\psi|\Lambda(|\psi\rangle\langle\psi|)), \tag{A21}$$

where the integral is taken over the Haar measure, it means the average fidelity of the ideal final state and the realistic final state over all pure states. Both the process fidelity and the average fidelity are well-defined metrics for quantifying the closeness of a quantum channel to the identity. Below, we refer to the fidelity to the process fidelity without further explanation. Also, note that in reality, a quantum gate, U, is noisy, and its noisy version can be expressed as the composite channel of $\mathcal{U}$ and its noise channel Λ. Then, the fidelity of U refers to the fidelity of its noise channel Λ.

3. Group twirling and representation theory

Representation theory is an effective tool for analyzing abstract groups and is especially useful in randomized benchmarking and randomized compiling. Below, we briefly introduce basic concepts in representation theory and refer to Refs. [50,51] for systematic introduction. Let G be a finite group and $G \in \mathsf{G}$ be an element of the group. The representation of G is defined as follows.

Definition 1 (Group representation). A map, ϕ, is called a representation of the group G on a linear space, V, if it is a group homomorphism mapping from G to $GL(V)$, where

$$\begin{aligned} \phi : \mathsf{G} &\to GL(V), \\ g &\mapsto \phi(G), \ \forall G \in \mathsf{G}. \end{aligned} \tag{A22}$$

Here, $GL(V)$ denotes the general linear group of V. The representation ϕ satisfies the condition of preserving multiplication that for all $G_1, G_2 \in \mathsf{G}$,

$$\phi(G_1)\phi(G_2) = \phi(G_1 G_2). \tag{A23}$$

Intuitively, representation is just using a matrix group to represent G. Below, we introduce the concept of irreducible representation. Given a representation, ϕ, on V, a linear subspace, $W \subseteq V$, is referred to as invariant if for all $w \in W$ and for all $G \in \mathsf{G}$,

$$\phi(G)w \in W. \tag{A24}$$

The restriction of ϕ to the invariant subspace W is called the subrepresentation of G on W. Note that any representation has a subrepresentation mapping all elements in G to 1 where $W = \{0\}$ and has itself as a subrepresentation where $W = V$. These two subrepresentations are both trivial. With the concept of subrepresentation, we define the irreducible representation as below.

Definition 2 (Irreducible representation). A representation ϕ of the group G on the linear space V is said to be irreducible if it only possesses trivial subrepresentations.

Maschke's theorem provides an interesting property stating that every representation ϕ of a finite group, G, can be decomposed into the direct sum of irreducible representations. For all $G \in \mathsf{G}$, the decomposition can be expressed as follows:

$$\phi(G) \simeq \bigoplus_{\phi_i \in R_{\mathsf{G}}} \mathbb{I}_{n_i \times n_i} \otimes \phi_i(G), \tag{A25}$$

Here, R_{G} denotes the set of all irreducible representations, and n_i represents the multiplicity of the equivalent irreducible representations of ϕ_i in ϕ.

The trace of a representation is called a character, which is defined below.

Definition 3 (Character function). Let ϕ be a representation over the group G. The character of ϕ is defined as the function $\chi_\phi : \mathsf{G} \to \mathbb{C}$ such that for every $G \in \mathsf{G}$,

$$\chi_\phi(G) = \text{tr}[\phi(G)]. \tag{A26}$$

Character function is a powerful tool in representation theory. We can define the inner product of two character functions as below.

Definition 4 (Inner product of character function). Let χ_1 and χ_2 be two character functions of a finite group, G, their inner product is defined as

$$\langle \chi_1 | \chi_2 \rangle = \frac{1}{|\mathsf{G}|} \sum_{G \in \mathsf{G}} \chi_1^*(G)\chi_2(G). \tag{A27}$$

There is a useful result for the inner product between two characters when χ_1 and χ_2 are both irreducible representations. Then $\langle \chi_1 | \chi_2 \rangle = 1$ if χ_1 and χ_2 are equivalent and $\langle \chi_1 | \chi_2 \rangle = 0$ otherwise. Note that Eq. (A25) tells us that

$$\chi_\phi(G) = \text{tr}(\phi(G)) = \sum_{\phi_i \in R_{\mathsf{G}}} n_i \chi_i(G), \tag{A28}$$

where $\chi_i(G) = \text{tr}(\phi_i(G))$. With the orthonormality of irreducible representations, we obtain $n_i = \langle \chi_i | \chi_\phi \rangle$. Then, we have the following lemma.

Lemma 3 (Multiplicity of irreducible representations). For any irreducible representation ϕ_i with character χ_i, the multiplicity of ϕ_i in ϕ is given by

$$n_i = \langle \chi_j | \chi_\phi \rangle. \tag{A29}$$

We also introduce a concept named multiplicity-free representation for further elaboration.

Definition 5 (Multiplicity-free representation). Given a representation, ϕ, of group G, if for any irreducible representation, ϕ_i, the multiplicity of ϕ_i in ϕ is 1, then we call ϕ a multiplicity-free representation.

Using the character function, we can also obtain the generalized projection formula utilized in character randomized benchmarking [37].

Lemma 4 (Generalized projection formula). Consider a finite group, G, and its representation ϕ. Let ϕ_i be an irreducible representation contained in ϕ with character χ_i. The projector onto the support space of ϕ_i can be expressed as

$$\Pi_i = \frac{d_i}{|\mathsf{G}|} \sum_{G \in \mathsf{G}} \chi_i(G)\phi(G), \tag{A30}$$

where $d_i = \dim \phi_i$ denotes the dimension of ϕ_i.

The representation theory is also useful for us to analyze the group twirling on a channel, Λ, over a group, G.

Definition 6 (Group Twirling). For a representation ϕ of the group G on the linear space V, the twirling of a linear map, $\Lambda : V \to V$ over G is defined as

$$\Lambda_{\mathsf{G}} = \frac{1}{|\mathsf{G}|} \sum_{G\in\mathsf{G}} \phi(G)^{\dagger} \Lambda \phi(G). \tag{A31}$$

To analyze the result of group twirling, we introduce Schur's lemma, which is essential in our further elaboration.

Lemma 5 (Schur's lemma). Let $\phi_1 : \mathsf{G} \to GL(V_1)$ and $\phi_2 : \mathsf{G} \to GL(V_2)$ be two arbitrary irreducible representations of group G and $A : V_1 \to V_2$ be a linear map from V_1 to V_2. Suppose for any $G \in \mathsf{G}$,

$$A\phi_1(G) = \phi_2(G)A. \tag{A32}$$

Then, A equals 0, mapping all vectors in V_1 to zero vector in V_2, if ϕ_1 and ϕ_2 are inequivalent irreducible representations; A is proportional to the identity operator if ϕ_1 and ϕ_2 are equivalent irreducible representations. Note that when ϕ_1 and ϕ_2 are equivalent, V_1 and V_2 are equivalent linear space, so the identity operator is well-defined.

Using Schur's Lemma, we can establish the following proposition.

Proposition 1. Let $\phi_1 : \mathsf{G} \to GL(V_1)$ and $\phi_2 : \mathsf{G} \to GL(V_2)$ be two arbitrary irreducible representations of group G and $A : V_1 \to V_2$ be an arbitrary linear map from V_1 to V_2. Define

$$A_G = \mathbb{E}_{G\in\mathsf{G}} \phi_2^{\dagger}(G) A \phi_1(G). \tag{A33}$$

Then,

$$A_G = 0, \tag{A34}$$

if ϕ_1 and ϕ_2 are inequivalent, and

$$A_G = \frac{\mathrm{tr}(A)}{\dim V_1} \mathbb{I}, \tag{A35}$$

if ϕ_1 and ϕ_2 are equivalent, where $\mathbb{I}$ is the identity operator mapping from V_1 to V_2.

For further elaboration and proof, we provide the Burnside theorem in the representation theory.

Lemma 6. Consider a finite group, G, and its all inequivalent irreducible representations $R_G = \{\phi_i\}$, then the cardinality of the twirling group, $|\mathsf{G}|$ can be given by

$$|\mathsf{G}| = \sum_{\phi_i \in R_G} \dim \phi_i^2. \tag{A36}$$

4. Randomized benchmarking and diagonalizability of twirled noise channel

Below, we introduce RB and character randomized benchmarking [37]. Let us start with a brief review of the standard RB, which aims to estimate the fidelity of a quantum gate group. We consider an n-qubit gate group, G, where each gate $G \in \mathsf{G}$ is assumed to be with a noise channel Λ in implementation. In reality, the noise channels for different gates can be different. Below, we simply consider the first-order approximation [7], which is valid and useful in experiments, and regard the noise channels for different gates in G as the same. One can express the noisy quantum gate $\widetilde{\mathcal{G}} = \Lambda\mathcal{G}$ as the composite of the noiseless gate G and its noise channel Λ for any gate $G \in \mathsf{G}$. Then, the task of RB is estimating the fidelity of Λ robust to SPAM error.

To achieve that, one always performs random gate sequences composed of m random gates from G and an inverse quantum gate where m is a prefixed integer called circuit depth,

$$\widetilde{S} = \widetilde{\mathcal{G}}_{\mathrm{inv}} \prod_{i=1}^{m} \widetilde{\mathcal{G}}_i = \Lambda\mathcal{G}_{\mathrm{inv}} \prod_{i=1}^{m} \Lambda\mathcal{G}_i, \tag{A37}$$

where $G_{\mathrm{inv}} = (\prod_{i=1}^{m} G_i)^{\dagger}$, $\mathcal{G}_i$ denotes the Liouville representation of G_i, and $\widetilde{\cdot}$ means the quantum gate is noisy.

In the sense of expectation, the random gate sequence $\widetilde{S}$ is equal to

$$\begin{aligned}\mathbb{E}_{\forall i, G_i\in\mathsf{G}} \widetilde{S} &= \Lambda \mathbb{E}_{\forall i, G_i\in\mathsf{G}} \prod_{i=1}^{m} \left(\prod_{j=1}^{i} \mathcal{G}_j\right)^{\dagger} \Lambda \left(\prod_{j=1}^{i} \mathcal{G}_j\right) \\ &= \Lambda (\mathbb{E}_{G\in\mathsf{G}} \mathcal{G}^{\dagger} \Lambda \mathcal{G})^m \\ &= \Lambda \Lambda_G^m. \end{aligned} \tag{A38}$$

The second line utilizes the fact that G is a group. Thus, $\prod_{j=1}^{i} \mathcal{G}_j$ independently and identically satisfy the uniform distribution on G for any $1 \leqslant j \leqslant m$.

If we use the same state preparation and measurement for all random sequences, then in the expectation, we can get

$$\begin{aligned} f(m) &= \mathbb{E}_{\forall i, G_i\in\mathsf{G}} \langle\langle M|\widetilde{S}|\rho\rangle\rangle \\ &= \langle\langle M|\Lambda\Lambda_G^m|\rho\rangle\rangle \\ &= \langle\langle M'|\Lambda_G^m|\rho\rangle\rangle, \end{aligned} \tag{A39}$$

where ρ and M are prefixed initial state and measurement, respectively, and M' is the measurement absorbing the noise of the inverse gate, satisfying $\langle\langle M'| = \langle\langle M|\Lambda$.

The next step is obtaining the trace of Λ_G robust to state preparation and measurement (SPAM) error. Note that the fidelity $F(\Lambda) = \frac{1}{d^2}\mathrm{tr}(\Lambda) = \frac{1}{d^2}\mathrm{tr}(\Lambda_G)$. Normally, Λ_G has few parameters as it has been twirled and becomes symmetric. These parameters can be evaluated robustly to SPAM error via obtaining $f(m)$ with different circuit depths m and employing exponential fitting. Then, the trace of Λ_G, or the fidelity, can be evaluated with the parameters of Λ_G.

Before the work of character RB [37], researchers mainly focus on a large group, G, like the Clifford group, to make Λ_G highly symmetric with few parameters. For Clifford RB [6,7], Λ_G is a depolarizing channel with a single parameter p satisfying $\Lambda_G(\rho) = p\rho + (1-p)\frac{\mathbb{I}_d}{d}$. In Pauli-Liouville representation, $\Lambda_G = |\sigma_0\rangle\rangle\langle\langle\sigma_0| + p\sum_{i\geqslant 1} |\sigma_i\rangle\rangle\langle\langle\sigma_i|$. Then, $f(m)$ has a single exponential decay expression

$$f(m) = Ap^m + B, \tag{A40}$$

where p is only relevant to Λ, and A and B are only relevant to SPAM. Thus, via single exponential fitting, one can get p along with the trace of Λ_G and the fidelity of Λ.

For CNOT dihedral RB [16], $\Lambda_G = \Pi_0 + p_Z\Pi_Z + p_X\Pi_X$ has two undetermined parameters, p_Z and p_X where $\Pi_0 = |\sigma_0\rangle\rangle\langle\langle\sigma_0|$, $\Pi_Z = \sum_{\sigma_z\in\{\frac{\mathbb{I}}{\sqrt{2}}, \frac{Z}{\sqrt{2}}\}^{\otimes n}/\sigma_0} |\sigma_z\rangle\rangle\langle\langle\sigma_z|$, and $\Pi_X =$

$\sum_{\sigma_x \in \mathsf{P}'_n/\{\frac{\mathbb{I}}{\sqrt{2}},\frac{Z}{\sqrt{2}}\}^{\otimes n}} |\sigma_x\rangle\rangle\langle\langle\sigma_x|$, then $f(m)$ is a double exponential decay function with parameters p_Z and p_X. One has to employ double exponential fitting to obtain $\mathrm{tr}(\Lambda_G)$ and its fidelity.

For a general group G, $f(m)$ is complex and generally not an exponential decay function [36]. But if Λ_G can be written as

$$\Lambda_G = \sum_i p_i \Pi_i, \tag{A41}$$

where $\{\Pi_i\}$ are mutually orthogonal projectors only dependent on G in the space of Liouville representation, we can obtain all decay parameters p_i with only single exponential fitting via the technique of character RB [37]. Note that Eq. (A41) can also be interpreted as that Λ_G is diagonal in the Pauli-Liouville representation up to a unitary transformation $\mathcal{T}$,

$$\Lambda_G = \mathcal{T}\left(\sum_i p_i \Pi_i^P\right)\mathcal{T}^\dagger, \tag{A42}$$

where Π_i^P is a projector, equaling the summation of the Pauli operator bases. The projector Π_i^P and the unitary transformation $\mathcal{T}$ are independent of the channel Λ and are only related to group G. We consider the case that $\mathcal{T}$ is a Liouville representation of a unitary gate T.

Note that in this case,

$$f(m) = \langle\langle M'|\Lambda_G^m|\rho\rangle\rangle = \sum_i \langle\langle M'|\Pi_i|\rho\rangle\rangle p_i^m \tag{A43}$$

is a multiple exponential decay function. If we directly fit $f(m)$ with a multiple exponential decay function, $\sum_i A_i p_i^m$, then the fitting process would consume massive computational resources, and the result is normally inaccurate. The key step of character RB is utilizing Lemma 4. Instead of implementing the gate sequence in Eq. (A37), in character RB, we select a projector Π_j from $\{\Pi_i\}$ and implement

$$\widetilde{S}' = \widetilde{\mathcal{G}}_{inv}\left(\prod_{i=1}^m \widetilde{\mathcal{G}}_i\right)\widetilde{\mathcal{G}}' = \Lambda\mathcal{G}_{inv}\left(\prod_{i=1}^m \Lambda\mathcal{G}_i\right)\mathcal{G}', \tag{A44}$$

where G' is named character gate and is independently sampled from a predetermined gate set, named character group, G'. We select a character function χ' associated with an irreducible representation ϕ' of G' and obtain a projector according to Lemma 4,

$$\Pi' = \frac{\dim \phi'}{|\mathsf{G}'|}\sum_{G'\in\mathsf{G}'} \chi'(G')\mathcal{G}', \tag{A45}$$

such that for any projector $\Pi_i \in \{\Pi_i\}$,

$$\Pi'\Pi_i = \delta_{ij}\Pi'. \tag{A46}$$

Meanwhile, we consider $\langle\langle M|\widetilde{S}'\dim\phi'\chi'(G')|\rho\rangle\rangle$ and in the expectation, this quantity equals

$$\begin{aligned} f_j(m) &= \mathbb{E}_{\forall i, G_i\in\mathsf{G}, G'\in\mathsf{G}'}\langle\langle M|\widetilde{S}'\dim\phi'\chi'(G')|\rho\rangle\rangle \\ &= \langle\langle M|\Lambda\Lambda_G^m\Pi'|\rho\rangle\rangle \\ &= \langle\langle M'|\Lambda_G^m\Pi'|\rho\rangle\rangle \\ &= \sum_i p_i^m\langle\langle M'|\Pi_i\Pi'|\rho\rangle\rangle \\ &= \langle\langle M'|\Pi'|\rho\rangle\rangle p_j^m. \end{aligned} \tag{A47}$$

Thus, p_j can be obtained via single exponential fitting and due to the arbitrariness of Π_j, all parameters of Λ_G can be obtained via single exponential fitting, and hence the fidelity can be evaluated accurately.

Note that in character RB, the key point is realizing projector Π'. Below we show that we can always realize Π' satisfying Eq. (A46). First, each projector can be decomposed as

$$\Pi_j = \sum_{k=1}^{\mathrm{tr}(\Pi_j)} |\mu_{j_k}\rangle\rangle\langle\langle\mu_{j_k}|, \tag{A48}$$

where $\{\mu_{j_k}, 1 \leqslant k \leqslant \mathrm{tr}(\Pi_j), \forall\Pi_j \in \{\Pi_i\}\}$ forms an orthonormal operator basis in Liouville representation. As normalized Pauli operators P'_n is also an orthonormal basis, there exists a unitary transformation linking the two bases. Then, given Π_j from $\{\Pi_i\}$, we select $|\mu_{j_1}\rangle\rangle\langle\langle\mu_{j_1}|$, which is equal to $\mathcal{T}|\sigma\rangle\rangle\langle\langle\sigma|\mathcal{T}^\dagger = |T\sigma T^\dagger\rangle\rangle\langle\langle T\sigma T^\dagger|$ where $|\sigma\rangle\rangle\langle\langle\sigma|$ is a Pauli operator basis and T is the unitary transformation linking two bases. For Pauli group P_n, we have the following equation:

$$|\sigma\rangle\rangle\langle\langle\sigma| = \mathbb{E}_{P\in\mathsf{P}_n}\chi_\sigma(P)\mathcal{P}, \tag{A49}$$

where $\chi_\sigma = (-1)^{\langle P,\sigma\rangle}$ equals 1 when P and σ commute and -1 otherwise. Then, we only need to choose $\mathsf{G}' = T\mathsf{P}_nT^\dagger$ and $\chi' = \chi_\sigma$ to realize

$$\Pi' = \mathcal{T}|\sigma\rangle\rangle\langle\langle\sigma|\mathcal{T}^\dagger = \mathbb{E}_{P\in\mathsf{P}_n}\chi'(P)\mathcal{T}\mathcal{P}\mathcal{T}^\dagger. \tag{A50}$$

Then, Π' satisfies Eq. (A46).

In summary, as long as Λ_G has an expression of Eq. (A41) or Eq. (A42), then with the technique of character RB, one can obtain $\mathrm{tr}(\Lambda_G)$ and the fidelity of Λ accurately with only single exponential fitting. In Appendix B, we would prove that if the Liouville representation of G is not multiplicity-free, as defined in Definition 5, then Λ_G cannot be diagonal for *arbitrary* noise channel Λ. If the Liouville representation of G is multiplicity-free, then Ref. [37] has shown that the character group G' can be chosen as a subgroup of G. In this case, each projector Π_i in Eq. (A41) relates to an irreducible representation subspace of G. Denote this irreducible representation as ϕ_i with character χ_i, then Π_i can be realized with

$$\Pi_i = \frac{\dim\phi_i}{|\mathsf{G}|}\sum_{G'\in\mathsf{G}}\chi_i(G')\mathcal{G}'. \tag{A51}$$

Thus, implementing character gate G' is not harder than the twirling group.

Besides adding character gates, there is another method to effectively realize the projector Π_i in Eq. (A41). If one can realize a measurement M such that $\sum_i p_i\langle\langle M'|\Pi_i = p_j\langle\langle M'|\Pi_j$, then Eq. (A43) will also reduce to $p_j^m\langle\langle M'|\Pi_j|\rho\rangle\rangle$. An accurate initial state ρ, satisfying $\sum_i p_i\Pi_i|\rho\rangle\rangle = p_j\Pi_j|\rho\rangle\rangle$, can also achieve that. Thus, if we have some information about SPAM, then we can realize the projection without the need to implement character gates.

Note that the realization of one projector allows one to get one of the diagonal elements in the twirled noise channel. To estimate the fidelity, one has to get multiple different diagonal terms in the twirled noise channel. Given the twirled noise

channel, $\Lambda_G = \sum_i p_i \Pi_i$, the fidelity of Λ is given by

$$F(\Lambda) = \sum_i \frac{\mathrm{tr}(\Pi_i)}{4^n} p_i. \tag{A52}$$

It is possible that the number of indexes i in the summation increases exponentially with respect to the number of qubits n, like the Pauli group. Nonetheless, in this case, one only needs to sample and estimate a constant number of diagonal elements to estimate the fidelity within acceptable error margins and confidence levels. Particularly, take an index set I in which element i is randomly sampled with probability $\frac{\mathrm{tr}(\Pi_i)}{4^n}$. Then, by Hoeffding inequality, the estimator $\frac{1}{|I|}\sum_{i\in I} p_i$ satisfies

$$\Pr\left(\left|\frac{1}{|I|}\sum_{i\in I} p_i - F(\Lambda)\right| \geqslant \epsilon\right) \leqslant 2\exp\left(-\frac{|I|\epsilon^2}{2}\right). \tag{A53}$$

Thus, $|I|$ only needs to take $2\epsilon^{-2}(\ln 2\delta^{-1})$ to make the estimation within error margin ϵ and confidence level $1-\delta$. The size of I is independent of the qubit count.

5. Interleaved randomized benchmarking

Above, we only introduce how to evaluate the fidelity of a quantum gate group G via randomized benchmarking. To obtain the gate fidelity of an individual target gate, U, one needs to utilize the technique of interleaved RB [38]. In Ref. [38], the target gate U is embedded into a group, G. Below, we call them the twirled gate and the twirling group, respectively. To enable interleaved RB, one implements two kinds of circuits. The first is just a random gate sequence in regular RB, extracting the average gate fidelity of the twirling group G. The second type of circuit is composed of random twirling gates from G interleaved with the target gate U, extracting the composition gate fidelity of the twirling group and the target gate. Comparing the two results, one can get the individual gate fidelity of the target gate. Specifically, suppose the noise channel for twirling group G is $\mathcal{E}$ and the noise channel for the target gate is Λ. Then with the RB methods introduced before one can obtain $F(\mathcal{E}) = \frac{\mathrm{tr}(\mathcal{E})}{d^2}$. After that, one implements an interleaved random gate sequence,

$$\widetilde{S}_i = \widetilde{\mathcal{G}}_{\mathrm{inv}} \prod_{i=1}^{m} \widetilde{\mathcal{U}}\widetilde{\mathcal{G}}_i = \mathcal{E}\mathcal{G}_{\mathrm{inv}} \prod_{i=1}^{m} \mathcal{U}\Lambda\mathcal{E}\mathcal{G}_i, \tag{A54}$$

where $G_{\mathrm{inv}} = (\prod_{i=1}^{m} U G_i)^\dagger$, $\mathcal{G}_i$ and $\mathcal{U}$ denote the Liouville representation of G_i and U, respectively, and $\widetilde{\cdot}$ means the quantum gate is noisy. Note that $\widetilde{\mathcal{U}} = \mathcal{U}\Lambda$ while $\widetilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$. The noise channels for U and gates in G are put in different positions, but due to the arbitrariness of Λ and $\mathcal{E}$, the difference does not put any restriction on the noise. In Ref. [38], the target gate U belongs to twirling group G, then under the expectation of sampling of G_i, Eq. (A54) is equal to

$$\begin{aligned}\mathbb{E}_{\forall i, G_i\in\mathsf{G}}\widetilde{S}_i &= \mathcal{E}\mathbb{E}_{\forall i, G_i\in\mathsf{G}} \prod_{i=1}^{m} \left(\left(\prod_{j=2}^{i} \mathcal{G}_j\mathcal{U}\right)\mathcal{G}_1\right)^\dagger \\ &\quad\times \Lambda\mathcal{E}\left(\left(\prod_{j=2}^{i} \mathcal{G}_j\mathcal{U}\right)\mathcal{G}_1\right) \\ &= \mathcal{E}(\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}^\dagger\Lambda\mathcal{E}\mathcal{G})^m = \mathcal{E}(\Lambda\mathcal{E})^m_G. \end{aligned} \tag{A55}$$

The second line utilizes the fact that G is a group. Thus, same as the discussion in the previous subsection, interleaved quantum circuits allow us to estimate the fidelity of $\Lambda\mathcal{E}$, that is, $F(\Lambda\mathcal{E}) = \frac{\mathrm{tr}(\Lambda\mathcal{E})}{d^2}$. In Ref. [38], the authors show how to estimate $F(\Lambda)$ from $F(\Lambda\mathcal{E})$ and $F(\mathcal{E})$. We review this technique as elaborated below.

Denote $\mathcal{E}_d$ to be the $\mathcal{E}$ twirled by the whole unitary group. It is known that $\mathcal{E}_d$ is a mixture of the identity channel and the depolarizing channel,

$$\mathcal{E}_d(\rho) = \frac{d^2F(\mathcal{E})-1}{d^2-1}\rho + \left(1 - \frac{d^2F(\mathcal{E})-1}{d^2-1}\right)\frac{\mathbb{I}}{d}. \tag{A56}$$

Now, let us investigate the quantity $|F(\Lambda\mathcal{E}) - F(\Lambda\mathcal{E}_d)|$, which is equal to

$$|F(\Lambda\mathcal{E}) - F(\Lambda\mathcal{E}_d)| = \left|F(\Lambda\mathcal{E}) - \frac{(d^2F(\Lambda)-1)(d^2F(\mathcal{E})-1)+d^2-1}{d^2(d^2-1)}\right|. \tag{A57}$$

Assuming $|F(\Lambda\mathcal{E}) - F(\Lambda\mathcal{E}_d)|$ is upper bounded by E, we can deduce $F(\Lambda)$ belongs to the following interval,

$$\left[\frac{d^2(F(\Lambda\mathcal{E})-E)-1}{d^2F(\mathcal{E})-1}\left(1-\frac{1}{d^2}\right)+\frac{1}{d^2}, \frac{d^2(F(\Lambda\mathcal{E})+E)-1}{d^2F(\mathcal{E})-1}\left(1-\frac{1}{d^2}\right)+\frac{1}{d^2}\right]. \tag{A58}$$

This interval has a mean value $\frac{d^2F(\Lambda\mathcal{E})-1}{d^2F(\mathcal{E})-1}(1-\frac{1}{d^2})+\frac{1}{d^2}$ and length $\frac{2(d^2-1)E}{d^2F(\mathcal{E})-1}$. With matrix analysis, the quantity E can be upper bounded by Eq. (A59) [38].

$$E \leqslant \min\begin{cases} 4(d+1)\sqrt{1-F(\mathcal{E})} + 2\frac{d+1}{d}(1-F(\mathcal{E})) \\ \frac{\left|d^2(F(\Lambda\mathcal{E})-F(\mathcal{E}))+2F(\mathcal{E})-F(\Lambda\mathcal{E})-1\right|+(d^2F(\mathcal{E})-1)(1-F(\mathcal{E}))}{d^2-1} \\ \frac{\left|d^2(F(\Lambda\mathcal{E})-F(\mathcal{E}))+2F(\mathcal{E})-F(\Lambda\mathcal{E})-1\right|+(d^2F(\mathcal{E})-1)(F(\mathcal{E})-F(\Lambda\mathcal{E}))}{d^2-1} \end{cases}. \tag{A59}$$

In general, the higher $F(\mathcal{E})$ is, the smaller E is, and the more accurate the estimation of $F(\Lambda)$ is. In Ref. [38], the target gate U belongs to G so the noisy levels, or the fidelities, of Λ and $\mathcal{E}$ are close, which is not beneficial for estimating $F(\Lambda)$. This is also one of the reasons that finding a small and practical twirling group G matters. In the main text, we focus on the interleaved circuit to estimate $F(\Lambda\mathcal{E})$, and $F(\Lambda)$ can be estimated via Eqs. (A58) and (A59). Note that though this estimation assumes the noise channel of the twirling group to be gate-independent, this assumption is good enough in a high-fidelity region and could also be relieved by using a higher-order fitting formula by more sophisticated analysis like Ref. [7].

6. Dihedral group and classically replaceable unitary operations

For further elaboration, in this part, we introduce the local dihedral group and CRU operations [33] along with their properties. The local dihedral group on n qubits is similar to the n-qubit Pauli group, which is defined as follows:

$$\mathsf{D}_n^m = \langle X, Z_m\rangle^{\otimes n} = \langle X_1, (Z_m)_1, X_2, (Z_m)_2, \cdots, X_n, (Z_m)_n\rangle, \tag{A60}$$

where $X = \left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$ is the Pauli X gate and $Z_m = \left(\begin{smallmatrix}1 & 0\\ 0 & e^{i\frac{2\pi}{m}}\end{smallmatrix}\right)$ is a phase gate with phase $\frac{2\pi}{m}$. Here, m is a positive integer. In general, the phase on different qubits can be different and we can define $\langle X_1, (Z_{m_1})_1, X_2, (Z_{m_2})_2, \cdots, X_n, (Z_{m_n})_n\rangle$, but below we only consider a simple case that $m_1 = m_2 = \cdots = m_n = m$. Same with the discussion of the Pauli group, we can define the projective local dihedral group via quotient by the center of $\langle X, Z_m\rangle^{\otimes n}$,

$$\mathsf{D}_m^m = \langle X, Z_m\rangle^{\otimes n}/\langle\omega_m\rangle, \tag{A61}$$

where $\omega_m = e^{i\frac{2\pi}{m}}$ and $\langle\omega_m\rangle = \{e^{i\frac{2k\pi}{m}}, 0 \leqslant k \leqslant m-1\}$. As the overall phase is not important, below we do not distinguish the local dihedral group and the projective local dihedral group and use the definition $\mathsf{D}_n^m = \langle X, Z_m\rangle^{\otimes n}/\langle\omega_m\rangle$.

Classically replaceable unitary operations, or incoherent unitary operations, are all unitary gates that can be moved after computational basis measurements and be replaced with classical postprocessing. Given the computational basis $|i\rangle$ on the Hilbert space $\mathcal{H}$, we can define the dephasing operation Δ. For any $\rho \in \mathcal{D}(\mathcal{H})$,

$$\Delta(\rho) = \sum_i |i\rangle\langle i|\rho|i\rangle\langle i|. \tag{A62}$$

Then, the CRU gate set has an expression,

$$\{U \text{ is unitary}|\mathcal{U}\Delta = \Delta\mathcal{U}\}. \tag{A63}$$

Note that we use the same notation Δ to represent its map representation and Liouville representation. It is shown that a CRU can be given by [33,52]

$$U = \sum_j e^{i\theta_j}|\sigma(j)\rangle\langle j|, \tag{A64}$$

where σ is a permutation over computational basis and θ_j is a phase in $[0, 2\pi)$. And any unitary gate having an expression of Eq. (A64) is a CRU. We can decompose U as

$$U = \sum_j |\sigma(j)\rangle\langle j| \sum_j e^{i\theta_j}|j\rangle\langle j|. \tag{A65}$$

Set $\Pi = \sum_j |\sigma(j)\rangle\langle j|$ and $W = \sum_j e^{i\theta_j}|j\rangle\langle j|$ and we can see any CRU can be written as the multiplication of a permutation matrix, Π, and a diagonal matrix, W, $U = \Pi W$. Note that any matrix with the form $\sum_j |\sigma(j)\rangle\langle j|$ for arbitrary permutation σ is defined to be a permutation matrix. Also, the multiplication of a permutation matrix and a diagonal matrix is always expressed as Eq. (A64) and is a CRU.

Lemma 7. Any classically replaceable unitary operation U can be decomposed as the multiplication of a permutation matrix, Π, and a diagonal matrix, W, $U = \Pi W$. Vice versa.

From the above lemma, we can obtain that the computational basis gate set is invariant under the action of a CRU.

Lemma 8. Diagonal matrices set or computational basis gate set is invariant under the action of any CRU.

Proof. For any CRU $U = \sum_j e^{i\theta_j}|\sigma(j)\rangle\langle j|$ and diagonal matrix $W = \sum_j e^{\phi_j}|j\rangle\langle j|$,

$$\begin{aligned} UWU^{-1} &= \sum_{jj'k} e^{i(\phi_k+\theta_j-\theta_{j'})}|\sigma(j)\rangle\langle j|k\rangle\langle k|j'\rangle\langle\sigma(j')| \\ &= \sum_j e^{i\phi_j}|\sigma(j)\rangle\langle\sigma(j)| \\ &= \sum_j e^{i\phi_{\sigma^{-1}(j)}}|j\rangle\langle j|, \end{aligned} \tag{A66}$$

which is a diagonal matrix. Proof is done. ■

In fact, for any CRU subgroup $\mathsf{G} \leqslant$ CRU, all diagonal matrices in G forms a subgroup, G_Z of G. Then, G_Z is invariant under the action of any gate in G, or equivalently, G_Z is a normal subgroup of G.

Lemma 9. Given an n-qubit CRU subgroup G, set the computational basis subgroup of G as $\mathsf{G}_Z = \{U \in \mathsf{G}|U \text{ is diagonal}\}$. Then, G_Z is a normal subgroup of G.

Proof. Obviously, G_Z is a subgroup of G. We only need to prove that under the conjugate action of any quantum gate G in G, G_Z is invariant. For any gate $G \in \mathsf{G}$, $G \in$ CRU, then $G\mathsf{G}_Z G^\dagger$ only contains diagonal matrices. As all diagonal matrices in G are contained in G_Z, $G\mathsf{G}_Z G^\dagger \subseteq \mathsf{G}_Z$. Also, $|G\mathsf{G}_Z G^\dagger| = |\mathsf{G}_Z|$, so $G\mathsf{G}_Z G^\dagger = \mathsf{G}_Z$. As G is an arbitrary gate in G, we prove that G_Z is the normal subgroup of G. Here, we complete the proof.

Below, we present a result about permutation matrices for further elaboration. The lemma tells us the structure of the permutation matrices, which can always be decomposed as the multiplication of Toffoli gates, CNOT gates, and Pauli X gates.

Lemma 10. All permutation matrices on n qubits can be generated by Pauli X gate on each qubit and all Toffoli gates $C^{n-1}X$ with $n-1$ control qubits and 1 target qubit.

Proof. The expression of permutation matrices on n qubits can be unified as follows:

$$\Pi_\sigma = \sum_{j\in\{0,1\}^n} |\sigma(j)\rangle\langle j|, \tag{A67}$$

where σ is an arbitrary permutation on $\{0,1\}^n$. The permutations themself form a permutation group with cardinality $2^n!$. As transpositions can generate the permutation group, we only need to prove that any transposition (s_1, s_2) can be generated by Pauli X gates and $C^{n-1}X$ gates for $s_1 \neq s_2 \in \{0,1\}^n$.

Denote X_k to be the Pauli X gate acting on kth qubit and set $X^s = \bigotimes_{i=1}^n X_i^{s_i}$ for $s \in \{0,1\}^n$. We also denote $C^{n-1}X_k$ to be the Toffoli gate with kth qubit as the target qubit and other qubits as the control qubits. Then, for any $s \in \{0,1\}^n$, $X^s C^{n-1}X_k (X^s)^\dagger$ would swap the basis $|1^n \oplus s\rangle$ and $|1^n \oplus 1_k \oplus s\rangle$ while keeping other bases fixed. Here, $1^n = 11\cdots1$ is the all-1 bit string and $1_k = 0^{k-1}10^{n-k}$ is the bit string with 1 in kth position and 0 in other positions. Thus, by taking k over 1 to n and s over $\{0,1\}^n$ we would obtain any transposition $(s, s \oplus 1_k)$. As $(s, s \oplus 1_{k1})(s, s \oplus 1_{k2})(s, s \oplus 1_{k1}) = (s \oplus 1_{k1}, s \oplus 1_{k2})$, we would also obtain any transposition $(s, s \oplus 1_{k1} \oplus 1_{k2} \ldots \oplus 1_{kl})$, which suffices to produce any transposition (s_1, s_2) for $s_1 \neq s_2 \in \{0,1\}^n$.

Note that the Toffoli gates with less than $n-1$ control qubits are also permutation matrices. It means that with single qubit Pauli X gates and $C^{n-1}X$ gates, one can construct any controlled-X gates C^kX with $1 \leqslant k \leqslant n-2$. ■

APPENDIX B: OPTIMAL TWIRLING GROUPS FOR MULTIQUBIT CONTROLLED PHASE GATES IN RANDOMIZED BENCHMARKING

In original interleaved RB [38], to benchmark an individual target quantum gate, U, one always embeds U in a large group, G, with a number of global quantum gates so that G has a strong twirling effect. However, the large size of the twirling group would make group sampling and computing inverse gates difficult, and plenty of global quantum gates in G would make the twirling gates hard to realize. The former difficulty is a classical computational problem, and the latter is a gate implementability problem. In this work, we focus on finding the smallest and the most local twirling group for a target gate.

Fortunately, Refs. [14,15,37] show that embedding the target gate U in the twirling group G is not necessary to characterize U. A Clifford gate can be effectively characterized with local Clifford twirling or Pauli twirling instead of global Clifford twirling. It gives hope that maybe we can choose a small twirling group for benchmarking a generic target quantum gate. Normally, we would like to choose a twirling group as small as possible, but the twirling group can not be arbitrarily small as well. Otherwise, the twirled noise channel is not symmetric enough, resulting in hard postprocessing and inaccurate fidelity estimation. In the main text, we have briefly introduced the requirements for twirling groups in RB and proposed Question 1. Below, we will present more detailed and complete requirements for twirling groups in RB and rederive Question 1. Moreover, in this part, we will provide the formal versions of the theorems and lemmas in the main text along with their proof. It is worth mentioning that when we write "a channel is diagonal," it always means that "the channel is diagonal up to a unitary transformation."

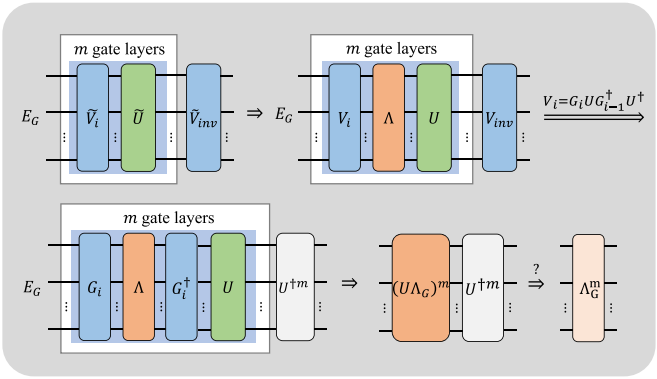


FIG. 6. Random twirling gates $V_1, V_2, \cdots, V_m$ interleaved with U in RB. The inverse gate $V_{\text{inv}} = (\prod_{i=1}^m UV_i)^\dagger$. $\tilde{\cdot}$ represents the noisy version of a quantum gate. Λ is the composite noise channel of V and U, and Λ_G is the G-twirled noise channel.

1. Requirements for twirling groups in randomized benchmarking

Below, we focus on how to estimate the fidelity of a target gate, U, with methods of RB. Same with the notations in the main text and in Appendix A, we express the noisy quantum gate $\tilde{\mathcal{U}} = \mathcal{U}\Lambda$ as a composite of the noiseless gate U and its noise channel Λ, where $\mathcal{U}$ denotes the Pauli-Liouville representation of U, and $\tilde{\cdot}$ represents the noisy version of a quantum gate. As mentioned in Appendix A 4, the key to enable RB is obtaining the powers of the G-twirled noise channel, Λ_G^m, where $m \in \mathbb{Z}+$, and ensuring Λ_G is diagonal up to a unitary transformation, or can be written as Eq. (A42). Then, with the technique of character RB, one can obtain $\text{tr}(\Lambda_G)$ and the fidelity of Λ. Here, G is the twirling group for tailoring Λ.

Different from the elaboration in the main text, below we distinguish two concepts of twirling gate set V and twirling group G. In the main text, we directly select a twirling group, G, and implement a random gate sequence composed of gates from G interleaved with target gate U to obtain Λ_G^m. But in fact, we have another way to obtain Λ_G^m as shown below.

To obtain Λ_G^m, we implement m twirling gates $V_i = G_i U G_{i-1}^\dagger U^\dagger$, sampled from the twirling gate set $\mathsf{V} = \mathsf{G}U\mathsf{G}U^\dagger$, interleaved with the target gate U as shown in Fig. 6. Here, G_i is uniformly and randomly sampled from the group G, and we set $G_0 = \mathbb{I}$. The circuit ends with the inverse gate $V_{\text{inv}} = (\prod_{i=1}^m UV_i)^\dagger = U^{\dagger m-1} G_m^\dagger U^\dagger$. The circuit ideally equals the identity, corresponding to $\Lambda_G = I$. Note that the twirling group G and the twirling gate set V are in general different and $\mathsf{G} \subseteq \mathsf{V}$. Although we only require Λ to be twirled by G, we need to realize gates in V to achieve that since the circuit involves gate U. The twirling gates should first eliminate the effect of the twirled gate U and then influence the noise channel Λ. Note that the inverse gate V_{inv} can always belong to V as long as we choose m such that $U^m = \mathbb{I}$. The cost to implement the inverse gate is nearly the same as other twirling gates.

Same with the main text and the arguments in Appendix A 5, we omit the noise from the twirling gates.

Then, the circuit is represented as

$$\begin{aligned}\widetilde{\mathcal{S}}_m &= \mathcal{V}_{\text{inv}} \prod_{i=1}^{m} \mathcal{U}\Lambda\mathcal{V}_i \\ &= \mathcal{U}^{\dagger m-1}\mathcal{G}_m^\dagger \mathcal{U}^\dagger \prod_{i=1}^{m} \mathcal{U}\Lambda\mathcal{G}_i\mathcal{U}\mathcal{G}_{i-1}^\dagger \mathcal{U}^\dagger \\ &= \mathcal{U}^{\dagger m} \prod_{i=1}^{m} \mathcal{U}\mathcal{G}_i^\dagger \Lambda \mathcal{G}_i. \end{aligned} \tag{B1}$$

In terms of expectation over sampling of G_i, Eq. (B1) becomes $\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m$. Compared to Λ_G^m, this formulation only requires the commutation relation,

$$\Lambda_G\mathcal{U} = \mathcal{U}\Lambda_G, \tag{B2}$$

which means that the symmetry of the group G is preserved under the action of U in the Liouville representation. Thus, to tailor U in RB, our task is finding the twirling group G satisfying Eq. (B2) and ensuring the diagonalizability of Λ_G while minimizing the size of the twirling gate set $\mathsf{V} = \mathsf{G}U\mathsf{G}U^\dagger$. The question is summarized below.

Question 3. Given a gate, U, find V such that $\mathsf{V} = \mathsf{G}U\mathsf{G}U^\dagger$ and the twirling group G satisfies

for any quantum channel Λ, $\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger$ is diagonal up to a unitary transformation independent of Λ, (B3)

and

$$\Lambda_G\mathcal{U} = \mathcal{U}\Lambda_G. \tag{B4}$$

Question 3 is a more refined problem for finding a twirling group in RB. However, in general, Eq. (B2) is challenging to characterize, so we substitute it with a more easily handled condition, that is,

$$U\mathsf{G}U^\dagger = \mathsf{G}. \tag{B5}$$

In this case, $\mathsf{V} = \mathsf{G}$ and Question 3 reduces to Question 1 in the main text, as shown below.

Question 1. Given a gate, U, find a twirling group, G such that,

for any quantum channel Λ, $\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger$ is diagonal up to a unitary transformation independent of Λ, (B6)

and

$$U\mathsf{G}U^\dagger = \mathsf{G}. \tag{B7}$$

As mentioned in the main text, the solution to Question 1 allows us to obtain the fidelity of $(\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}$, which is a lower bound of the fidelity of Λ. Below, we present the proof.

Lemma 11. Given a target gate, U, and a twirling group, G satisfying Question 1, then for any positive integer m, the fidelity of $(\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}$ is a lower bound of the fidelity of Λ. Mathematically,

$$F((\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}) \leqslant F(\Lambda), \tag{B8}$$

or equivalently,

$$\text{tr}((\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}) \leqslant \text{tr}(\Lambda). \tag{B9}$$

Proof. As $\text{tr}(\Lambda) = \text{tr}(\Lambda_G)$, we only need to prove $\text{tr}((\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}) \leqslant \text{tr}(\Lambda_G)$.

Note that Λ_G has the expression of Eq. (A41) and can be written as

$$\Lambda_G = \sum_{i=1}^{k} p_i\Pi_i. \tag{B10}$$

Also, the condition $U\mathsf{G}U^\dagger = \mathsf{G}$ ensures that $\mathcal{U}^\dagger\Lambda_G\mathcal{U} = (\mathcal{U}^\dagger\Lambda_G\mathcal{U})_G$ as proven in Lemma 12, which means we can express $\mathcal{U}^\dagger\Lambda_G\mathcal{U}$ as

$$\mathcal{U}^\dagger\Lambda_G\mathcal{U} = \sum_{i=1}^{k} q_i\Pi_i. \tag{B11}$$

As $\mathcal{U}$ is unitary and does not change the spectrum of Λ_G, the value of q_i must be equal to one of elements from $\{p_i, 1 \leqslant i \leqslant k\}$. Thus, we can construct a map $f_1 : \{p_i, 1 \leqslant i \leqslant k\} \to \{q_i, 1 \leqslant i \leqslant k\}$. Similarly, as $\mathcal{U}(\mathcal{U}^\dagger\Lambda_G\mathcal{U})\mathcal{U}^\dagger = \Lambda_G$, we can construct a map $f_2 : \{q_i, 1 \leqslant i \leqslant k\} \to \{p_i, 1 \leqslant i \leqslant k\}$ so that the compositions of f_1 and f_2 are identity maps, $f_1 \circ f_2(q_i) = q_i$ and $f_2 \circ f_1(p_i) = p_i$. Then, f_1 is just a permutation on $\{p_i, 1 \leqslant i \leqslant k\}$ and $\{q_i, 1 \leqslant i \leqslant k\}$ is equivalent to $\{p_i, 1 \leqslant i \leqslant k\}$ after permutation. Thus, the diagonal terms of $\mathcal{U}^\dagger\Lambda_G\mathcal{U}\Lambda_G$ are $\{p_ip_{\sigma(i)}\}$ where σ is a permutation on $\{1, 2, \cdots, k\}$.

Similarly, the diagonal terms of $\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m = (\mathcal{U}^\dagger\cdots(\mathcal{U}^\dagger(\mathcal{U}^\dagger\Lambda_G\mathcal{U})\Lambda_G\mathcal{U})\Lambda_G\cdots\mathcal{U})\Lambda_G$ are $\{p_ip_{\sigma_1(i)}\cdots p_{\sigma_{m-1}(i)}\}$ where $\sigma_1, \cdots, \sigma_{m-1}$ are permutations on $\{1, 2, \cdots, k\}$. Thus, with rearrangement inequality,

$$\begin{aligned}\text{tr}((\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}) &= \sum_{i=1}^{k}(p_ip_{\sigma_1(i)}\cdots p_{\sigma_{m-1}(i)})^{\frac{1}{m}} \\ &\leqslant \sum_{i=1}^{k} p_i \\ &= \text{tr}(\Lambda_G). \end{aligned} \tag{B12}$$

Here is the proof. ■

Thus, solve Question 1, and then we can provide a lower bound of the fidelity of the target gate. Based on the proof of Lemma 11, this lower bound is close to the real fidelity as long as the diagonal terms of the noise channel in the Pauli-Liouville representation are close to each other. And the lower bound is saturated for a global depolarizing noise. More importantly, if the target gate U is a multiqubit controlled phase gate, then the twirling group satisfying $U\mathsf{G}U^\dagger = \mathsf{G}$ would also satisfy $\Lambda_G\mathcal{U} = \mathcal{U}\Lambda_G$, which we will show in Theorem 2. As a consequence, we can accurately estimate the fidelities of multiqubit controlled phase gates instead of only providing a lower bound.

Lemma 12. Given a unitary gate, U, and a unitary subgroup, G, if $U\mathsf{G}U^\dagger = \mathsf{G}$, then for any quantum channel Λ, $(\mathcal{U}\Lambda_G\mathcal{U}^\dagger)_G = \mathcal{U}\Lambda_G\mathcal{U}^\dagger$.

Proof. Provided with $U\mathsf{G}U^\dagger = \mathsf{G}$, we have

$$\begin{aligned}(\mathcal{U}\Lambda_G\mathcal{U}^\dagger)_G &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\mathcal{U}\Lambda_G\mathcal{U}^\dagger\mathcal{G}^\dagger \\ &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{U}(\mathcal{U}^\dagger\mathcal{G}\mathcal{U})\Lambda_G(\mathcal{U}^\dagger\mathcal{G}^\dagger\mathcal{U})\mathcal{U}^\dagger \\ &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{U}\mathcal{G}\Lambda_G\mathcal{G}^\dagger\mathcal{U}^\dagger \\ &= \mathcal{U}\Lambda_G\mathcal{U}^\dagger. \end{aligned} \tag{B13}$$

The equality in the third line comes from the condition $U\mathsf{G}U^\dagger = \mathsf{G}$. ■

2. Proof of Lemma 1

In the main text, we present the lemma that for a finite n-qubit unitary subgroup G, if for any quantum channel Λ, its $\mathsf{G}-$twirled channel, Λ_G, is diagonal in the Pauli-Liouville representation up to a unitary transformation independent of Λ, then the Pauli-Liouville representation of G is multiplicity-free. Also, the cardinality of the twirling group $|\mathsf{G}| \geqslant 4^n$. Below, we present the proof of this lemma. It is worth mentioning that, in the proof, we consider a more generic scenario: if Λ_G is diagonal up to an invertible matrix transformation rather than a unitary transformation, then the Pauli-Liouville representation of G is multiplicity-free. The result is more generic, and the proof is stronger here.

For convenience, we rewrite the Lemma 1 below.

Lemma 1. For a finite n-qubit unitary subgroup, G, if for any quantum channel Λ, its $\mathsf{G}-$twirled channel, Λ_G, is diagonal in the Pauli-Liouville representation up to an invertible matrix transformation independent of Λ, then the Pauli-Liouville representation of G is multiplicity-free. As a corollary, the cardinality of the twirling group $|\mathsf{G}| \geqslant 4^n$.

Proof. Let us begin with the irreducible representation decomposition of the Pauli Liouville representation of G. For each element $G \in \mathsf{G}$, $\mathcal{G}$ can be decomposed as the direct sum of irreducible representations up to an isomorphism $\mathcal{V}$,

$$\mathcal{V}\mathcal{G}\mathcal{V}^{-1} = \bigoplus_{i=1}^{k} \mathbb{I}_{n_i\times n_i} \otimes \phi_i(G), \tag{B14}$$

where $\mathcal{V}$ is an invertible matrix, ϕ_i denotes the irreducible representation of G, and n_i denotes its multiplicity in $\mathcal{G}$. The term k records the number of inequivalent irreducible representations that $\mathcal{G}$ contains. It is worth mentioning that the basis making $\mathcal{G}$ block-diagonal may not be Pauli operators $\{\mathbb{I}, X, Y, Z\}$. That is why we put $\mathcal{V}$ and $\mathcal{V}^{-1}$ in the left side of Eq. (B14). As any unitary channel has an invariant subspace $\{\mathbb{I}\}$, the Pauli-Liouville representation of a unitary channel must be block-diagonal as follows:

$$\mathcal{G} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{\mathcal{G}} \end{pmatrix}. \tag{B15}$$

Thus, the form of the basis transformation matrix, $\mathcal{V}$, can also be constrained like Eq. (B15). That means $\mathcal{V}$ only changes the nonidentity basis while keeping the basis $\mathbb{I}$ invariant. In addition, without loss of generality, we set the irreducible representation in the subspace spanned by $\{\mathbb{I}\}$ as the trivial representation, that is, mapping all group elements to 1.

Given a channel, Λ, its twirling over group G is

$$\begin{aligned}\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^{-1} &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{V}^{-1}\left(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G)\right)\mathcal{V}\Lambda\mathcal{V}^{-1}\left(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G)\right)^{-1}\mathcal{V} \\ &= \mathcal{V}^{-1}\left[\mathbb{E}_{G\in\mathsf{G}}\left(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G)\right)\mathcal{V}\Lambda\mathcal{V}^{-1}\left(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G)\right)^{-1}\right]\mathcal{V}.\end{aligned} \tag{B16}$$

Focusing on the calculation in the square brackets in Eq. (B16), we denote $\Lambda' = \mathcal{V}\Lambda\mathcal{V}^{-1}$ and decompose it corresponding to the block partition of the right-hand side in Eq. (B14). That is,

$$\Lambda' = \begin{pmatrix} \Lambda'^{11} & \Lambda'^{12} & \cdots & \Lambda'^{1k} \\ \Lambda'^{21} & \Lambda'^{22} & \cdots & \Lambda'^{2n_k} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda'^{k1} & \Lambda'^{k2} & \cdots & \Lambda'^{kk} \end{pmatrix}, \tag{B17}$$

where we set

$$\Lambda'^{ii} = \begin{pmatrix} \Lambda'^{i}_{11} & \Lambda'^{i}_{12} & \cdots & \Lambda'^{i}_{1n_i} \\ \Lambda'^{i}_{21} & \Lambda'^{i}_{22} & \cdots & \Lambda'^{i}_{2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda'^{i}_{n_i1} & \Lambda'^{i}_{n_i2} & \cdots & \Lambda'^{i}_{n_in_i} \end{pmatrix}. \tag{B18}$$

Then, we get Eq. (B19).

$$\begin{aligned}&\mathbb{E}_{G\in\mathsf{G}}\left(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G)\right)\Lambda'\left(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G)\right)^{-1} \\ &= \mathbb{E}_{G\in\mathsf{G}}\bigoplus_{i=1}^{k}\left[\begin{pmatrix}\phi_i(G) & 0 & \cdots & 0\\ 0 & \phi_i(G) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \phi_i(G)\end{pmatrix}\begin{pmatrix}\Lambda'^{i}_{11} & \Lambda'^{i}_{12} & \cdots & \Lambda'^{i}_{1n_i}\\ \Lambda'^{i}_{21} & \Lambda'^{i}_{22} & \cdots & \Lambda'^{i}_{2n_i}\\ \vdots & \vdots & \ddots & \vdots\\ \Lambda'^{i}_{n_i1} & \Lambda'^{i}_{n_i2} & \cdots & \Lambda'^{i}_{n_in_i}\end{pmatrix}\begin{pmatrix}\phi_i^{-1}(G) & 0 & \cdots & 0\\ 0 & \phi_i^{-1}(G) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \phi_i^{-1}(G)\end{pmatrix}\right]\end{aligned}$$

$$= \mathbb{E}_{G\in\mathsf{G}} \bigoplus_{i=1}^{k} \begin{pmatrix} \phi_i(G)\Lambda'^i_{11}\phi_i^{-1}(G) & \phi_i(G)\Lambda'^i_{12}\phi_i^{-1}(G) & \cdots & \phi_i(G)\Lambda'^i_{1n_i}\phi_i^{-1}(G) \\ \phi_i(G)\Lambda'^i_{21}\phi_i^{-1}(G) & \phi_i(G)\Lambda'^i_{22}\phi_i^{-1}(G) & \cdots & \phi_i(G)\Lambda'^i_{2n_i}\phi_i^{-1}(G) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_i(G)\Lambda'^i_{n_i1}\phi_i^{-1}(G) & \phi_i(G)\Lambda'^i_{n_i2}\phi_i^{-1}(G) & \cdots & \phi_i(G)\Lambda'^i_{n_in_i}\phi_i^{-1}(G). \end{pmatrix}$$

$$= \bigoplus_{i=1}^{k} \begin{pmatrix} \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{11}\phi_i^{-1}(G) & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{12}\phi_i^{-1}(G) & \cdots & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{1n_i}\phi_i^{-1}(G) \\ \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{21}\phi_i^{-1}(G) & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{22}\phi_i^{-1}(G) & \cdots & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{2n_i}\phi_i^{-1}(G) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{n_i1}\phi_i^{-1}(G) & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{n_i2}\phi_i^{-1}(G) & \cdots & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^i_{n_in_i}\phi_i^{-1}(G) \end{pmatrix}, \tag{B19}$$

By Schur's lemma, for any irreducible representation ϕ_i and matrix A, the twirling of A over G would be proportional to identity,

$$\mathbb{E}_{G\in\mathsf{G}}\phi_i(G)A\phi_i^{-1}(G) = \mathrm{tr}(A)\frac{\mathbb{I}_{d_i}}{d_i}, \tag{B20}$$

where $d_i = \dim \phi_i$. We set

$$\Lambda'_i = \begin{pmatrix} \mathrm{tr}\left(\Lambda'^i_{11}\right) & \mathrm{tr}\left(\Lambda'^i_{12}\right) & \cdots & \mathrm{tr}\left(\Lambda'^i_{1n_i}\right) \\ \mathrm{tr}\left(\Lambda'^i_{21}\right) & \mathrm{tr}\left(\Lambda'^i_{22}\right) & \cdots & \mathrm{tr}\left(\Lambda'^i_{2n_i}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}\left(\Lambda'^i_{n_i1}\right) & \mathrm{tr}\left(\Lambda'^i_{n_i2}\right) & \cdots & \mathrm{tr}\left(\Lambda'^i_{n_in_i}\right) \end{pmatrix}. \tag{B21}$$

Thus, we conclude that

$$\mathbb{E}_{G\in\mathsf{G}}\left(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G)\right)\Lambda'\left(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G)\right)^{-1}$$

$$= \bigoplus_{i=1}^{k}\Lambda'_i\otimes\frac{\mathbb{I}_{d_i}}{d_i}, \tag{B22}$$

and

$$\Lambda_G = \mathcal{V}^{-1}\left(\bigoplus_{i=1}^{k}\Lambda'_i\otimes\frac{\mathbb{I}_{d_i}}{d_i}\right)\mathcal{V}. \tag{B23}$$

Note that Λ_G is diagonal up to an invertible matrix transformation independent of Λ. There exists a fixed invertible matrix transformation $\mathcal{V}'$ such that

$$\mathcal{V}'\Lambda_G\mathcal{V}'^{\dagger} = \sum_i p_i\Pi_i, \tag{B24}$$

where Π_i is a projector independent of Λ and $\mathcal{V}'$ is also independent of Λ.

Notice that Eq. (B24) is a linear function acting on Λ. If for any channel Λ, Λ_G is diagonal, then the linear combination of any set of quantum channels would also be diagonal after G-twirling. Note that the linear span of quantum channels is the set of all trace-preserving (TP) maps [53]. Thus, for any TP map with form $\Lambda = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{t} & \mathbf{T}_\Lambda \end{pmatrix}$, $\mathcal{V}'\Lambda_G\mathcal{V}'^{\dagger}$ must be diagonal. It further requires that for any index i, Λ'_i defined in Eq. (B21) is diagonal up to an invertible matrix transformation. Recall that Λ'_i is defined from Λ', which equals $\mathcal{V}\Lambda\mathcal{V}^{-1}$. As $\mathcal{V}$ has the form $\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_\mathcal{V} \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{t} & \mathbf{T}_\Lambda \end{pmatrix}$ is an arbitrary TP map, $\Lambda' = \mathcal{V}\Lambda\mathcal{V}^{-1} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{T}_\mathcal{V}\mathbf{t} & \mathbf{T}_\mathcal{V}\mathbf{T}_\Lambda\mathbf{T}_\mathcal{V}^{-1} \end{pmatrix}$ is also an arbitrary TP map. Denote $\mathbf{t}' = \mathbf{T}_\mathcal{V}\mathbf{t}$ and $\mathbf{T}_{\Lambda'} = \mathbf{T}_\mathcal{V}\mathbf{T}_\Lambda\mathbf{T}_\mathcal{V}^{-1}$, the elements of $\mathbf{t}'$ and $\mathbf{T}_{\Lambda'}$ would be arbitrary. Back to considering Λ'_i, when $i = 1$, ϕ_1 is the trivial irreducible representation, then $\Lambda'_1 = \Lambda'^i$ can be written as follows:

$$\Lambda'_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \mathbf{t}'_1 & \mathbf{T}_{\Lambda'_{11}} & \cdots & \mathbf{T}_{\Lambda'_{1,n_1-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{t}'_{n_1-1} & \mathbf{T}_{\Lambda'_{n_1-1,1}} & \cdots & \mathbf{T}_{\Lambda'_{n_1-1,n_1-1}} \end{pmatrix}, \tag{B25}$$

where the matrix elements are arbitrary except for the first line. If $n_1 > 1$, then there exists a matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \oplus \mathbb{I}_{n_1-2}$ that cannot be diagonalized. Thus, n_1 must take 1. Similarly, when $i \geqslant 2$, Λ'_i can take any matrix as Λ' is an arbitrary TP map. If $n_i > 1$ in this case, then we can also find a matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \oplus \mathbb{I}_{n_i-2}$ that cannot be diagonalized. The above arguments indicate that $\forall i$, $n_i = 1$. The irreducible representation decomposition of Pauli-Liouville representation of G must be multiplicity-free to make Λ_G diagonal for any channel Λ. This completes the proof of the first conclusion in the lemma.

A direct corollary of the nonmultiplicity condition of G in Liouville representation is the cardinality of the twirling group $|\mathsf{G}| \geqslant 4^n$. This can be obtained by the Burnside theorem, as follows:

$$\begin{aligned} |\mathsf{G}| &= \sum_{\phi_i\in R_G}\dim^2\phi_i \\ &\geqslant \sum_{i=1}^{k}\dim^2\phi_i \\ &\geqslant \sum_{i=1}^{k}\dim\phi_i \\ &= 4^n, \end{aligned} \tag{B26}$$

where R_G records all inequivalent irreducible representations of G. ■

3. Systematic twirling group construction for generic quantum gates

Before we go through the proof of the main theorem, we discuss how to construct twirling groups for generic quantum gates satisfying the conditions in Question 1.

Notice that if G contains a subgroup H that would make arbitrary channels diagonal, up to a unitary transformation,

via twirling, then G would also enjoy this property as $\Lambda_G = (\Lambda_G)_H$, which is shown below.

Lemma 13. If G contains a subgroup H that would twirl any noise channel into a diagonal channel, up to a unitary transformation, in the Pauli-Liouville representation, then G would also enjoy this property. Mathematically,

$$\mathsf{H} \subset \mathsf{G}, \forall \Lambda,\ \Lambda_H \text{ is diagonal} \Rightarrow \forall \Lambda,\ \Lambda_G \text{ is diagonal}. \quad \text{(B27)}$$

Proof.

$$\begin{aligned}(\Lambda_G)_H &= \mathbb{E}_{G_h \in \mathsf{H}} \mathbb{E}_{G \in \mathsf{G}} \mathcal{G}_h \mathcal{G} \Lambda (\mathcal{G}_h \mathcal{G})^\dagger \\ &= \mathbb{E}_{G \in \mathsf{G}} \mathcal{G} \Lambda \mathcal{G}^\dagger \\ &= \Lambda_G. \end{aligned} \quad \text{(B28)}$$

Proof is done. ■

As Pauli group P_n can twirl any channel into a Pauli channel, a simple solution to Question 1 is just the smallest group containing Pauli group and normalized by target U, which is constructed and spanned by continuously applying U on P_n until no new element is generated. It is worth noting that, except for Pauli group P_n, local dihedral group $\mathsf{D}_n^m = \langle X, Z_m \rangle^{\otimes n}$ can also twirl any channel into a Pauli channel. In fact, P_n is a special case of D_n^m when $m = 2$. For specific target gate U like Z_m, substituting P_n with D_n^m may lead to a smaller twirling group G though D_n^m is in general larger than P_n. In reality, one can select an optimal m to obtain a better choice of G. The above discussions can be summarized below.

Corollary 1. If n-qubit Pauli group $\mathsf{P}_n \subseteq \mathsf{G}$, then Λ_G is a Pauli channel and is diagonal in the Pauli-Liouville representation.

Corollary 2. If n-qubit local dihedral group $\mathsf{D}_n^m \subseteq \mathsf{G}$, then Λ_G is a Pauli channel and is diagonal in the Pauli-Liouville representation.

Example 1 (Simple twirling group construction). Given an n-qubit unitary, U, a simple solution to Question 1 is the smallest group containing D_n^m and normalized by U where $m \geqslant 2$ is a positive integer. Concretely, the twirling group G can be constructed as follows:

$$\begin{aligned}\mathsf{G} &= \left\langle \bigcup_{l \in \mathbb{N}} U^l \mathsf{D}_n^m (U^\dagger)^l \right\rangle \\ &= \left\{ (U^{l_1} D_1 (U^\dagger)^{l_1})^{k_1} (U^{l_2} D_2 (U^\dagger)^{l_2})^{k_2} \cdots, \right. \\ &\quad \left. \forall i, l_i, k_i \in \mathbb{Z}, D_i \in \mathsf{D}_n^m \right\} \\ &= \left\{ U^{l_1} D_1 (U^\dagger)^{l_1} U^{l_2} D_2 (U^\dagger)^{l_2} \cdots, \forall i, l_i \in \mathbb{Z}, D_i \in \mathsf{D}_n^m \right\}. \end{aligned} \quad \text{(B29)}$$

Here, D_n^m is the n-qubit local dihedral group $\langle X, Z_m \rangle^{\otimes n}$. In practice, we select an optimal m to make $|\mathsf{G}|$ as small as possible.

4. Proof of Lemma 2

Below, we focus on the case that the twirling group is a CRU subgroup and prove Lemma 2 and Theorem 1 in the main text. We first provide a more formal and more mathematical version of Lemma 2.

Lemma 2 (Formal Version). For a finite n-qubit CRU subgroup G, set the Z-basis subgroup of G as $\mathsf{G}_Z = \{U \in \mathsf{G} | U \text{ is Z basis}\}$, which is a normal subgroup of G by Lemma 9. If for any quantum channel Λ, its twirling over group G, Λ_G, is diagonal up to a unitary transformation in the Pauli-Liouville representation, then the quotient group G/G_Z can interchange any two computational basis states. In another word, G/G_Z contains set $\mathsf{S} = \{\Pi_{\mathbf{i}} | \Pi_{\mathbf{i}} = \Pi_{\mathbf{i}}^X \Pi_{\mathbf{i}}^C, \mathbf{i} = i_1 i_2 ... i_n \in \{0,1\}^n, \Pi_{\mathbf{i}} = X_1^{i_1} X_2^{i_2} ... X_n^{i_n} \in \mathsf{X}, \Pi_{\mathbf{i}}^C \in \mathsf{C}^{[n-1]}\mathsf{X}\}$. Here, $\mathsf{X} = \langle X \rangle$ is the group generated by Pauli X gates on all qubits, $\mathsf{C}^{[n-1]}\mathsf{X} = \langle CX, CCX, C^{n-1}X \rangle$ is the group generated by CNOT and multiqubit Toffoli gates on all qubits.

Note that $\langle \cdot \rangle$ denotes the group generated by $\cdot$. Same with before, we simply use $\langle X \rangle$ to represent group $\langle X_1, X_2, \cdots, X_n \rangle$ and $\langle CX, CCX, C^{n-1}X \rangle$ to represent $\langle CX_{12}, \cdots CCX_{123}, \cdots, C^{n-1}X_{1,2,\cdots n} \rangle$ while subscripts label the qubits acted upon. Below we provide the proof of Lemma 2.

Proof. From Lemma 8, we obtain that the space spanned by $\{\mathbb{I}, Z\}^{\otimes n}$ is an invariant subspace of G. By rearranging Pauli operator bases and put $\{\mathbb{I}, Z\}^{\otimes n}$ forward, the Pauli-Liouville representation of G would be in a block-diagonal form

$$\mathcal{G} = \mathcal{G}_Z \bigoplus \mathcal{G}_\perp, \quad \text{(B30)}$$

where $\dim \mathcal{G}_Z = 2^n$ and $\dim \mathcal{G}_\perp = 4^n - 2^n$. Correspondingly, we represent channel Λ in a block-diagonal form in the rearranged basis,

$$\Lambda = \begin{pmatrix} \Lambda_Z & \Lambda_{Z\perp} \\ \Lambda_{\perp Z} & \Lambda_\perp \end{pmatrix}. \quad \text{(B31)}$$

Thus, the twirling of Λ over G is

$$\begin{aligned}\Lambda_G &= \mathbb{E}_{G \in \mathsf{G}} \mathcal{G} \Lambda \mathcal{G}^{-1} \\ &= \mathbb{E}_{G \in \mathsf{G}} \begin{pmatrix} \mathcal{G}_Z \Lambda_Z \mathcal{G}_Z^{-1} & \mathcal{G}_Z \Lambda_{Z\perp} \mathcal{G}_\perp^{-1} \\ \mathcal{G}_\perp \Lambda_{\perp Z} \mathcal{G}_Z^{-1} & \mathcal{G}_\perp \Lambda_\perp \mathcal{G}_\perp^{-1} \end{pmatrix}. \end{aligned} \quad \text{(B32)}$$

Due to the direct sum structure, the irreducible representation decomposition of $\mathcal{G}$ is composed of decompositions of $\mathcal{G}_Z$ and $\mathcal{G}_\perp$. From Lemma 1, all of the irreducible representations contained in $\mathcal{G}$ must be inequivalent to make Λ_G diagonal. Therefore, $\mathcal{G}_Z$ and $\mathcal{G}_\perp$ do not have any equivalent irreducible representation in common, which leads to $\mathbb{E}_{G \in \mathsf{G}} \mathcal{G}_Z \Lambda_{Z\perp} \mathcal{G}_\perp^{-1} = 0$, $\mathbb{E}_{G \in \mathsf{G}} \mathcal{G}_\perp \Lambda_{\perp Z} \mathcal{G}_Z^{-1} = 0$, and $\Lambda_G = (\mathbb{E}_{G \in \mathsf{G}} \mathcal{G}_Z \Lambda_Z \mathcal{G}_Z^{-1}) \oplus (\mathbb{E}_{G \in \mathsf{G}} \mathcal{G}_\perp \Lambda_\perp \mathcal{G}_\perp^{-1})$. To make Λ_G diagonal, we require that two blocks in it are both diagonal. Below, we focus on the first block.

Note that G_Z is a normal subgroup of G. Consider its coset, or quotient group $\mathsf{G}/\mathsf{G}_Z = \{\Pi_i \mathsf{G}_Z, 1 \leqslant i \leqslant |\mathsf{G}/\mathsf{G}_Z|\}$, where $\forall i$, Π_i is a representative element. Without loss of generality, we set Π_1 as $\mathbb{I}_{2^n}$. Then the other representative elements must be outside G_Z. Then we separate the twirling of Λ_Z over G into two parts. One is the twirling over G_Z, and the other is the twirling over quotient group G/G_Z. As Z-basis gates are all identity in the basis of $\{\mathbb{I}, Z\}^{\otimes n}$ in Liouville representation, they do not influence the twirled channel in that subspace. For instance, the Liouville representation of the CS gate is shown in Fig. 7. In the subspace spanned by $\{\mathbb{I}, Z_1, Z_2, Z_1 Z_2\}$, CS gate is equal to identity. Certainly, if the CS gate is a twirling gate, then it contributes nothing to the twirling in this subspace. The same are the gates in G_Z.

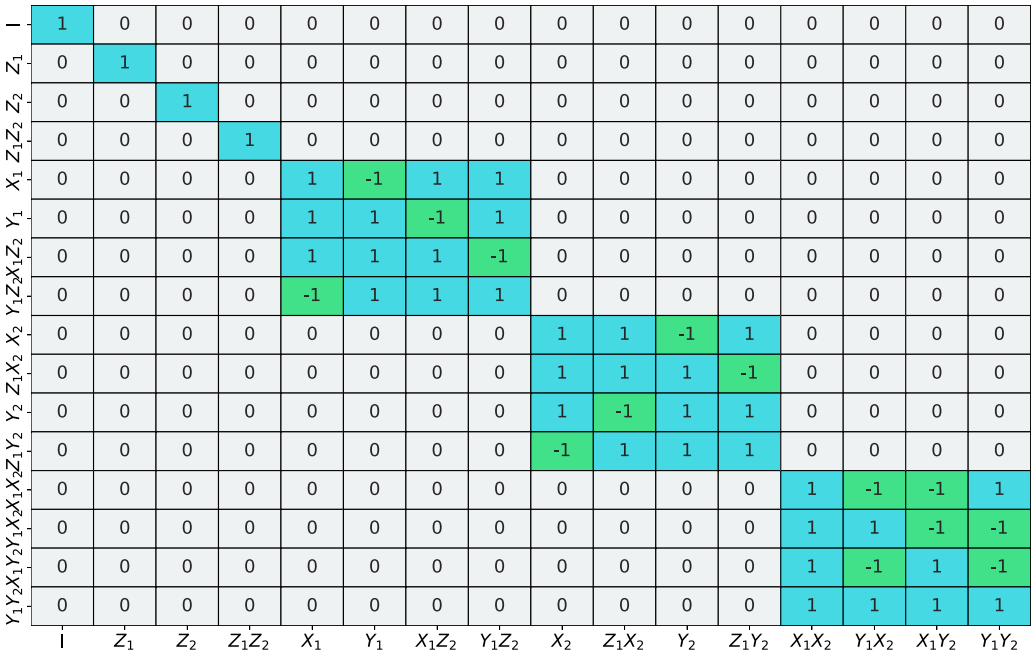


FIG. 7. Pauli Liouville representation of CS gate.

Thus, the only contribution for twirling in this subspace comes from the quotient group. With the similar arguments in Lemma 1, we can directly obtain that $|\mathsf{G}/\mathsf{G}_Z| \geqslant \dim \Lambda_Z = 2^n$, the result in Corollary 3. Specifically,

$$\begin{aligned}\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}_Z\Lambda_Z\mathcal{G}_Z^{-1} &= \mathbb{E}_{1\leqslant i\leqslant|\mathsf{G}/\mathsf{G}_Z|}\Pi_{iZ}\big(\mathbb{E}_{W\in\mathsf{G}_Z}\mathcal{W}_Z\Lambda_Z\mathcal{W}_Z^{-1}\big)\Pi_{iZ}^{-1}\\ &= \mathbb{E}_{1\leqslant i\leqslant|\mathsf{G}/\mathsf{G}_Z|}\Pi_{iZ}\Lambda_Z\Pi_{iZ}^{-1},\end{aligned} \tag{B33}$$

where the subscript Z denotes the sub-representation of Pauli-Liouville representation in $\{\mathbb{I}, Z\}^{\otimes n}$. It can be verified that $\{\Pi_{iZ}, 1 \leqslant i \leqslant |\mathsf{G}/\mathsf{G}_Z|\}$ is a representation of quotient group G/G_Z. Therefore, one can decompose Π_{iZ} with irreducible representations of G/G_Z,

$$\Pi_{iZ} = \bigoplus_{j=1}^{k}\phi_j(\Pi_i). \tag{B34}$$

The twirled channel $\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}_Z\Lambda_Z\mathcal{G}_Z^{-1}$ would be block diagonal corresponding to the irreducible representation decomposition of Π_{iZ}. With the same arguments in Lemma 1, Π_{iZ} must be multiplicity-free. Then

$$\begin{aligned}|\mathsf{G}/\mathsf{G}_Z| &= \sum_{j\in R_{G/Z}}\dim^2\phi_j\\ &\geqslant \sum_{j=1}^{k}\dim^2\phi_j\\ &\geqslant \sum_{j=1}^{k}\dim\phi_j\\ &= 2^n,\end{aligned} \tag{B35}$$

where $R_{G/Z}$ records all inequivalent irreducible representations of G/G_Z.

To obtain the result in Lemma 2, we further study the irreducible representation decomposition of quotient group G/G_Z in the space spanned by $\{\mathbb{I}, Z\}^{\otimes n}$. As mentioned before, Π_{iZ} must be multiplicity-free, which means each irreducible representation can appear at most once in Π_{iZ}. Focusing on the trivial irreducible representation, its multiplicity in Π_{iZ} can be obtained via Lemma 3 and is given by

$$m_t = \mathbb{E}_{\Pi\in\mathsf{G}/\mathsf{G}_Z}1\cdot\mathrm{tr}(\Pi_Z), \tag{B36}$$

where 1 and $\mathrm{tr}(\Pi_Z)$ are characters of trivial irreducible representation and representation in space $\{\mathbb{I}, Z\}^{\otimes n}$, respectively. Through direct calculation, Eq. (B36) can be simplified to

$$\begin{aligned}m_t &= \mathbb{E}_\Pi\frac{1}{2^n}\sum_{W\in\{\mathbb{I},\mathbb{Z}^{\otimes n}\}}\mathrm{tr}(W\Pi W\Pi^\dagger)\\ &= \frac{1}{2^n}\mathbb{E}_\Pi\sum_{i_1,i_2,\cdots,i_n=0}^{1}\mathrm{tr}\left(\left(\prod_{j=1}^{n}Z_j^{i_j}\right)\Pi\left(\prod_{j=1}^{n}Z_j^{i_j}\right)\Pi^\dagger\right)\\ &= \frac{1}{2^n}\mathbb{E}_\Pi\sum_{i_1,i_2,\cdots,i_n=0}^{1}\mathrm{tr}\left(\left(\prod_{j=1}^{n}Z_j^{i_j}\right)\left(\prod_{j=1}^{n}\Pi Z_j^{i_j}\Pi^\dagger\right)\right)\\ &= \frac{1}{2^n}\mathbb{E}_\Pi\sum_{i_1,i_2,\cdots,i_n=0}^{1}\mathrm{tr}\left(\left(\prod_{j=1}^{n}Z_j^{i_j}\right)\left(\prod_{j=1}^{n}\Pi Z_j^{i_j}\Pi^\dagger\right)\right)\\ &= \frac{1}{2^n}\mathbb{E}_\Pi\sum_{i_1,i_2,\cdots,i_n=0}^{1}\mathrm{tr}\left(\left(\prod_{j=1}^{n}Z_j^{i_j}\Pi Z_j^{i_j}\Pi^\dagger\right)\right)\\ &= \frac{1}{2^n}\mathbb{E}_\Pi\mathrm{tr}\left(\left(\prod_{j=1}^{n}\sum_{i_j=0}^{1}Z_j^{i_j}\Pi Z_j^{i_j}\Pi^\dagger\right)\right)\\ &= \frac{1}{2^n}\mathbb{E}_\Pi\mathrm{tr}\left(\prod_{j=1}^{n}(\mathbb{I}+Z_j\Pi Z_j\Pi^\dagger)\right).\end{aligned} \tag{B37}$$

As $\Pi \in \mathsf{G}/\mathsf{G}_Z$ is a permutation matrix, $\Pi Z_j \Pi^\dagger$, $Z_j \Pi Z_j \Pi^\dagger$, and $\mathbb{I} + Z_j \Pi Z_j \Pi^\dagger$ are all diagonal matrices. Therefore, Eq. (B37) can be further simplified to

$$\begin{aligned} m_t &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n} \prod_{j=1}^{n} \langle \mathbf{i}|\mathbb{I} + Z_j \Pi Z_j \Pi^\dagger|\mathbf{i}\rangle \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n} \prod_{j=1}^{n} (1 + \langle \mathbf{i}|Z_j \Pi Z_j \Pi^\dagger|\mathbf{i}\rangle) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n} \prod_{j=1}^{n} (1 + (-1)^{\mathbf{i}_j} \langle \mathbf{i}|\Pi Z_j \Pi^\dagger|\mathbf{i}\rangle) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n} \prod_{j=1}^{n} (1 + (-1)^{\mathbf{i}_j + \pi(\mathbf{i})_j}) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n} 2^n \delta_{\pi(\mathbf{i})=\mathbf{i}} \\ &= \mathbb{E}_\Pi N_\Pi \\ &= O_{\mathsf{G}/\mathsf{G}_Z \to \{0,1\}^n}. \end{aligned} \tag{B38}$$

In the fourth line, we utilize that a permutation matrix would transform a bit string into another bit string and denote $|\pi(\mathbf{i})\rangle = \Pi^\dagger|\mathbf{i}\rangle$. Here, π is a permutation acting on $\{0,1\}^n$ and can be viewed as a representation of Π. In the sixth line, $N_\Pi = \sum_{\mathbf{i}\in\{0,1\}^n} \delta_{\pi(\mathbf{i})=\mathbf{i}}$ denotes the number of fixed points for π acting on $\{0,1\}^n$. Via Burnside's lemma, $\mathbb{E}_\Pi N_\Pi$ equals to the number of orbits for group G/G_Z acting on $\{0,1\}^n$ which we denote as $O_{\mathsf{G}/\mathsf{G}_Z \to \{0,1\}^n}$.

As trivial irreducible representation can appear at most once, m_t cannot be larger than 1, or the number of orbits cannot be larger than 1. It implies that gates of G/G_Z can interchange any two computational states or any two bit strings in $\{0,1\}^n$. Then, we obtain the results in the informal version of Lemma 2.

Now, we obtain that G/G_Z can transform 0^n to any other bit string in $\{0,1\}^n$. Recall that G/G_Z is a subgroup of $\langle X, C^{n-1}X\rangle$, each element in G/G_Z can be written as $\Pi_i^X \Pi_i^C$, where $\Pi_i^X \in \mathsf{X} = \langle X\rangle$, $\Pi_i^C \in \langle CX, CCX, \cdots, C^{n-1}X\rangle$. As any element in $\langle CX, CCX, \cdots, C^{n-1}X\rangle$ has no effect on bit string 0^n, $\Pi_i^X \Pi_i^C$ would simply transform $|0^n\rangle$ to $\Pi_i^X|0^n\rangle$. Thus, Π_i^X must take over all elements in X to transform 0^n to all bit strings in $\{0,1\}^n$. Here, we complete the proof of Lemma 2. ■

In the main text, we have discussed the advantages of choosing the CRU subgroup as a twirling group in virtually implementing the inverse gate. Below, we discuss that it might be sufficient only to consider CRU twirling gates for tailoring diagonal gates. We notice that in most RB protocols, the twirling groups contain a CRU subgroup that can twirl the noise channel diagonal, including Clifford group [6], generalized matchgate group [18], and CNOT dihedral group [16]. We conjecture that if any finite group G can make arbitrary noise channels diagonal via twirling, then under the equivalence of unitary transformation, it has a CRU subgroup G_C also achieving that. A concrete example is G as the Clifford group and G_C as the Pauli group. Thus, for diagonal gate U in the computational basis, considering G in CRU highly likely suffices to find the optimal solution. It is worth noting that the above discussion is only conjecture, and we hope in the future, people can find the smallest group in the whole unitary group for tailoring generic quantum gates.

From Lemma 2, we can directly obtain a lower bound for the cardinality of the quotient group, as shown in the following corollary.

Corollary 3. For a finite n-qubit CRU subgroup G, set the computational basis subgroup of G as $\mathsf{G}_Z = \{U \in \mathsf{G} | U \text{ is computational basis}\}$. By Lemma 8, G_Z is a normal subgroup of G. If for any quantum channel Λ, its twirling over group G, Λ_G, is diagonal up to a unitary transformation in the Pauli-Liouville representation, then the cardinality of the quotient group $|\mathsf{G}/\mathsf{G}_Z| \geqslant 2^n$.

5. Proof of Theorem 1

Below, we provide the proof of Theorem 1. The key point is utilizing that G contains specific permutation gates and G is normalized by target gate U. We first present a lemma, telling us what is the smallest group containing a given permutation group and normalized by a given computational basis diagonal gate U.

Lemma 14. Given a permutation group, Π, and a diagonal gate, U, in the computational basis, the smallest group containing Π and normalized by U is given by

$$\mathsf{G} = \Pi \ltimes \mathsf{W}, \tag{B39}$$

where $\ltimes$ denotes semiproduct and $\mathsf{W} = \langle \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \Pi\}\rangle$.

Proof. We first clarify the meaning of semiproduct. It means that

(1) Any element in $\Pi \ltimes \mathsf{W}$ has an expression of ΠW where $\Pi \in \Pi$ and $W \in \mathsf{W}$.

(2) W is a normal subgroup of $\Pi \ltimes \mathsf{W}$.

It is easy to verify the second condition that W is a normal subgroup with the first condition. We only need to verify that for any Π and Π', $\Pi'^\dagger \Pi^\dagger U \Pi U^\dagger \Pi' \in \mathsf{W}$. This can be seen via the following equation:

$$\Pi'^\dagger \Pi^\dagger U \Pi U^\dagger \Pi' = (\Pi\Pi')^\dagger U (\Pi\Pi') U^\dagger (\Pi'^\dagger U \Pi' U^\dagger)^\dagger. \tag{B40}$$

As $(\Pi\Pi')^\dagger U(\Pi\Pi')U^\dagger$ and $\Pi'^\dagger U \Pi' U^\dagger$ are both elements in W, we successfully show the soundness of the second condition.

Now we turn to the proof of the lemma. Obviously, $\Pi \ltimes \mathsf{W}$ contains Π. It is also normalized by U:

Given an element ΠW where $\Pi \in \Pi$ and $W \in \mathsf{W}$,

$$\begin{aligned} U \Pi W U^\dagger &= U \Pi U^\dagger W \\ &= \Pi(\Pi^\dagger U \Pi U^\dagger) W. \end{aligned} \tag{B41}$$

The first line comes from the fact that W and U are both diagonal gates. As $\Pi^\dagger U \Pi U^\dagger$ and W both belong to W, $U \Pi W U^\dagger$ has an expression of $\Pi W'$ where $W' \in \mathsf{W}$. Combining the condition 1, we obtain that $\Pi \ltimes \mathsf{W}$ is normalized by U.

Meanwhile, for any group G containing Π and normalized by U, $U \Pi U^\dagger \in \mathsf{G}$ and $\Pi^\dagger U \Pi U^\dagger \in \mathsf{G}$. Then we know Π and W both belong to G. As a consequence, G must contain $\Pi \ltimes \mathsf{W}$. Combining with the arguments before, we prove that

$\Pi \ltimes \mathsf{W}$ is the smallest group containing Π and normalized by U. ■

With Lemma 14, we can easily prove Theorem 1 in the main text. For convenience, we rewrite the theorem below.

Theorem 1. The optimal twirling group G in CRU for the multiqubit controlled phase gate, $U = C^n Z_m$, with $n \geqslant 1, m \geqslant 2$, is the smallest group containing X and normalized by U, given by

$$\mathsf{G} = \left\{ \Pi \left(\prod_{i=1}^{t} (\Pi_i^\dagger U \Pi_i U^\dagger)^{l_i} \right) | \Pi \in \mathsf{X}, t \in \mathbb{N}, \right. \\ \left. \forall i, l_i \in \pm 1, \Pi_i \in \mathsf{X} \right\}. \tag{B42}$$

Proof. From Lemma 2, we can deduce that G at least contains permutation group $\langle \mathsf{S} \rangle$ where $\mathsf{S} = \{\Pi_i^X \Pi_i^C, \forall \Pi_i^X \in \mathsf{X}, \exists \Pi_i^C \in \mathsf{C}^{[n-1]}\mathsf{X}\}$. Using Lemma 14, we obtain that any twirling group G as a solution to Question 1 must satisfy

$$\langle \mathsf{S} \rangle \ltimes \langle \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle \mathsf{S} \rangle\} \rangle \leqslant \mathsf{G}. \tag{B43}$$

If $U = C^{n-1} Z_m = \begin{pmatrix} \mathbb{I}_{2^{n-1}} & \mathbf{0} \\ \mathbf{0} & e^{i\frac{2\pi}{m}} \end{pmatrix}$, then we can show that $\langle \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle \mathsf{X} \rangle\} \rangle \leqslant \langle \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle \mathsf{S} \rangle\} \rangle$. For brevity, we denote $\mathsf{W} = \langle \{\Pi'^\dagger U \Pi' U^\dagger, \Pi' \in \langle \mathsf{S} \rangle\} \rangle$ and $\mathsf{W}_X = \langle \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle \mathsf{X} \rangle\} \rangle$. For any generator $\Pi^\dagger U \Pi U^\dagger$ in W_X where $\Pi \in \mathsf{X}$, $\Pi^\dagger U \Pi$ is a diagonal matrix while only one diagonal element is not 1 but equals $e^{i\frac{2\pi}{m}}$. As the permutation elements in S can interchange any two computational bases, there must exist an element $\Pi' \in \mathsf{S}$ such that

$$\Pi'^\dagger U \Pi' = \Pi^\dagger U \Pi. \tag{B44}$$

Then,

$$\Pi'^\dagger U \Pi' U^\dagger = \Pi^\dagger U \Pi U^\dagger. \tag{B45}$$

Thus, any generator of W_X belongs to W. Then, we obtain that $\mathsf{W}_X \leqslant \mathsf{W}$ for $C^{n-1} Z_m$. This result also applies to the diagonal matrix in which only one diagonal element differs from the others.

Note that the smallest group containing X and normalized by U is $\mathsf{X} \ltimes \mathsf{W}_X$. As X is no larger and no global than $\langle \mathsf{S} \rangle$ and W_X is a subset of W, $\mathsf{X} \ltimes \mathsf{W}_X$ is smaller and more local than $\langle \mathsf{S} \rangle \ltimes \langle \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle \mathsf{S} \rangle\} \rangle$, which implies that the twirling group cannot be better than $\mathsf{X} \ltimes \mathsf{W}_X$. Below, we would show that for $U = C^n Z_m$ with $n \geqslant 1, m \geqslant 2$, $\mathsf{X} \ltimes \mathsf{W}_X$ suffices to tailor $C^n Z_m$. Combining the two sides, we would obtain that the optimal twirling group for $C^n Z_m$ is $\mathsf{X} \ltimes \mathsf{W}_X$, just shown in Theorem 1.

In the following, we show $\mathsf{X} \ltimes \mathsf{W}_X$ satisfies Question 1 for multiqubit controlled phase gate $U = C^n Z_m$. From the definition of $\mathsf{X} \ltimes \mathsf{W}_X$, we know that $\mathsf{X} \ltimes \mathsf{W}_X$ is normalized by U. Thus, we only need to verify that $\mathsf{X} \ltimes \mathsf{W}_X$ can make arbitrary channels diagonal via twirling. With corollaries 1 and 2, it is sufficient to show that Pauli Z gate or Z_m gate on each qubit belongs to W_X. Below, we show that this is right for $U = C^n Z_m$ with $n \geqslant 1, m \geqslant 2$.

Recall that $\mathsf{W}_X = \langle \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle \mathsf{X} \rangle\} \rangle$, in the following we analyze the generators of W_X, or $\Pi^\dagger U \Pi U^\dagger$, in detail. For certain $\Pi \in \langle X \rangle$, let us define its pattern to be a 0-1 bit string, s_j^Π, such that $s_j^\Pi = 1$ if the jth qubit of Π is I and $s_j^\Pi = 0$ if the jth qubit of Π is X. Then, the matrix representation of $\Pi^\dagger U \Pi U^\dagger$ for $U = C^n Z_m$ is

$$\Pi^\dagger C^n Z_m \Pi C^n Z_m^\dagger = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & e^{i\frac{2\pi}{m}} & & & \\ & & & \ddots & & \\ & & & & e^{-i\frac{2\pi}{m}} \end{pmatrix}, \tag{B46}$$

where the (s^Π, s^Π) entry is $e^{i\frac{2\pi}{m}}$, the $(2^{n+1}-1, 2^{n+1}-1)$ entry is $e^{-i\frac{2\pi}{m}}$, and other diagonal entries are 1. As generators are all diagonal, all elements in W_X are diagonal in the computational basis. Moreover, the diagonal elements would be power of $e^{i\frac{2\pi}{m}}$. Define an injective map $\phi : \mathsf{W}_X \to \mathbb{Z}_m^{2^{n+1}}$ where $\mathbb{Z}_m = \{0, 1, \cdots, m-1\}$,

$$\phi(w) = -i\frac{m}{2\pi} \begin{pmatrix} \ln[w]_{0,0} \\ \vdots \\ \ln[w]_{2^{n+1}-1, 2^{n+1}-1} \end{pmatrix}. \tag{B47}$$

Note that the entries of $\phi(w)$ will always belong to $\{0, 1, \cdots, m-1\}$. With map ϕ, we express the gate in W_X with a vector. For example, $\phi(\Pi^\dagger C^n Z_m \Pi C^n Z_m^\dagger) = e_{s^\Pi} - e_{2^{n+1}-1}$, where we define e_i to be the basis vector whose ith entry is 1 and other entries are 0. In this way, the group multiplication is turned into integer vector addition, and the group generation is equivalent to the linear combination of vectors. As Π can take any element in X, vectors $v_0 = e_0 - e_{2^{n+1}-1}, \cdots, v_{2^{n+1}-2} = e_{2^{n+1}-2} - e_{2^{n+1}-1}$, are all basis vectors. Note that the overall phase of a quantum gate is not important, so any vectors differing by multiples of $v_{2^{n+1}-1} = (1 \quad \cdots \quad 1)^T$ are equivalent. Next, we show how to construct vectors associated with Z or Z_m gates with vectors $v_0, \cdots, v_{2^{n+1}-1}$.

As different qubits are symmetric under arbitrary permutation, we only need to construct Z gate or Z_m gate on the first qubit. If m is odd, then we show $Z_m \in \mathsf{W}_X$. Gate Z_m on first qubit corresponds to vector $\phi((Z_m)_1) = (0 \quad \cdots \quad 0 \quad 1 \quad \cdots \quad 1)^T$. Define

$$u = -(v_0 + v_1 + \cdots + v_{2^{n+1}-2}) + (2^{n+1}-1)v_{2^{n+1}-1} \\ = (0 \quad \cdots \quad 2^{n+1})^T. \tag{B48}$$

When m is odd, $\gcd(2^{n+1}, m) = 1$, vector u is equivalent to $(0 \quad \cdots \quad 1)^T = e_{2^{n+1}-1}$. Then, we can also obtain $e_i = v_i + u$ for $0 \leqslant i \leqslant 2^{n+1} - 2$. With e_i for $0 \leqslant i \leqslant 2^{n+1} - 1$, $\phi((Z_m)_1)$ can certainly be constructed by

$$\phi((Z_m)_1) = e_{2^{n+1}-2^n} + e_{2^{n+1}-2^{n+1}} + \cdots + e_{2^{n+1}-1}. \tag{B49}$$

Thus, if m is odd, then we can ensure that phase gates Z_m on all qubits belong to W_X. In this case, $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ definitely satisfies the requirements in Question 1.

When m is even, we show $Z \in \mathsf{W}_X$. We first express $m = q2^k$ where $q \geqslant 1$ is odd and k is a positive integer. Gate Z on first qubit corresponds to vector $\phi(Z_1) = (0 \quad \cdots \quad 0 \quad q2^{k-1} \quad \cdots \quad q2^{k-1})^T$. As $\gcd(2^{n+1}, m) = 2^{\min(n+1,k)}$, u can be expressed as $(0 \quad \cdots \quad 2^{\min(n+1,k)})^T$.

Then, $\phi(Z_1)$ can be constructed by

$$\begin{aligned}\phi(Z_1) &= q2^{k-1}(e_{2^{n+1}-2^n} + e_{2^{n+1}-2^n+1} + \cdots + e_{2^{n+1}-1}) \\ &= q2^{k-1}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{k-1}(2^n-1)e_{2^{n+1}-1} + q2^{k-1}e_{2^{n+1}-1} \\ &= q2^{k-1}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{n+k-1}e_{2^{n+1}-1} \\ &= q2^{k-1}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{n+k-1-\min(n+1,k)}u. \end{aligned} \tag{B50}$$

As $n \geqslant 1$, $n+k-1-\min(n+1,k) = \max(n-1,k-2)$ is always a nonnegative integer. Thus, $q2^{n+k-1-\min(n+1,k)}$ is an integer. As a consequence, $\phi(Z_1)$ can be constructed with linear combination of $v_0, \cdots, v_{2^{n+1}-2}$, and u, which is equivalent to $Z_1 \in \mathsf{W}_X$. Thus, if m is even, then we can ensure that Pauli Z gates on all qubits belong to W_X. In this case, $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ also satisfies the requirements in Question 1. Proof is done. ■

6. Twirling groups for multiqubit controlled phase gates

Below, we prove the following theorem, showing the twirling groups constructed for multiqubit controlled phase gates in the previous subsection not only satisfies $U\mathsf{G}U^\dagger = \mathsf{G}$, but also satisfies $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$.

Theorem 2. For $U = C^nZ_m$ with $n \geqslant 1, m \geqslant 2$ and $nm \neq 2$, the twirling group $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ satisfies $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$.

For such quantum gates, our protocol can provide their fidelity estimation accurately instead of providing lower bounds. To prove Theorem 2, we present the following lemma.

Lemma 15. As long as $\langle X, CZ, S\rangle \leqslant \mathsf{G}$ and U is diagonal in the computational basis, the equation $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ holds.

Proof. In Pauli-Liouville representation, $\mathcal{U}$ is block-diagonal. Below, we investigate the diagonal blocks of $\mathcal{U}$. By definition,

$$\begin{aligned}\mathcal{U}_{ij} &= \mathrm{tr}(\sigma_i U(\sigma_j)) \\ &= d\,\mathrm{tr}(\sigma_i U\sigma_j U^\dagger\sigma_j^\dagger\sigma_j) \\ &= d\,\mathrm{tr}(\sigma_j\sigma_i U\sigma_j U^\dagger\sigma_j^\dagger). \end{aligned} \tag{B51}$$

As U is diagonal in the computational basis or Z basis and σ_j is proportional to a Pauli operator, the right part $U\sigma_jU^\dagger\sigma_j^\dagger$ is also diagonal in the computational basis. Thus, when the left part $\sigma_j\sigma_i$ is not diagonal, the trace of the above expression is 0. In general, we can express that $\sigma_i = \frac{1}{\sqrt{d}}X_{\mathbf{a}_i}Z_{\mathbf{b}_i}$ and $\sigma_j = \frac{1}{\sqrt{d}}X_{\mathbf{a}_j}Z_{\mathbf{b}_j}$ where $X_{\mathbf{a}_i}, X_{\mathbf{a}_j} \in \{\mathbb{I}, X\}^{\otimes n}$ and $Z_{\mathbf{b}_i}, Z_{\mathbf{b}_j} \in \{\mathbb{I}, Z\}^{\otimes n}$. The notation $\mathbf{a}_i$ is a bit string from $\{0,1\}^n$ meaning $X_{\mathbf{a}_i} = \prod_{k=1}^n X_k^{(\mathbf{a}_i)_k}$. The same are for $\mathbf{a}_j$, $\mathbf{b}_i$, and $\mathbf{b}_j$. If $X_{\mathbf{a}_i} \neq X_{\mathbf{a}_j}$, then $\sigma_j\sigma_i$ is not diagonal and hence $\mathcal{U}_{ij} = 0$. Thus, for any operator $X_{\mathbf{a}} \in \{\mathbb{I}, X\}^{\otimes n}$, there is a corresponding block with dimension $2^n \times 2^n$ in the Liouville representation of the diagonal matrix U. The linear space associated with the block is spanned by bases $\{X_{\mathbf{a}} \cdot I^{\otimes n}, X_{\mathbf{a}} \cdot (I^{\otimes n-1} \otimes Z), \cdots, X_{\mathbf{a}} \cdot Z^{\otimes n}\}$ and we denote the set of this bases as $X_{\mathbf{a}}\mathsf{Z}$. It is worth noting that the block spanned by Z is equal to identity with dimension 2^n, which means this block can be further decomposed into the direct sum of 2^n small blocks with dimensions 1.

Recall that Λ_G is also diagonal in Liouville representation, and all of its blocks are proportional to identities. If Λ_G has the same blocks as $\mathcal{U}$, then they will commute with each other and hence $\mathcal{U}\Lambda_G\mathcal{U}^{-1} = \Lambda_G$ holds. Now we show that $\langle X, CZ, S\rangle \leqslant \mathbf{G}$ is enough for Λ_G to have the same blocks as $\mathcal{U}$. Since $\langle X, Z\rangle \leqslant \langle X, CZ, S\rangle \leqslant \mathsf{G}$, Λ_G must be diagonal in Pauli-Liouville representation, $\Lambda_G = \sum_i \lambda_i |\sigma_i\rangle\rangle\langle\langle\sigma_i|$. Note that if there exists an element $G \in \mathsf{G}$ such that $\sigma_i = G\sigma_jG^\dagger$, then σ_i and σ_j are symmetric under the group action of G, which results in $\lambda_i = \lambda_j$, or equivalently, σ_i and σ_j are in the same diagonal block [16]. Note that any $G \in \langle CZ, S\rangle \leqslant \mathsf{G}$ can be decomposed as $G = WV$ where $W \in \langle CZ\rangle, V \in \langle S\rangle$, then for any $X_{\mathbf{a}} \in \{\mathbb{I}, X\}^{\otimes n}/\mathbb{I}^{\otimes n}$,

$$\begin{aligned}GX_{\mathbf{a}}G^\dagger &= WVX_{\mathbf{a}}V^\dagger W^\dagger \\ &= X_{\mathbf{a}}(X_{\mathbf{a}}^\dagger W X_{\mathbf{a}})(X_{\mathbf{a}}^\dagger V X_{\mathbf{a}})V^\dagger W^\dagger \\ &= X_{\mathbf{a}}(X_{\mathbf{a}}^\dagger W X_{\mathbf{a}} W^\dagger)(X_{\mathbf{a}}^\dagger V X_{\mathbf{a}} V^\dagger). \end{aligned} \tag{B52}$$

Below we express $X_{\mathbf{a}} = \bigotimes_{i=1}^n X_i^{a_i}$ where $a_i \in \{0,1\}$ and we simply denote $\mathbf{a} = (a_1, a_2, \cdots, a_n)^T \in \{0,1\}^n$. As $X_{\mathbf{a}}$ is not equal to identity $\mathbb{I}^{\otimes n}$, there is at least an element in $\{a_1, a_2, \cdots, a_n\}$ to be nonzero. Suppose $a_i = 1$. For any $Z_{\mathbf{b}} \in \{\mathbb{I}, Z\}^{\otimes n}$ where $\mathbf{b} = (b_1, b_2, \cdots, b_n)^T \in \{0,1\}^n$, we can choose $W = \prod_{1\leqslant j\leqslant n, j\neq i} CZ_{ij}^{b_j}$, $V = S_i^{\sum_{1\leqslant j\leqslant n} a_jb_j}$ such that

$$\begin{aligned}&X_{\mathbf{a}}(X_{\mathbf{a}}^\dagger W X_{\mathbf{a}} W^\dagger)(X_{\mathbf{a}}^\dagger V X_{\mathbf{a}} V^\dagger) \\ &= X_{\mathbf{a}}\left(\prod_{1\leqslant j\leqslant n, j\neq i} Z_j^{b_j}Z_i^{a_jb_j}\right)(Z_i^{\sum_{1\leqslant j\leqslant n} a_jb_j}) \\ &= X_{\mathbf{a}}\left(\prod_{1\leqslant j\leqslant n, j\neq i} Z_j^{b_j}\right)(Z_i^{a_ib_i+2\sum_{1\leqslant j\leqslant n, j\neq i} a_jb_j}) \\ &= X_{\mathbf{a}}\prod_{1\leqslant j\leqslant n} Z_j^{b_j} \\ &= X_{\mathbf{a}}Z_{\mathbf{b}}. \end{aligned} \tag{B53}$$

Here we utilize the following two identities:

$$X_iCZ_{i,j}X_iCZ_{i,j}^\dagger = Z_j, \tag{B54}$$

$$X_iS_iX_iS_i^\dagger = Z_i. \tag{B55}$$

Then, for any $X_{\mathbf{a}} \in \{\mathbb{I}, X\}^{\otimes n}/\mathbb{I}^{\otimes n}$ and $Z_{\mathbf{b}} \in \{\mathbb{I}, Z\}^{\otimes n}$, we can find a unitary $G = WV$, such that

$$GX_{\mathbf{a}}G^\dagger = X_{\mathbf{a}}Z_{\mathbf{b}}. \tag{B56}$$

It means that the linear space spanned by $X_{\mathbf{a}}\mathsf{Z}$ is a diagonal block for Λ_G, which is the same as $\mathcal{U}$. In the space spanned by Z, Λ_G and $\mathcal{U}$ are both fully diagonal with 2^n one-dimensional blocks. Thus, we prove that $\langle X, CZ, S\rangle \leqslant \mathsf{G}$ suffices to make Λ_G commute with the diagonal matrix $\mathcal{U}$ and complete the proof of Lemma 15. ■

Now we present the proof of Theorem 2.

Proof. The situation of $m = 2$ and that of $m \geqslant 3$ are different. Below, we first discuss the case that $m \geqslant 3$. The main technique is following the discussion in the proof of Theorem 1 and mapping W_X to $\mathbb{Z}_m^{2^{n+1}}$ via injective map ϕ in Eq. (B47).

Suppose $m \geqslant 3$. In the case that m is odd for $U = C^n Z_m$, from the proof of Theorem 1, we obtain that $\phi(\mathsf{W}_X)$ is spanned by $\{e_i, 0 \leqslant i \leqslant 2^{n+1} - 1\}$. It means that $\phi(\mathsf{W}_X)$ is equal to $\mathbb{Z}_m^{2^{n+1}}$. Note that $\phi(C^n Z_m) = e_{2^{n+1}-1}$, so in this case the target gate U itself belongs to W_X and hence U belongs to the twirling group $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$. When $U \in \mathsf{G}$, then

$$\begin{aligned}\mathcal{U}\Lambda_G\mathcal{U}^\dagger &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{U}\mathcal{G}\Lambda\mathcal{G}^\dagger\mathcal{U}^\dagger \\ &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \\ &= \Lambda_G. \end{aligned} \tag{B57}$$

The commutation between $\mathcal{U}$ and Λ_G is satisfied automatically.

Below, we discuss the case that m is even and $m \geqslant 3$ for $U = C^n Z_m$. Based on Lemma 15, we study whether CZ and S gates belong to W_X or not. By symmetry, we only need to study that for the phase gate on the first qubit S_1 and the controlled-phase gate on the first two qubits CZ_{12}. Note that the phase gate S can be expressed as $S = Z_m^{m/4}$ and CZ can be expressed as $CZ = CZ_m^{m/2}$. We express $m = q2^k$ where $q \geqslant 1$ is odd and $k \geqslant 1$ is a positive integer. Note that the bases of $\phi(\mathsf{W}_X)$ are $\{v_i = e_i - e_{2^{n+1}-1}, 0 \leqslant i \leqslant 2^{n+1} - 2\}$ along with $u = 2^{\min(n+1,k)} e_{2^{n+1}-1}$. Then, $\phi(S_1)$ can be constructed by

$$\begin{aligned}\phi(S_1) &= q2^{k-2}(e_{2^{n+1}-2^n} + e_{2^{n+1}-2^n+1} + \cdots + e_{2^{n+1}-1}) \\ &= q2^{k-2}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{k-2}(2^n - 1)e_{2^{n+1}-1} + q2^{k-2}e_{2^{n+1}-1} \\ &= q2^{k-2}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{n+k-2}e_{2^{n+1}-1} \\ &= q2^{k-2}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{n+k-2-\min(n+1,k)}u \\ &= q2^{k-2}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{\max(k-3,n-2)}u. \end{aligned} \tag{B58}$$

Meanwhile, $\phi(CZ_{12})$ can be constructed by

$$\begin{aligned}\phi(CZ_{12}) &= q2^{k-1}(e_{2^{n+1}-2^{n-1}} + e_{2^{n+1}-2^{n-1}+1} + \cdots + e_{2^{n+1}-1}) \\ &= q2^{k-1}(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{k-1}(2^{n-1} - 1)e_{2^{n+1}-1} + q2^{k-1}e_{2^{n+1}-1} \\ &= q2^{k-1}(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{n+k-2}e_{2^{n+1}-1} \\ &= q2^{k-1}(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{n+k-2-\min(n+1,k)}u \\ &= q2^{k-1}(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + q2^{\max(k-3,n-2)}u. \end{aligned} \tag{B59}$$

The constructions Eq. (B58) and Eq. (B59) are valid only when the coefficients of the bases are integers. Obviously, $k \geqslant 3$, and $k = 2$ and $n \geqslant 2$ can make the coefficients to be integers. In this case, S_1 and CZ_{12} belong to W_X and furthermore, $\langle CZ, X, S\rangle \leqslant \mathsf{G}$. Based on Lemma 15, $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ is satisfied. The exceptions are the case that $k = 1$ and that $n = 1$ and $k = 2$. Below, we discuss the two kinds of exceptions in detail.

When $k = 1$, the target gate $U = C^n Z_m$ and $m = 2q$. In this case, $u = 2e_{2^{n+1}-1}$, and $CZ_{m/2}$ and Z_m gates belong to G, which can be derived through the following equations:

$$\begin{aligned}\phi((Z_m)_1) &= e_{2^{n+1}-2^n} + e_{2^{n+1}-2^n+1} + \cdots + e_{2^{n+1}-1} \\ &= v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2} \\ &\quad + (2^n - 1)e_{2^{n+1}-1} + e_{2^{n+1}-1} \\ &= v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2} + 2^n e_{2^{n+1}-1} \\ &= v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2} + 2^{n-1}u, \end{aligned} \tag{B60}$$

$$\begin{aligned}\phi((CZ_{m/2})_{12}) &= 2(e_{2^{n+1}-2^{n-1}} + e_{2^{n+1}-2^{n-1}+1} + \cdots + e_{2^{n+1}-1}) \\ &= 2(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + 2(2^{n-1} - 1)e_{2^{n+1}-1} + 2e_{2^{n+1}-1} \\ &= 2(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + 2^n e_{2^{n+1}-1} \\ &= 2(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) \\ &\quad + 2^{n-1}u. \end{aligned} \tag{B61}$$

Thus, when $k = 1$, $\langle CZ_{m/2}, Z_m\rangle \leqslant \mathsf{W}_X$ and $\mathsf{X} \ltimes \langle CZ_{m/2}, Z_m\rangle \leqslant \mathsf{G}$. Notice that $m \geqslant 3$ and $m/2$ is odd, so $m/2 \geqslant 3$ holds. It can be verified that the irreducible representation decomposition of the Liouville representation for group $\langle CZ_m, Z_m, X\rangle$, where $m \geqslant 3$ and m is odd, and group $\langle CZ, S, X\rangle$ are the same. As $\mathsf{X} \ltimes \langle CZ_{m/2}, Z_{m/2}\rangle \leqslant \mathsf{X} \ltimes \langle CZ_{m/2}, Z_m\rangle \leqslant \mathsf{G}$ and $m/2 \geqslant 3$, after being twirled by G, the twirled noise channel Λ_G would have the same blocks as $\mathcal{U}$, and $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ is satisfied.

When $n = 1$ and $k = 2$, the target gate $U = CZ_m$ and $m = 4q$. The bases for $\phi(\mathsf{W}_X)$ are $v_0 = (1, 0, 0, -1)^T$, $v_1 = (0, 1, 0, -1)^T$, $v_2 = (0, 0, 1, -1)^T$, and $u = (0, 0, 0, 4)$. Note that $\phi(CZ \cdot S_1) = (0, 0, q, 3q)^T = v_2 + u$ and $\phi(CZ \cdot S_2) = (0, q, 0, 3q)^T = v_1 + u$. Thus, in this case, $\langle CZ \cdot S_1, CZ \cdot S_2\rangle \in \mathsf{W}_X \leqslant \mathsf{G}$. In the proof of Theorem 1, we have already shown that Pauli group $\mathsf{P}_2 \in \mathsf{G}$. Thus, the twirled noise channel would be diagonal in Liouville representation, $\Lambda_G = \sum_i \lambda_i |\sigma_i\rangle\rangle\langle\langle\sigma_i|$. Also, we have the following identities:

$$CZ \cdot S_1 X_1 (CZ \cdot S_1)^\dagger = X_1 Z_1 Z_2, \tag{B62}$$

$$CZ \cdot S_2 X_1 (CZ \cdot S_2)^\dagger = X_1 Z_2, \tag{B63}$$

$$(CZ \cdot S_2 \cdot CZ \cdot S_1) X_1 (CZ \cdot S_2 \cdot CZ \cdot S_1)^\dagger = X_1 Z_1. \tag{B64}$$

Under the twirling of $\langle CZ \cdot S_1, CZ \cdot S_2\rangle$, X_1, X_1Z_1, X_1Z_2, and $X_1Z_1Z_2$ would be symmetric. Their corresponding Pauli fidelities λ_i would be the same after the twirling. As the twirling group contains $\langle CZ \cdot S_1, CZ \cdot S_2\rangle$, the twirled noise channel Λ_G would be diagonal and proportional to identity in the space spanned by $X_1\mathsf{Z}$. The cases for spaces spanned by $X_2\mathsf{Z}$ or $X_1X_2\mathsf{Z}$ are the same. In summary, Λ_G has diagonal blocks in spaces spanned by $X_1\mathsf{Z}$, $X_2\mathsf{Z}$ and $X_1X_2\mathsf{Z}$, respectively. Thus, in this case, $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ is also satisfied.

At last, we analyze the case that $m = 2$ and $n \geqslant 2$. The case that $m = 2$ and $n = 1$ is just CZ and obviously in this

case $\mathcal{U}\Lambda_G\mathcal{U}^\dagger \neq \Lambda_G$. When the target gate is C^nZ with $n \geqslant 2$, the twirling group is $\mathsf{G} = X \ltimes W_X = \langle C^{n-1}Z, \cdots, Z, X\rangle$. Different from a generic diagonal gate, the diagonal blocks spanned by $X_{\mathbf{a}}\mathsf{Z}$ of C^nZ gate in Pauli-Liouville representation can be further divided into two small blocks. We will show that any channel twirled by group $\langle CZ, Z, X\rangle$ would have the same diagonal blocks as C^nZ in Liouville representation. As $\langle C^{n-1}Z, \cdots, Z, X\rangle$ contains $\langle CZ, Z, X\rangle$, $\Lambda_{\langle C^{n-1}Z,\cdots,Z,X\rangle}$ would commute with C^nZ.

Similar to Eq. (B51), by setting $\sigma_i = \frac{1}{\sqrt{d}}X_{\mathbf{a}_i}Z_{\mathbf{b}_i}$ and $\sigma_j = \frac{1}{\sqrt{d}}X_{\mathbf{a}_j}Z_{\mathbf{b}_j}$ where $X_{\mathbf{a}_i}, X_{\mathbf{a}_j} \in \{\mathbb{I}, X\}^{\otimes n}$ and $Z_{\mathbf{b}_i}, Z_{\mathbf{b}_j} \in \{\mathbb{I}, Z\}^{\otimes n}$, we get the matrix element of the Liouville representation of $U = C^nZ$,

$$\begin{aligned}\mathcal{U}_{ji} &= d\,\mathrm{tr}\left(\sigma_i\sigma_j C^nZ\sigma_i C^nZ\sigma_i^\dagger\right)\\ &= \frac{1}{d}\mathrm{tr}\left(X_{\mathbf{a}_i}Z_{\mathbf{b}_i}X_{\mathbf{a}_j}Z_{\mathbf{b}_j}W\right)\\ &\propto \frac{1}{d}\mathrm{tr}\left(X_{\mathbf{a}_i}X_{\mathbf{a}_j}Z_{\mathbf{b}_i}Z_{\mathbf{b}_j}W\right),\\ W &= C^nZX_{\mathbf{a}_i}C^nZX_{\mathbf{a}_i}.\end{aligned} \tag{B65}$$

Notice that to make $\mathcal{U}_{ji}$ nonzero, we at least require $X_{\mathbf{a}_i} = X_{\mathbf{a}_j}$. Define support of $X_{\mathbf{a}_i}$ to be the set, $\mathrm{supp}(X_{\mathbf{a}_i}) = \{i|X_{\mathbf{a}_i} \text{ on qubit } i \text{ is } X\}$, which is essentially the location of 1 in $\mathbf{a}_i$. Notice that two of the diagonal elements of W are -1 while others are all 1. Suppose these two elements are W_{i_1,i_1} and W_{i_2,i_2}. Without loss of generality, i_2 can be set as $2^{n+1}-1$ as the element of the last row and last column is -1. To make $\mathcal{U}_{ji}$ nonzero, we furthermore require that $(Z_{\mathbf{b}_i}Z_{\mathbf{b}_j})_{i_1,i_1} = (Z_{\mathbf{b}_i}Z_{\mathbf{b}_j})_{i_2,i_2}$, or $\langle i_1|Z_{\mathbf{b}_i}Z_{\mathbf{b}_j}|i_1\rangle = \langle i_2|Z_{\mathbf{b}_i}Z_{\mathbf{b}_j}|i_2\rangle$. Note that $|i_2\rangle = X_{\mathbf{a}_i}|i_1\rangle$. Thus, to make $\mathcal{U}_{ji}$ nonzero, we require that $X_{\mathbf{a}_i}$ and $Z_{\mathbf{b}_i}Z_{\mathbf{b}_j}$ commute, which is equivalent to that $Z_{\mathbf{b}_i}Z_{\mathbf{b}_j}$ has even number of Z gates on qubits of $\mathrm{supp}(X_{\mathbf{a}_i})$. Hence, the block spanned by $X_{\mathbf{a}_i}\mathsf{Z}$ can be further split into two blocks when $m = 2$. One block is spanned by $X_{\mathbf{a}_i}\mathsf{Z}_e = \{X_{\mathbf{a}_i}Z_{\mathbf{b}_k}, \mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 0\}$ and the other block is spanned by $X_{\mathbf{a}_i}\mathsf{Z}_o = \{X_{\mathbf{a}_i}Z_{\mathbf{b}_k}, \mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 1\}$. Here, $\mathbf{a}_i \cap \mathbf{b}_k$ is the bitwise and operation between $\mathbf{a}_i$ and $\mathbf{b}_k$, $|\mathbf{a}_i \cap \mathbf{b}_k|$ is the weight of $\mathbf{a}_i \cap \mathbf{b}_k$, and $\mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 0$ means that there are even number of Z gates of $Z_{\mathbf{b}_k}$ acting on qubits of $\mathrm{supp}(X_{\mathbf{a}_i})$. Notice that $\mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 0$ is equivalent to that $\sum_{1\leqslant j\leqslant n}(\mathbf{a}_i)_j(\mathbf{b}_k)_j$ is even and $\mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 1$ is equivalent to that $\sum_{1\leqslant j\leqslant n}(\mathbf{a}_i)_j(\mathbf{b}_k)_j$ is odd. These relations are useful in the following proof.

Now we obtain that the Liouville representation of $\mathcal{U}$ is spanned by blocks with bases $X_{\mathbf{a}}\mathsf{Z}_e$ and $X_{\mathbf{a}}\mathsf{Z}_o$. To show $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$, we also need to check whether Λ_G has the same diagonal blocks with $\mathcal{U}$ and whether Λ_G is proportional to identity when restricting in these blocks. The proof is similar to that in Lemma 15. Recall that $\Lambda_G = \sum_i \lambda_i|\sigma_i\rangle\rangle\langle\langle\sigma_i|$ in Liouville representation. Also, if there exists an element $G \in \mathsf{G}$ such that $\sigma_i = G\sigma_jG^\dagger$, then $\lambda_i = \lambda_j$, and σ_i and σ_j would be in the same blocks. Then, we only need to check whether any element in $X_{\mathbf{a}}\mathsf{Z}_e$ can be generated from $X_{\mathbf{a}}$ by conjugate action of elements in G.

Given $X_{\mathbf{a}} \in \{\mathbb{I}, X\}^{\otimes n}/\mathbb{I}^{\otimes n}$, choose i from $\mathrm{supp}(X_{\mathbf{a}})$. For any $Z_{\mathbf{b}} \in \{\mathbb{I}, Z\}^{\otimes n}$ where $X_{\mathbf{a}}Z_{\mathbf{b}} \in X_{\mathbf{a}}\mathsf{Z}_e$, we can choose $W = \prod_{j\neq i} CZ_{ij}^{b_j}$, such that

$$\begin{aligned}WX_{\mathbf{a}}W^\dagger &= X_{\mathbf{a}}(X_{\mathbf{a}}^\dagger WX_{\mathbf{a}}W^\dagger)\\ &= X_{\mathbf{a}}\left(\prod_{1\leqslant j\leqslant n, j\neq i} Z_j^{b_j}\right)Z_i^{\sum_{1\leqslant j\leqslant n, j\neq i}a_jb_j}\\ &= X_{\mathbf{a}}\left(\prod_{1\leqslant j\leqslant n, j\neq i} Z_j^{b_j}\right)Z_i^{(\sum_{1\leqslant j\leqslant n}a_jb_j)-a_ib_i}\\ &= X_{\mathbf{a}}\left(\prod_{1\leqslant j\leqslant n, j\neq i} Z_j^{b_j}\right)Z_i^{-a_ib_i}\\ &= X_{\mathbf{a}}\left(\prod_{1\leqslant j\leqslant n, j\neq i} Z_j^{b_j}\right)Z_i^{b_i}\\ &= X_{\mathbf{a}}Z_{\mathbf{b}}.\end{aligned} \tag{B66}$$

Here, in the fourth line, we utilize $X_{\mathbf{a}}Z_{\mathbf{b}} \in X_{\mathbf{a}}\mathsf{Z}_e$, which means $\sum_{1\leqslant j\leqslant n}a_jb_j$ is even. In the fifth line, we use the condition that $a_i = 1$. Thus, we show that Λ_G is proportional to identity when restricting in the block spanned by $X_{\mathbf{a}}\mathsf{Z}_e$. From a similar argument, it can be verified that this is also true for block spanned by $X_{\mathbf{a}}\mathsf{Z}_o$. Then, we successfully show that $\mathcal{U}$ and Λ_G have the same diagonal blocks, and $\mathcal{U}\Lambda_G\mathcal{U}^{-1} = \Lambda_G$ when $m = 2$.

As a conclusion, the twirling group $X \ltimes \mathsf{W}_X$ satisfies $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ for $U = C^nZ_m$ with $n = 1, m \geqslant 3$ or $n \geqslant 2, m \geqslant 2$. Proof is done. ■

7. Structure of W_X

In this part, we present the concrete form of W_X. Following the discussion of W_X in previous subsections, we map W_X into $\mathbb{Z}_m^{2^{n+1}}$ via Eq. (B47).

We first discuss the cardinality of W_X. The image $\phi(\mathsf{W}_X)$ is the linear span of bases $\{v_0 = e_0 - e_{2^{n+1}-1}, \cdots, v_{2^{n+1}-2} = e_{2^{n+1}-2} - e_{2^{n+1}-1}\}$. If considering the factor of global phase, the image would be the linear span of bases $\{v_0 = e_0 - e_{2^{n+1}-1}, \cdots, v_{2^{n+1}-2} = e_{2^{n+1}-2} - e_{2^{n+1}-1}, u = 2^{\min(n+1,k)}e_{2^{n+1}-1}\}$, and we denote this enlarged image to be $\phi'(\mathsf{W}_X)$. In $\phi'(\mathsf{W}_X)$, any two vectors differing by $v_{2^{n+1}-1} = (1 \quad \cdots \quad 1)^T$ correspond to the same gate. If $n+1 \geqslant k$, then u is $(0 \quad \cdots \quad 2^k)^T$. Now we consider the cardinality of $\phi'(\mathsf{W}_X)$. One can take arbitrary values from $\mathbb{Z}_m$ to fix the coefficients of the first $2^{n+1}-1$ bases. After that, the coefficient of the last basis u only has $\frac{m}{2^k}$ inequivalent choices from $\mathbb{Z}_m$ as $(a + \frac{m}{2^k})u = au$. Hence, the cardinality of $\phi'(\mathsf{W}_X)$ is

$$|\phi'(\mathsf{W}_X)| = m^{2^{n+1}-1} \cdot \frac{m}{2^k} = m^{2^{n+1}}/2^k. \tag{B67}$$

For any $\phi(w)$ belongs to $\phi'(\mathsf{W}_X)$, we can also find $\phi(w) + (1 \quad \cdots \quad 1)^T, \cdots, \phi(w) + (m-1)\cdot(1 \quad \cdots \quad 1)^T$ in $\phi'(\mathsf{W}_X)$. All these m terms correspond to the same gate. Thus, the cardinality of $\phi(\mathsf{W}_X)$ is just $|\phi'(\mathsf{W}_X)|$ divided by m,

excluding the redundancy of the global phase,

$$|\mathsf{W}_X| = |\phi(\mathsf{W}_X)| = \frac{1}{m}|\phi'(\mathsf{W}_X)| = m^{2^{n+1}-1}/2^k. \tag{B68}$$

Now we investigate the structure of $\mathsf{W}_X = \langle \Pi^\dagger C^n Z_m \Pi C^n Z_m^\dagger \rangle$. We first analyze the simple case of $m = q$ and $m = 2q$ where q is an odd number. Then, we use induction to analyze the general case that $m = q2^k$.

(1) Case 1: $m = q$.

In this case, $k = 0$, $\gcd(m, 2^{n+1}) = 1$, the vector $u = e_{2^{n+1}-1}$, and $\phi'(\mathsf{W}_X) = \mathbb{Z}_m^{2^{n+1}}$. Clearly, for any $0 \leqslant l \leqslant n$, $\phi(C^l Z_m) \in \phi'(\mathsf{W}_X)$, so we can obtain that $\langle C^n Z_m, C^{n-1} Z_m, \cdots, C Z_m, Z_m \rangle \leqslant \mathsf{W}_X$. Since $\langle C^n Z_m, C^{n-1} Z_m, \cdots, C Z_m, Z_m \rangle = \langle C^n Z_m \rangle \times \langle C^{n-1} Z_m \rangle \times \cdots \times \langle Z_m \rangle$, $|\langle C^n Z_m, C^{n-1} Z_m, \cdots, C Z_m, Z_m \rangle| = m^{\binom{n+1}{n+1}} \times m^{\binom{n+1}{n}} \times \cdots \times m^{\binom{n+1}{1}} = m^{2^{n+1}-1} = |\mathsf{W}_X|$. The cardinalities of W_X and its subgroup $\langle C^n Z_m, C^{n-1} Z_m, \cdots, C Z_m, Z_m \rangle$ are the same, hence, we can conclude that

$$\begin{aligned} \mathsf{W}_X &= \langle C^n Z_m, C^{n-1} Z_m, \cdots, C Z_m, Z_m \rangle \\ &= \langle C^n Z_m \rangle \times \langle C^{n-1} Z_m \rangle \times \cdots \times \langle Z_m \rangle. \end{aligned} \tag{B69}$$

(2) Case 2: $m = 2q$.

In this case, $k = 1$, $\gcd(m, 2^{n+1}) = 2$, and $u = 2e_{2^{n+1}-1}$. We can use the following formula to construct all $C^l Z_m$ with $0 \leqslant l \leqslant n-1$,

$$\begin{aligned} C^l Z_m = \phi^{-1}(& v_{(2^{l+1}-1)*2^{n-l}} + v_{(2^{l+1}-1)*2^{n-l}+1} \\ &+ \cdots + v_{2^{n+1}-2} + 2^{n-l-1} \cdot u). \end{aligned} \tag{B70}$$

But this construction is not valid for $C^n Z_m$ as it requires $\frac{1}{2}u$. We can only construct $C^n Z_m^2 = C^n Z_{m/2}$. This implies that $\langle C^n Z_{m/2}, C^{n-1} Z_m, C^{n-2} Z_m, \cdots \rangle \leqslant \mathsf{W}_X$. Since $\langle C^n Z_{m/2}, C^{n-1} Z_m, \cdots \rangle = \langle C^n Z_{m/2} \rangle \times \langle C^{n-1} Z_m \rangle \times \cdots \times \langle Z_m \rangle$, $|\langle C^n Z_{m/2}, C^{n-1} Z_m, \cdots, C Z_m, Z_m \rangle| = (m/2)^{\binom{n+1}{n+1}} \times m^{\binom{n+1}{n}} \times \cdots \times m^{\binom{n+1}{1}} = m^{2^{n+1}-1}/2 = |\mathsf{W}_X|$. Thus, in this case,

$$\begin{aligned} \mathsf{W}_X &= \langle C^n Z_{m/2}, C^{n-1} Z_m, \cdots, C Z_m, Z_m \rangle \\ &= \langle C^n Z_{m/2} \rangle \times \langle C^{n-1} Z_m \rangle \times \cdots \times \langle Z_m \rangle. \end{aligned} \tag{B71}$$

(3) Case 3.1: $m = q2^k$ with $k \leqslant n$.

We first give the result and prove it by induction.

Theorem 3. If $m = q2^k$ and $k \leqslant n$, then the structure of W_X can be expressed in the following recursion:

$$\begin{aligned} \mathsf{W}_X &= \langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots, Z_m \rangle \times \langle A_k \rangle, \\ A_0 &= \{\mathbb{I}\}, A_1 = \{C^n Z_{m/2}\}, \\ \forall 2 \leqslant k \leqslant n, A_k &= \{C^{n-k+1} Z_{m/2}, g C^{n-k+1} Z_m | g \in A_{k-1}\}. \end{aligned} \tag{B72}$$

Proof. The proof idea is showing $|\mathsf{W}_X| = |\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots, Z_m \rangle \times \langle A_k \rangle|$ and $\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots, Z_m \rangle \times \langle A_k \rangle \leqslant \mathsf{W}_X$.

Let us start with studying the relation between $\langle A_{k-1} \rangle$ and $\langle A_k \rangle$. For any $g \in \langle A_{k-1} \rangle$, it can be decomposed as $g = g_1^{\lambda_1} g_2^{\lambda_2} \cdots$ where $g_1, g_2, \cdots$ all belong to A_{k-1} and $\lambda_1, \lambda_2, \cdots$ are integers. Since $g_1 C^{n-k+1} Z_m, g_2 C^{n-k+1} Z_m, \cdots \in A_k$, we can conclude that $g' = (g_1 C^{n-k+1} Z_m)^{\lambda_1} (g_2 C^{n-k+1} Z_m)^{\lambda_2} \cdots = g C^{n-k+1} Z_m^{\sum_i \lambda_i} \in \langle A_k \rangle$. Suppose $\sum_i \lambda_i$ is an even number, since $C^{n-k+1} Z_{m/2} = C^{n-k+1} Z_m^2 \in A_k$, the set $\{g, g C^{n-k+1} Z_m^2, g C^{n-k+1} Z_m^4, \cdots\} \subseteq \langle A_k \rangle$. Now we want to argue that any gate has the form of $g C^{n-k+1} Z_m^{2p+1}$ cannot be in $\langle A_k \rangle$. This claim can be proved by contradiction. If $g C^{n-k+1} Z_m^{2p+1} \in \langle A_k \rangle$, combining with $g C^{n-k+1} Z_m^{2p} \in \langle A_k \rangle$, then it is clear that $C^{n-k+1} Z_m = (g C^{n-k+1} Z_m^{2p})^\dagger g C^{n-k+1} Z_m^{2p+1} \in \langle A_k \rangle$. It leads to a contradiction as $C^{n-k+1} Z_m$ cannot be constructed with $\{v_i, 0 \leqslant i \leqslant 2^{n+1} - 2\}$ and $u = 2^k e_{2^{n+1}-1}$. Similarly, if $\sum_i \lambda_i$ is an odd number, then we can see that all elements have the form $g C^{n-k+1} Z_m^{2p+1}$ will belong to $\langle A_k \rangle$ while elements within form of $g C^{n-k+1} Z_m^{2p}$ will not. In either case, one element g in $\langle A_{k-1} \rangle$ will contribute to $|\langle C^{n-k+1} Z_{m/2} \rangle| = \frac{1}{2}|\langle C^{n-k+1} Z_m \rangle|$ elements in $\langle A_k \rangle$, which implies that $|\langle A_k \rangle| = \frac{1}{2}|\langle A_{k-1} \rangle| \times |\langle C^{n-k+1} Z_m \rangle|$. By iteration, the cardinality of $\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots \rangle \times \langle A_k \rangle$ is

$$\begin{aligned} &|\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots \rangle \times \langle A_k \rangle| \\ &= \frac{1}{2}|\langle C^{n-k+1} Z_m, C^{n-k} Z_m, \cdots \rangle \times \langle A_{k-1} \rangle| \\ &= \frac{1}{2^k}|\langle C^n Z_m, C^{n-1} Z_m, \cdots \rangle| \\ &= m^{2^{n+1}-1}/2^k. \end{aligned} \tag{B73}$$

This is exactly equal to $|\mathsf{W}_X|$. To complete the proof, we only need to argue that $\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots \rangle \times \langle A_k \rangle \leqslant \mathsf{W}_X$. Note that $\phi'(\mathsf{W}_X)$ is spanned by bases $v_i, 1 \leqslant i \leqslant 2^{n+1} - 2$ and u. We only need to investigate whether the generators of $\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots \rangle \times \langle A_k \rangle$ belong to the preimage set of ϕ.

For $C^{n-k-l} Z_m$ with $0 \leqslant l \leqslant n-k$, it can be expressed in the form of

$$\begin{aligned} C^{n-k-l} Z_m &= \phi^{-1}\left(2^l (0 \quad \cdots \quad 2^k)^T + \sum_{i=(2^{n+1-k-l}-1)\cdot 2^{k+l}}^{2^{n+1}-2} v_i\right) \\ &= \phi^{-1}\left(2^l u + \sum_{i=(2^{n+1-k-l}-1)\cdot 2^{k+l}}^{2^{n+1}-2} v_i\right). \end{aligned} \tag{B74}$$

Thus, $C^{n-k-l} Z_m$ belongs to W_X. Next we prove $\langle A_k \rangle \leqslant \mathsf{W}_X$, or equivalently, $A_k \subseteq \mathsf{W}_X$, by induction. Denote $\mathsf{W}_X^k = \langle \Pi^\dagger C^n Z_{q2^k} \Pi C^n Z_{q2^k}^\dagger \rangle$. Suppose $A_{k-1} \subseteq \mathsf{W}_X^{k-1}$. Then, any element $g \in A_{k-1}$ has the following expression:

$$g = \phi^{-1}\left((0 \quad \cdots \quad 2^{k-1})^T + \sum_{i=0}^{2^{n+1}-2} c_i v_i\right). \tag{B75}$$

Note that $C^{n-k+1} Z_m$ can also be expressed in the above form as

$$C^{n-k+1} Z_m = \phi^{-1}\left((0 \quad \cdots \quad 2^{k-1})^T + \sum_{i=(2^{n-k+2}-1)\cdot 2^{k-1}}^{2^{n+1}-2} v_i\right). \tag{B76}$$

This implies that $C^{n-k+1}Z_m^2$ and $C^{n-k+1}Z_m g$ can be expressed as follows:

$$C^{n-k+1}Z_m^2 = \phi^{-1}\left((0 \quad \cdots \quad 2^k)^T + \sum_{i=(2^{n-k+2}-1)\cdot 2^{k-1}}^{2^{n+1}-2} 2v_i\right), \tag{B77}$$

$$gC^{n-k+1}Z_m = \phi^{-1}\left((0 \quad \cdots \quad 2^k)^T + \sum_{i=0}^{(2^{n-k+2}-1)\cdot 2^{k-1}-1} c_i v_i + \sum_{i=(2^{n-k+2}-1)\cdot 2^{k-1}}^{2^{n+1}-2} (c_i+1)v_i\right). \tag{B78}$$

Therefore, any element $g' \in A_k$ has the expression

$$g' = \phi^{-1}\left((0 \quad \cdots \quad 2^k)^T + \sum_{i=0}^{2^{n+1}-2} c_i v_i\right). \tag{B79}$$

It means that all elements of A_k can be constructed with u and v_i, and belong to W_X^k. Now we have shown that $A_{k-1} \subseteq \mathsf{W}_X^{k-1}$ implies $A_k \subseteq \mathsf{W}_X^k$. For the induction argument to hold, we only need to check the initial condition, i.e., $A_1 \subseteq \mathsf{W}_X^1$. This reduces to Case 2. Thus, we prove that $A_k \subseteq \mathsf{W}_X$ and $\langle A_k\rangle \leqslant \mathsf{W}_X$. Hence, $\langle C^{n-k}Z_m, C^{n-k-1}Z_m, \cdots, Z_m\rangle \times \langle A_k\rangle \leqslant \mathsf{W}_X$. Combining with Eq. (B73), we obtain that $\mathsf{W}_X = \langle C^{n-k}Z_m, C^{n-k-1}Z_m, \cdots\rangle \times \langle A_k\rangle$ and proof is done. ■

(4) Case 3.2: $m = q2^k$ with $k > n$.

In this case, $\gcd(2^{n+1}, m) = 2^{n+1}$ and $\mathsf{W}_X = \langle A_{n+1}\rangle$ as described in the previous case.

Below we present examples of the twirling groups for target gate U to be C^nZ and CZ_m, which have been mentioned in the main text. In the next subsection, we will analyze their sample complexity and computational complexity.

Example 2. The optimal twirling group G in CRU for multiqubit controlled Z gate C^nZ is

$$\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle = \mathsf{X} \ltimes \langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z\rangle. \tag{B80}$$

Example 3. The optimal twirling group G in CRU for controlled phase gate CZ_m is

$$\langle CZ_m, Z_m, X\rangle = \mathsf{X} \ltimes \langle CZ_m, Z_m\rangle, \tag{B81}$$

if m is odd, and

$$\langle CZ_{m/2}, Z_m, X\rangle = \mathsf{X} \ltimes \langle CZ_{m/2}, Z_m\rangle, \tag{B82}$$

if m is even and $m/2$ is odd, and

$$\left\langle CZ_{m/4}, Z_{m/2}, X, CZ_{m/2}Z_m^1, CZ_{m/2}Z_m^2\right\rangle = \mathsf{X} \ltimes \left\langle CZ_{m/4}, Z_{m/2}, CZ_{m/2}Z_m^1, CZ_{m/2}Z_m^2\right\rangle, \tag{B83}$$

if $m/2$ is even. It is worth mentioning that $\langle CZ_{m/4}, Z_{m/2}, CZ_{m/2}Z_m^1, CZ_{m/2}Z_m^2\rangle$ is a subgroup of $\langle CZ_{m/2}, Z_m\rangle$.

8. Complexity analysis

To enable randomized benchmarking or any other quantum information tasks with twirling groups, one needs to sample from the group and compute the multiplication of group elements. It is necessary to analyze the sample complexity and the computational complexity of the twirling group. The sample complexity is directly related to the cardinality of the corresponding twirling group. The computational complexity is related to both the group structure and the algorithm for computing the multiplication. We will provide a group multiplication algorithm and present its complexity as the upper bound of the group computational complexity. In the discussion below, we distinguish the number of qubits N and the number of controlled qubits n. In general, $n \leqslant N-1$.

To benchmark C^nZ gate, the twirling group is $\mathsf{G}_{C^nZ} = \langle C^{n-1}Z, C^{n-2}Z\cdots, CZ, Z, X\rangle$ shown in Example 2. Note that this group has a semidirect product structure,

$$\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle = \langle X\rangle \rtimes (\langle C^{n-1}Z\rangle \times \langle C^{n-2}Z\rangle \times \cdots \langle Z\rangle). \tag{B84}$$

This means that any element in G_{C^nZ} can be written in the form of $\Pi W_{n-1}W_{n-2}\cdots W_1W_0$ where $\Pi \in \langle X\rangle$ and $W_l \in \langle C^lZ\rangle$. To sample from this twirling group G_{C^nZ}, we just need to sample independently from $\langle C^{n-1}Z\rangle$, $\langle C^{n-2}Z\rangle$, $\cdots$, $\langle Z\rangle$, and $\langle X\rangle$. Since the orders of all generators are 2, we can use a binary string whose length equals the number of generators of the group to represent an arbitrary group element. Then, we can sample the group element by sampling the binary string. For instance, the generators of $\langle CZ\rangle$ are $\{CZ_{1,2}, CZ_{1,3}, \cdots, CZ_{1,N}, \cdots, CZ_{2,3}, \cdots, CZ_{N-1,N}\}$. So a binary string with length $\binom{N}{2}$ is enough for sampling from $\langle CZ\rangle$. The total length needed to sample G_{C^nZ} is given by

$$\log|\mathsf{G}_{C^nZ}| = \log|\langle X\rangle \rtimes (\langle C^{n-1}Z\rangle \times \cdots \times \langle CZ\rangle \times \langle Z\rangle)| = n + \sum_{i=1}^{n}\binom{N}{i} = O(N^n), \tag{B85}$$

which is indeed the sample complexity. For computing the inverse gates, we will utilize the following identity. Given a C^nZ gate acting on qubits $i_1, \cdots, i_{n+1}$ and a subset of $\{i_1, \cdots, i_{n+1}\}$, I, then

$$C^nZ_{i_1,i_2,\cdots,i_{n+1}}\Pi_{i\in I}X_i = \Pi_{i\in I}X_i\Pi_{S\subseteq I}C^{n-|S|}Z_{\{i_1,\cdots,i_{n+1}\}\setminus S}. \tag{B86}$$

So, the multiplication of two elements in the group is calculated by

$$\begin{aligned}
&\Pi^{(1)}W_{n-1}^{(1)}\cdots W_0^{(1)}\Pi^{(2)}W_{n-1}^{(2)}\cdots W_0^{(2)}\\
&\quad = (-1)^{\Pi^{(2)}\cdot W_0^{(1)}}\Pi^{(1)}W_{n-1}^{(1)}\cdots W_1^{(1)}\Pi^{(2)}W_0^{(1)}W_{n-1}^{(2)}\cdots W_0^{(2)}\\
&\quad = (-1)^{\Pi^{(2)}\cdot W_0^{(1)}}\Pi^{(1)}W_{n-1}^{(1)}\cdots\\
&W_2^{(1)}\Pi^{(2)}W_0'W_1^{(1)}W_0^{(1)}W_{n-1}^{(2)}\cdots W_0^{(2)}\\
&\quad = (-1)^{\Pi^{(2)}\cdot W_0^{(1)}}\Pi^{(1)}\Pi^{(2)}W_{n-2}'W_{n-1}^{(1)}\cdots\\
&W_0'W_1^{(1)}W_0^{(1)}W_{n-1}^{(2)}W_{n-2}^{(2)}\cdots W_1^{(2)}W_0^{(2)}.
\end{aligned} \tag{B87}$$

The $W_0', \cdots, W_{n-2}'$ in the third and fourth lines are indeed the additional controlled-Z gates and multiqubit controlled-Z gates in Eq. (B86).

Now, we give a brief analysis of the total complexity. The key point is utilizing Eq. (B86) to shift $\Pi^{(2)}$ gate to the left. We first fix an integer l and focus on C^lZ. Now we try to study

the swapping between C^lZ and $\Pi^{(2)}$. By Eq. (B86), the total complexity for shifting $\Pi^{(2)}$ across a C^lZ gate is $O(2^{l+1})$ since the cardinality of subset I is at most $l+1$ and the number of subsets of I is at most 2^{l+1}. As the number of C^lZ gates on N qubits is at most $\binom{N}{l+1}$, the total complexity for shifting $\Pi^{(2)}$ to the leftmost is

$$\text{group multiplication complexity} = \sum_{l=0}^{n-1} \binom{N}{l+1} \cdot 2^{l+1}. \tag{B88}$$

The right part is bounded by N^n. Hence, the total group multiplication complexity is $O(N^n)$. Notice that this bound is quite untight. For example, if we let $n = N$, then the right part of Eq. (B88) is indeed $3^n - 1$.

The inverse of an element $\Pi W_{n-1}\cdots W_0$ is $W_{n-1}\cdots W_0\Pi$, which equals the multiplication of element $W_{n-1}\cdots W_1 1_{\langle X\rangle}$ and $1_{\langle C^{n-1}Z\rangle}\cdots 1_{\langle Z\rangle}\Pi$, where 1_{G} represents the identity in group G and is indeed the all-zero string in the binary string representation of group elements. The computational complexity for computing inverse gate is then equal to the complexity for group element multiplication, i.e., $O(N^n)$.

To benchmark CZ_m gate, the twirling group by our protocol would always be a subgroup of $\mathsf{G}_{CZ_m} = \langle CZ_m, Z_m, X\rangle$. For simplicity, we only analyze the complexity of $\langle CZ_m, Z_m, X\rangle$. This complexity would be the upper bound of the complexity of the twirling group shown in Example 3. $\langle CZ_m, Z_m, X\rangle$ has a similar direct product structure like $\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle$,

$$\mathsf{G}_{CZ_m} = \langle X\rangle \rtimes (\langle CZ_m\rangle \times \langle Z_m\rangle). \tag{B89}$$

Thus, each element in G_{CZ_m} can be expressed as $W_2W_1\Pi$ where $W_2 \in \langle CZ_m\rangle$, $W_1 \in \langle Z_m\rangle$ and $\Pi \in \langle X\rangle$. In this case, the orders of generators CZ_m and Z_m are both m. We need to use $\lceil \log m\rceil$ bits to record their multiplicities and to uniquely identify one group element. Then, to represent an element in $\langle CZ_m\rangle$, we should use a bit string with length $\lceil \log m\rceil N(N-1)/2$. The total number of bits to express a group element is

$$\begin{aligned}\log|\langle X\rangle \rtimes (\langle CZ_m\rangle \times \langle Z_m\rangle)| &= \left(\frac{N(N-1)}{2} + N\right)\lceil \log m\rceil + N \\ &= O(N^2 \log m),\end{aligned} \tag{B90}$$

which is indeed the sample complexity. The equation Eq. (B86) can be extended for CZ_m in the following way:

$$CZ_m^k X_i = X_i CZ_m^{-k} (Z_m)_j^k, \tag{B91}$$

where the CZ_m here is on qubit i, j. By a similar argument like C^nZ, we can see that the group multiplication complexity, and thus inverse computation complexity, is $O(N^2 \log m)$.

The group size and computational complexity for the CNOT dihedral group have been elaborated in Ref. [16]. So, we omit the details here. Note that for CZ_m, their results only apply to $m = 2^k$, and ours admit m taking arbitrary positive integers. The scaling of the complexity results has been summarized and listed in the main text. In Fig. 8, we also show an accurate result of the size comparison between our group and the CNOT dihedral group by considering $N = n + 1$. The size of our group increases slower with respect to the qubit number than that of the CNOT dihedral group.

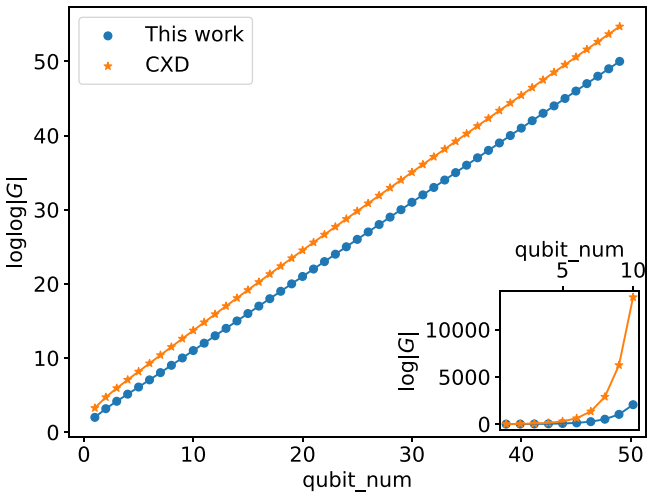


FIG. 8. Size comparison between the twirling groups for C^nZ in our method and in Ref. [16], associated with blue and orange patterns, respectively. $|\mathsf{G}|$ represents the group size. If taking a double logarithmic scale, then both kinds of groups increased nearly linearly with respect to the qubit number, but the CNOT dihedral group increased faster. The smaller figure demonstrates the results in the logarithm scale.

Below, we discuss the hierarchy of twirling groups, which may be additionally interesting. In the main text, we have mentioned that by choosing the twirled gate to be C^nZ_m with increasing n and m, the size of the twirling group and the computational complexity of the group multiplication would keep increasing. The two complexities are closely related to the classical simulability of the twirling group, and the hierarchy of the groups actually forms a computational complexity hierarchy. The results may be useful in other fields like quantum complexity theory and quantum simulation.

Here, we discuss the hierarchy of groups themselves rather than the hierarchy of their sample and computational complexities. Specifically, we focus on the group of $\langle C^nZ, C^{n-2}Z, \cdots, CZ, X, Z_{2^k}\rangle$. Starting from Pauli group $\langle X, Z\rangle$, there are two ways to build up the hierarchy. One way is adding the number of sides of dihedral groups, i.e., increasing the number k in $\langle X, Z_{2^k}\rangle$. The other way is adding multiqubit controlled Z gates C^nZ or even adding multiqubit controlled X gates C^nX. Interestingly, C^nZ can be contained in the CNOT-dihedral group $\langle CX, X, Z_{2^k}\rangle$ by choosing $k = n+1$ [16]. The hierarchy is summarized as follows:

$$\begin{aligned}&\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z_{2^k}\rangle \\ &\quad\leqslant \langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z_{2^{k+1}}\rangle \cdots \\ &\quad\leqslant \langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z(\theta)\rangle \cdots ;\end{aligned} \tag{B92}$$

$$\begin{aligned}\langle C^{n-1}X, X, Z_{2^k}\rangle &\leqslant \langle C^{n-1}X, X, Z_{2^{k+1}}\rangle \cdots \\ &\leqslant \langle C^{n-1}X, X, Z(\theta)\rangle \cdots \\ &\leqslant \mathrm{CRU}_n,\end{aligned} \tag{B93}$$

$$\begin{aligned}\langle X, Z_{2^k}\rangle &\leqslant \langle CZ, X, Z_{2^k}\rangle \\ &\leqslant \langle CCZ, CZ, X, Z_{2^k}\rangle \cdots,\end{aligned} \tag{B94}$$

$$\begin{aligned}\langle X, Z_{2^k}\rangle &\leqslant \langle CX, X, Z_{2^k}\rangle \\ &\leqslant \langle CCX, X, Z_{2^k}\rangle \cdots \\ &\leqslant \mathrm{CRU}, \end{aligned} \tag{B95}$$

$$\langle C^{k-1}Z, C^{k-2}Z, \cdots, CZ, X, Z_{2^k}\rangle \leqslant \langle CX, X, Z_{2^k}\rangle. \tag{B96}$$

Here, CRU_n represents the CRU on n qubits and CRU represents $\cup_n \mathrm{CRU}_n$. As shown in the above equations, the upper limit of the hierarchy is CRU. The key point is that $\langle C^{n-1}X, X\rangle$ contains all permutations over the computational bases $\{0,1\}^n$ based on Lemma 10. Thus, $\langle C^{n-1}X, X, Z(\theta)\rangle$ contains all quantum gates like ΠW where Π is an arbitrary permutation gate, and W is an arbitrary diagonal gate. This is exactly the necessary and sufficient condition for CRU. Also, from the group hierarchy, it is clearer that the twirling group of our protocol is smaller than that in Ref. [16] for multiqubit controlled-Z gates C^nZ. It is straightforward by investigating $\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z\rangle \leqslant \langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z_{2^k}\rangle \leqslant \langle CX, X, Z_{2^k}\rangle$.

APPENDIX C: SIMULATION

In this part, we present additional details of the simulation, including the setting of the noise model, the benchmarking protocol using the ZX-SPAM setting, and additional benchmarking results. We denote the total number of qubits as N, which is 2 for the CS gate and $n+1$ for the C^nZ gate.

1. Noise model

In summary, to simulate a noisy quantum gate, we consider local dephasing noise, local amplitude damping noise, and unitary noise. The strength of the former two noises depends on the time to implement the gate. To calculate this time, we first decompose this gate into a series of CNOT gates, Pauli X gates, multiqubit controlled phase gates, and single-qubit phase gates. We sum the time to implement each part to obtain the total time of implementing a gate, denoted as T_{gate}. Then, the former two types of noise are simulated as follows.

(1) Local dephasing channel Λ_d. The dephasing channel on qubit i, where $1 \leqslant i \leqslant N$, is defined as

$$\Lambda_d^i(\rho) = K_0^i \rho K_0^{i\dagger} + K_1^i \rho K_1^{i\dagger}. \tag{C1}$$

Here,

$$K_0^i \equiv \begin{pmatrix} \sqrt{p_i} & 0 \\ 0 & \sqrt{p_i} \end{pmatrix}, \quad K_1^i \equiv \begin{pmatrix} \sqrt{1-p_i} & 0 \\ 0 & -\sqrt{1-p_i} \end{pmatrix}, \tag{C2}$$

where p_i denotes the dephasing strength. The local dephasing channel is defined as

$$\Lambda_d = \bigotimes_{i=1}^{N} \Lambda_d^i. \tag{C3}$$

In the simulation, we set all $p_i, 1 \leqslant i \leqslant N$ to be the same, which equals

$$p_i = e^{-\frac{T_{gate}}{T_2}}, \tag{C4}$$

where T_{gate} denotes the time to implement a gate and depends on the concrete gate and $T_2 = 15\,000$ ns.

(2) Local amplitude damping channel Λ_a. The amplitude damping channel on qubit i, where $1 \leqslant i \leqslant N$, is defined as

$$\Lambda_a^i(\rho) = K_0^i \rho K_0^{i\dagger} + K_1^i \rho K_1^{i\dagger}. \tag{C5}$$

Here,

$$K_0^i \equiv \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{q_i} \end{pmatrix}, \quad K_1^i \equiv \begin{pmatrix} 0 & \sqrt{1-q_i} \\ 0 & 0 \end{pmatrix}. \tag{C6}$$

The parameter q_i denotes the noise strength of the amplitude damping channel. The local amplitude damping channel on N qubits is defined as

$$\Lambda_a = \bigotimes_{i=1}^{N} \Lambda_a^i. \tag{C7}$$

In the simulation, we set all $q_i, 1 \leqslant i \leqslant N$ to be the same, which equals

$$q_i = e^{-\frac{T_{\mathrm{gate}}}{T_1}}, \tag{C8}$$

where T_{gate} depends on the concrete gate and $T_1 = 25\,000$ ns.

To simulate the unitary noise for each gate, we also utilize the decomposition of the gate and consider unitary errors for each decomposed part. Then, the noisy gate is simulated by sequentially acting local dephasing noise, local amplitude damping noise, and gate with unitary errors.

We introduce the calculation of T_{gate} and the unitary noise for each gate below. Recall that all of the twirling gates and the twirled gate belong to CRU. A CRU gate can be separated into a permutation part and a diagonal part. The time T_{gate} is evaluated by the sum of times of two parts. The unitary errors are also considered separately. We discuss the noise models for the two parts respectively.

(1) The permutation part is $\langle CX, X\rangle$ for the CNOT dihedral group and $\langle X\rangle$ for the other simulated groups. Note that one can always implement a number of CNOT gates, and then implement one layer of Pauli X gates to realize an element in $\langle CX, X\rangle$; $\langle X\rangle$ is realized by one layer of Pauli X gates.

For the part constructed by CNOT gates, denoted as 'CNOT part', we decompose the circuit based on the algorithm in Ref. [54] and get a synthesis of the circuit with CNOT gates and SWAP gates. We calculate the circuit depth of the "CNOT part," denoted as "d_{CNOT}," by counting the minimum layer to realize all CNOT gates. If two CNOT gates can be realized parallelly, then they are in the same layer. The SWAP gates do not contribute to the counting of circuit depth since one can always virtually implement SWAP gates by manually changing the label of qubits. Then, the time to implement the "CNOT part" is

$$t_{\mathrm{CNOT}} = d_{\mathrm{CNOT}} \times 2 \times 30 \text{ ns}. \tag{C9}$$

If there are no gates in the "CNOT part," then t_{CNOT} is set as 0. For each decomposed CNOT gate, we consider its unitary noise as

$$\mathrm{Controlled}-\exp(i\delta X), \tag{C10}$$

where $\delta = 0.005\pi$ and X is the Pauli X gate.

Similarly, the circuit depth of the Pauli X layer is 1 as long as one qubit is acted nontrivially by a Pauli X gate and 0

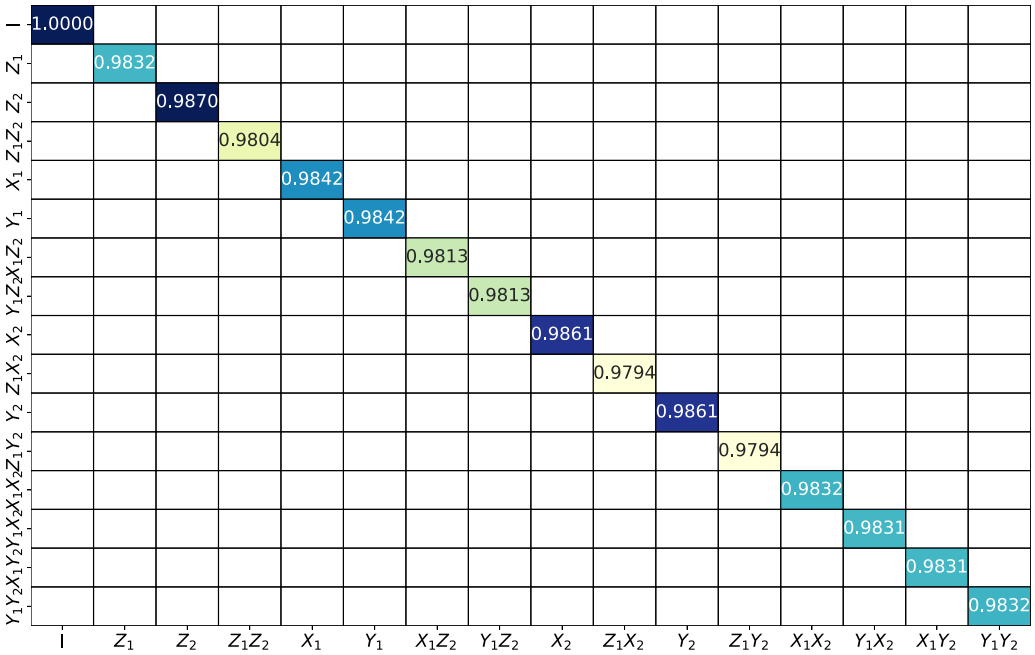


FIG. 9. Pauli Liouville representation of Λ_{P} where Λ is the noise channel of CS gate and P denotes the Pauli group. The values in vacant squares are all 0, and we omit them. The two axes record the bases of the Pauli-Liouville representation.

otherwise, which is denoted as d_X. The time to implement the Pauli X layer is

$$t_X = d_x \times 30 \text{ ns}. \tag{C11}$$

The unitary error for each decomposed Pauli X gate is

$$\exp(i\delta X), \tag{C12}$$

where $\delta = 0.005\pi$. The total time of the permutation part is $t_{\text{perm}} = t_{\text{CNOT}} + t_X$.

(2) The diagonal part of a CRU gate, W, can be generated by multiqubit controlled phase gates and single-qubit Z rotation:

$$W = \prod_{r=1}^{N} \prod_{i \in I_r} U_i^r, \tag{C13}$$

where $U_i^r = C^{r-1}Z_l$ is a multiqubit controlled phase gate with $r-1$ controlled qubit number, and l denotes the phase. When $r = 1$, U_i^r is a single-qubit Z rotation. The notation I_r is an index set. When $I_{r'}$ is empty, there is no multiqubit controlled phase gate with $r'-1$ controlled qubit number in the decomposition of W. Since U_i^r commutes with each other, we consider all these gates can be implemented at the same time in the simulation. Thus, the time to implement W is set as

$$t_{\text{diag}} = \max(\{r | I_r \neq \emptyset\}) \times 30 \text{ ns}. \tag{C14}$$

This is the time to implement the multiqubit controlled phase gate with the largest controlled qubit number. The unitary error is a multiple-Z coupling, considered for each multiqubit controlled phase gate in the decomposition. For $C^{r-1}Z_l = e^{i\frac{2\pi}{l}|1\rangle\langle 1|^{\otimes r}}$ gate, the unitary error is

$$U_Z = e^{i\delta r \sum_{z\in\{0,1\}^r} 2^{|z|-r}|z\rangle\langle z|}, \tag{C15}$$

where $\delta = 0.005\pi$, and $|z|$ is the number of 1 in z. The state with more 1 is influenced more.

Besides the noise of the gate, we also set errors for state preparation and measurement (SPAM). In our simulation experiments, the initial states are $U|0\rangle^{\otimes N}$ where U is a local Clifford gate. In the simulation, we suppose the local Clifford gate does not bring errors, and the error of the initial state comes from the noise when preparing $|0\rangle^{\otimes n}$. We assume that on each qubit, the state $|0\rangle$ would turn into $|1\rangle$ with probability 0.02, resulting from the thermal excitation. We take the measurement to be noiseless as there is always a measurement error correction in experiments.

2. Benchmarking procedure with the ZX-SPAM

Below, we present the benchmarking procedure with the ZX-SPAM when the twirling group contains the CNOT dihedral group $\text{CZD} = \langle CZ, X, S\rangle$. Before the introduction of the concrete benchmarking procedure, we explain why the ZX-SPAM suffices to extract all different diagonal terms when the twirling group contains the CNOT dihedral group.

We take CS gate for an example and present the Liouville representation of its twirled noise channel under several different twirling groups, including Pauli group $\mathsf{P} = \langle X, Z\rangle$, CZ Pauli group $\text{CZP} = \langle CZ, X, Z\rangle$, CZ dihedral group $\text{CZD} = \langle CZ, X, S\rangle$, and CNOT dihedral group $\text{CXD} = \langle CX, X, T\rangle$. The results are shown in Figs. 9–12.

The matrix elements of the original noise channel Λ of CS gate in the Pauli-Liouville representation are all nonzero and contain complex numbers. After twirled by the Pauli group, there are only diagonal terms left to be nonzero, as shown in Fig. 9. If we use a larger twirling group, the CZ Pauli group, for any $X_a \in \{X_1, X_2, X_1X_2\}$, then the matrix elements corresponding to X_a and $X_aZ_1Z_2$ would be averaged, and the elements of X_aZ_1 and X_aZ_2 would be also averaged. The matrix elements of Λ_{CZP} are shown in

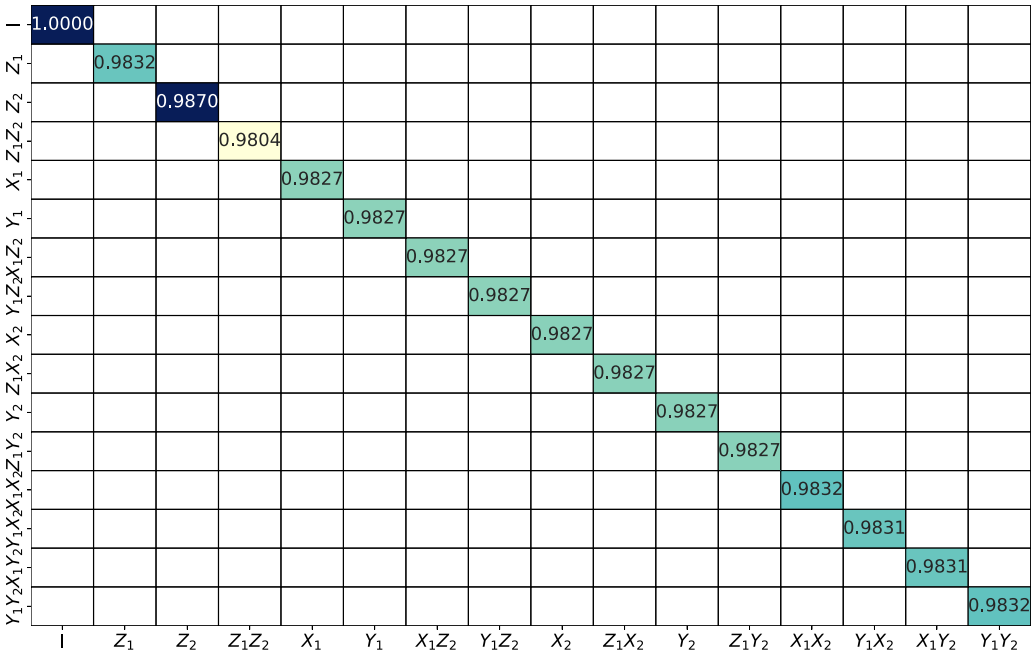


FIG. 10. Pauli Liouville representation of Λ_{CZP} where Λ is the noise channel of CS gate and CZP denotes the twirling group $\langle CZ, Z, X\rangle$. The values in vacant squares are all 0, and we omit them. Note that the elements corresponding to X_1X_2 and Y_1Y_2 are the same, and the elements corresponding to X_1Y_2 and X_1Y_2 are the same. But the two values are different. Generally, the elements corresponding to X_1 and Y_1 are different, and the elements corresponding to X_2 and Y_2 are also different. However, in this case, the noise channel is very special, and the difference can hardly be seen by accident.

Fig. 10. Its diagonal blocks are associated with the bases set in $\{\{\mathbb{I}\}, \{Z_1\}, \{Z_2\}, \{Z_1Z_2\}, \{X_aZ_1, X_aZ_2\}, \{X_a, X_aZ_1Z_2\}|X_a \in \{X_1, X_2, X_1X_2\}\}$. Thus, the number of different diagonal terms is 7, excluding the one associated with $\{\mathbb{I}\}$. Nonetheless, via the ZX-SPAM, one can at most obtain $2(2^N - 1) = 6$ different diagonal terms,

$$\mathrm{tr}(P\Lambda(P))/2^N, P \in (\{\mathbb{I}, Z\}^{\otimes N} \bigcup \{\mathbb{I}, X\}^{\otimes N})\backslash\mathbb{I}^{\otimes N}. \quad \text{(C16)}$$

Thus, the twirling of the CZ Pauli group does not suffice.

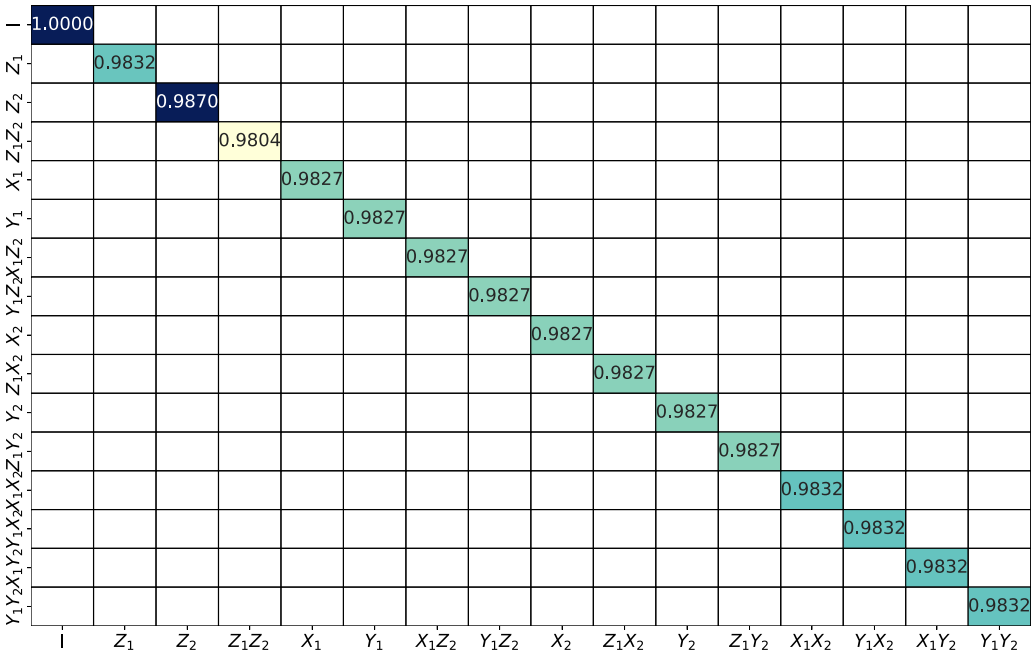


FIG. 11. Pauli Liouville representation of Λ_{CZD} where Λ is the noise channel of CS gate and CZD denotes the twirling group $\langle CZ, Z, S\rangle$. The values in vacant squares are all 0, and we omit them. The matrix elements corresponding to $X_a\mathsf{Z}$ are all averaged respectively for $X_a \in \{X_1, X_2, X_1X_2\}$.

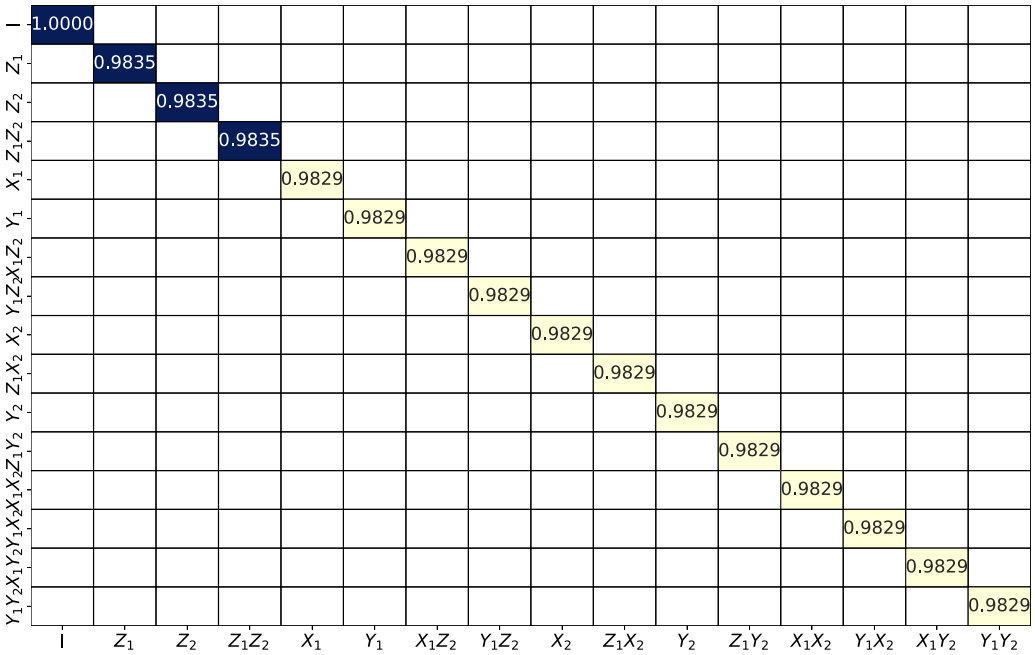
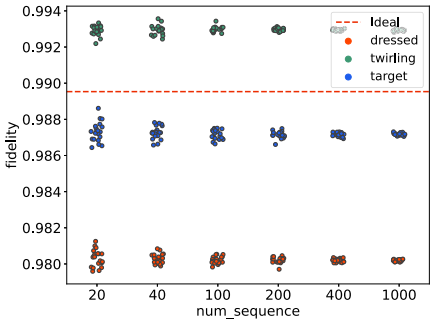
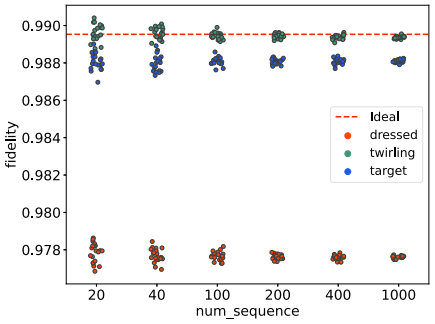


FIG. 12. Pauli Liouville representation of Λ_{CXD} where Λ is the noise channel of CS gate and CXD denotes the twirling group $\langle CX, X, T\rangle$. The values in vacant squares are all 0, and we omit them. The second to the fourth diagonal elements are averaged, and the last 12 diagonal elements are also averaged.

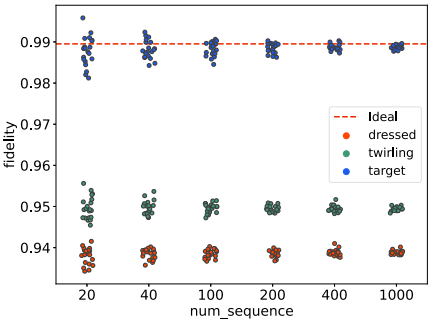
If we further use a bit larger twirling group, CZ dihedral group, then the matrix elements corresponding to $X_a\mathsf{Z}$ would be all averaged as shown in Fig. 11. The number of different diagonal terms is just $2(2^N - 1) = 6$. In this case, using the ZX-SPAM, we can extract the diagonal terms of Λ^m_{CZD} and hence obtain the diagonal terms of Λ_{CZD} via fitting. If using



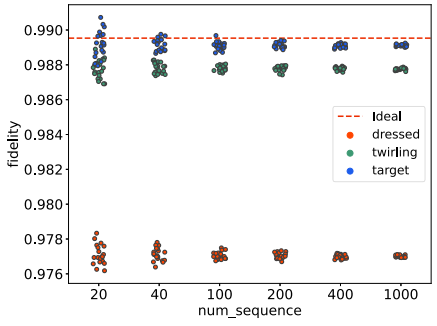
(a) CS, Pauli, random SPAM



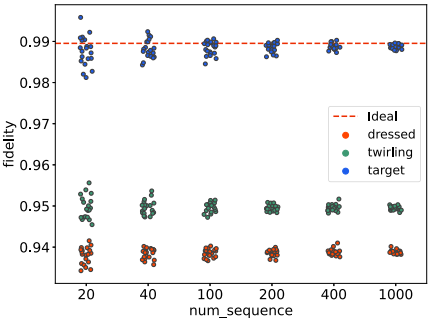
(b) CS, CZD, random SPAM



(c) CS, CXD, two SPAM



(d) CS, CZD, ZX-SPAM



(e) CS, CXD, ZX-SPAM

FIG. 13. The results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the CS gate with five benchmarking methods. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity.

the CNOT dihedral group for twirling, then the twirled noise channel Λ_{CXD} would be more symmetric than Λ_{CZD}. The elements corresponding to Z_1, Z_2, and Z_1Z_2 will be averaged and the elements corresponding to $\{X_a\mathsf{Z}|X_a \in \{X_1, X_2, X_1X_2\}\}$ will also be averaged. Thus, Λ_{CXD} only has two parameters as shown in Fig. 12. Using the ZX-SPAM can also extract all different diagonal terms of Λ_{CXD}.

In Box 1, we introduce the detailed benchmarking procedure with the ZX-SPAM. The twirling group contains the CZ dihedral group.

Box 1: Procedures for benchmarking with the ZX-SPAM

(1) First initialize the state, $|\psi_0\rangle = |0\rangle^{\otimes N}$ or $|\psi_1\rangle = |+\rangle^{\otimes N}$, where N is the number of qubits.

(2) Choose a positive integer, M, and choose two sets of positive integers $\{m_1, m_2, ..., m_M\}$ and $\{K_1, K_2, ..., K_M\}$. Here, $\{m_1, m_2, ..., m_M\}$ is the set of circuit depths and for $1 \leqslant i \leqslant m$, K_i is the number of sampled sequences when circuit depth equals m_i.

(3) For each integer $1 \leqslant i \leqslant M$, uniformly and randomly sample $2m_i$ gates from twirling group G, where $\mathsf{G} \supseteq \langle CZ, S, X\rangle$, for K_i times, which we denote $\{G_{j,1}, G_{j,2}, \cdots, G_{j,2m_i}\}$, $1 \leqslant j \leqslant K_i$.

(4) For each integer $1 \leqslant i \leqslant M$ and $1 \leqslant j \leqslant K_i$, implement gate sequence

$$\widetilde{S}(j, m_i) = \widetilde{U}_{\text{inv}} \prod_{t=1}^{m_i} U^\dagger G_{j,2t} U G_{j,2t-1}, \tag{C17}$$

where $\widetilde{\cdot}$ represents the noisy version of the quantum gate and $U_{\text{inv}} = (\prod_{t=1}^{m_i} U^\dagger G_{j,2t} U G_{j,2t-1})^\dagger$; U is the target gate.

(5) For initial state $|\psi_0\rangle = |0\rangle^{\otimes N}$, measuring all observables from $\{\mathbb{I}, Z\}^{\otimes N}$ of the final state via Z-basis measurement. For initial state $|\psi_1\rangle = |+\rangle^{\otimes N}$, measuring all observables from $\{\mathbb{I}, X\}^{\otimes N}$ of the final state via X-basis measurement. That is, for each $Q_k \in \{\mathbb{I}, Z\}^{\otimes N}$, estimate

$$f_Z(j, m_i, k) = \text{tr}(\widetilde{Q}_k \widetilde{S}(j, m_i)(\widetilde{\rho}_0)). \tag{C18}$$

For each $P_k \in \{\mathbb{I}, X\}^{\otimes N}$, estimate

$$f_X(j, m_i, k) = \text{tr}(\widetilde{P}_k \widetilde{S}(j, m_i)(\widetilde{\rho}_1)). \tag{C19}$$

Here, $\widetilde{\rho}_0$ and $\widetilde{\rho}_1$ are noisy versions of $|0\rangle^{\otimes N}$ and $|+\rangle^{\otimes N}$, respectively.

(6) Average the results of different gate sequences and obtain

$$f_Z(m_i, k) = \frac{1}{K_i} \sum_{j=1}^{K_i} f_Z(j, m_i, k), \tag{C20}$$

$$f_X(m_i, k) = \frac{1}{K_i} \sum_{j=1}^{K_i} f_X(j, m_i, k). \tag{C21}$$

(7) For each Q_k, fit $f_Z(m_i, k)$ to function

$$f(m_i) = A\lambda_{Z,k}^{m_i} \tag{C22}$$

and obtain $\lambda_{Z,k}$. For each P_k, fit $f_X(m_i, k)$ to function

$$f(m_i) = A\lambda_{X,k}^{m_i} \tag{C23}$$

and obtain $\lambda_{X,k}$.

(8) Estimate the fidelity of the target gate via

$$F = \frac{\sum_k \lambda_{Z,k} + 2^N\left(\sum_k \lambda_{X,k} - 1\right)}{4^N}. \tag{C24}$$

(9) If one wants to separate further the noise of the twirling gates and the target gate, then execute the following step. Replace the target gate with identity and repeat the above processes to estimate the fidelity of the twirling groups, F_G. Then, estimate the fidelity of U with F and F_G by

$$F_U = \frac{d^2F - 1}{d^2F_G - 1}\left(1 - \frac{1}{d^2}\right) + \frac{1}{d^2}. \tag{C25}$$

Note that in the above benchmarking procedure, we use the circuit structure in Ref. [15] instead of Ref. [14]. That means we implement random twirling gates interleaved with U and $U^\dagger$ instead of being just interleaved with U. For C^nZ gate, $U = U_{\text{inv}}$ and this modification does not influence anything. For more general case that U is C^nZ_m gate like CS gate, under a mild assumption that U and $U^\dagger$ has the same noise Λ, the above procedure can estimate the fidelity of $\sqrt{\mathcal{U}^\dagger\Lambda_G\mathcal{U}\Lambda_G}$, which is equal to the fidelity of Λ. In this modified procedure, the circuit depth only needs to be chosen as the multiples of 2. While within the circuit structure in Fig. 6, just like Ref. [14], the circuit depth needs to be chosen as the multiples of the order of the target gate to ensure the inverse gate belonging to the twirling group. For instance, the order of C^nZ_m is m and can be large for C^nZ_m with large m. In this case, the modified benchmarking procedure in Box 1 can offer an advantage of shorter circuit depth in circuit implementation. One can also use the circuit structure in Fig. 6 to benchmark the noise channel.

3. Fidelities of dressed and twirling gates

In this part, we show the full results of the dressed fidelity, or the fidelity of the composite noise channel of the target gate and the twirling group, the fidelity of the twirling groups, and the fidelity of the target gates. Depending on the twirling group and the SPAM setting, we have five benchmarking methods:

(1) Pauli, random SPAM.

(2) CZ dihedral group (for the CS gate) or $\langle C^{n-1}Z, \cdots, CZ, X, Z\rangle$ (for the C^nZ gate), random SPAM.

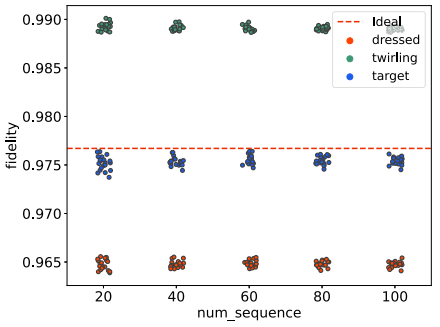
(3) CNOT dihedral group, two SPAM.

(4) $\langle C^{n-1}Z, \cdots, CZ, X, S\rangle$, ZX-SPAM.

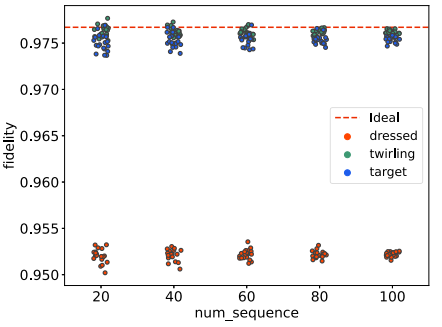
(5) CNOT dihedral group, ZX-SPAM.

For simplicity, the CZ dihedral group or $\langle C^{n-1}Z, \cdots, CZ, X, S\rangle$ is denoted as CZD. The group $\langle C^{n-1}Z, \cdots, CZ, X, Z\rangle$ is denoted as CZP. The CNOT dihedral group is denoted as CXD.

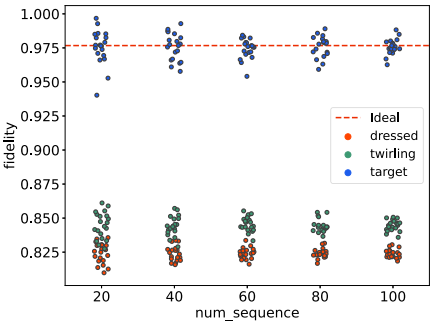
Recall that the benchmarking results in the main text cover the CS, CCZ, $CCCZ$, and $CCCCZ$ gates. We list the



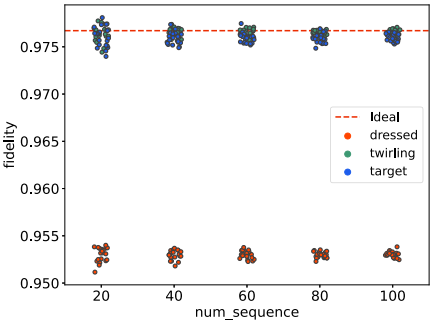
(a) CCZ, Pauli, random SPAM



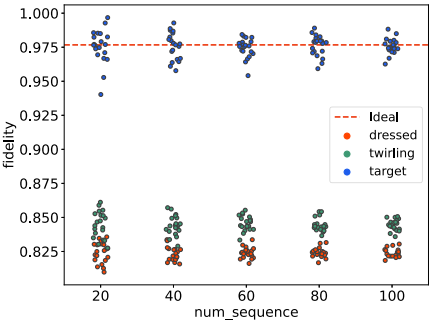
(b) CCZ, CZP, random SPAM



(c) CCZ, CXD, two SPAM

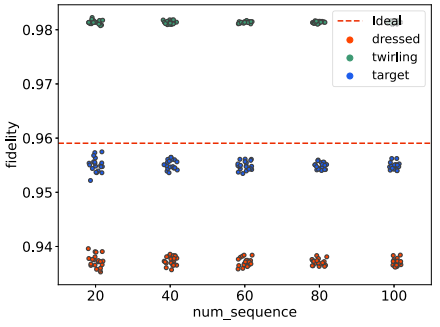


(d) CCZ, CZD, ZX-SPAM

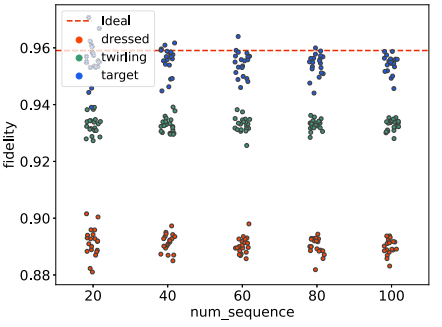


(e) CCZ, CXD, ZX-SPAM

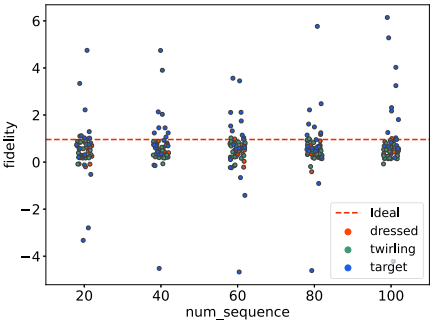
FIG. 14. The results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the CCZ gate with five benchmarking methods. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity.



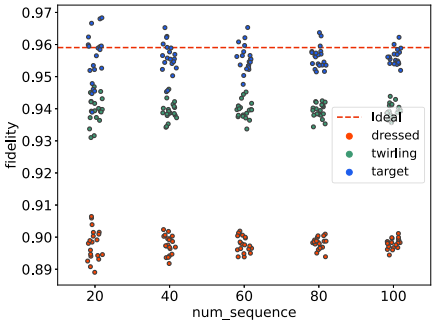
(a) $CCCZ$, Pauli, random SPAM



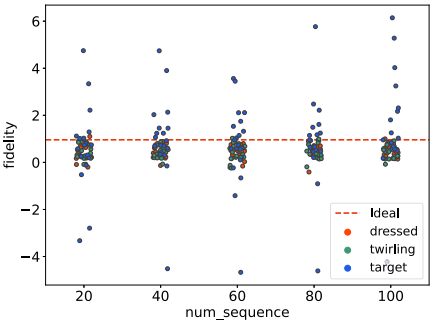
(b) $CCCZ$, CZP, random SPAM



(c) $CCCZ$, CXD, two SPAM

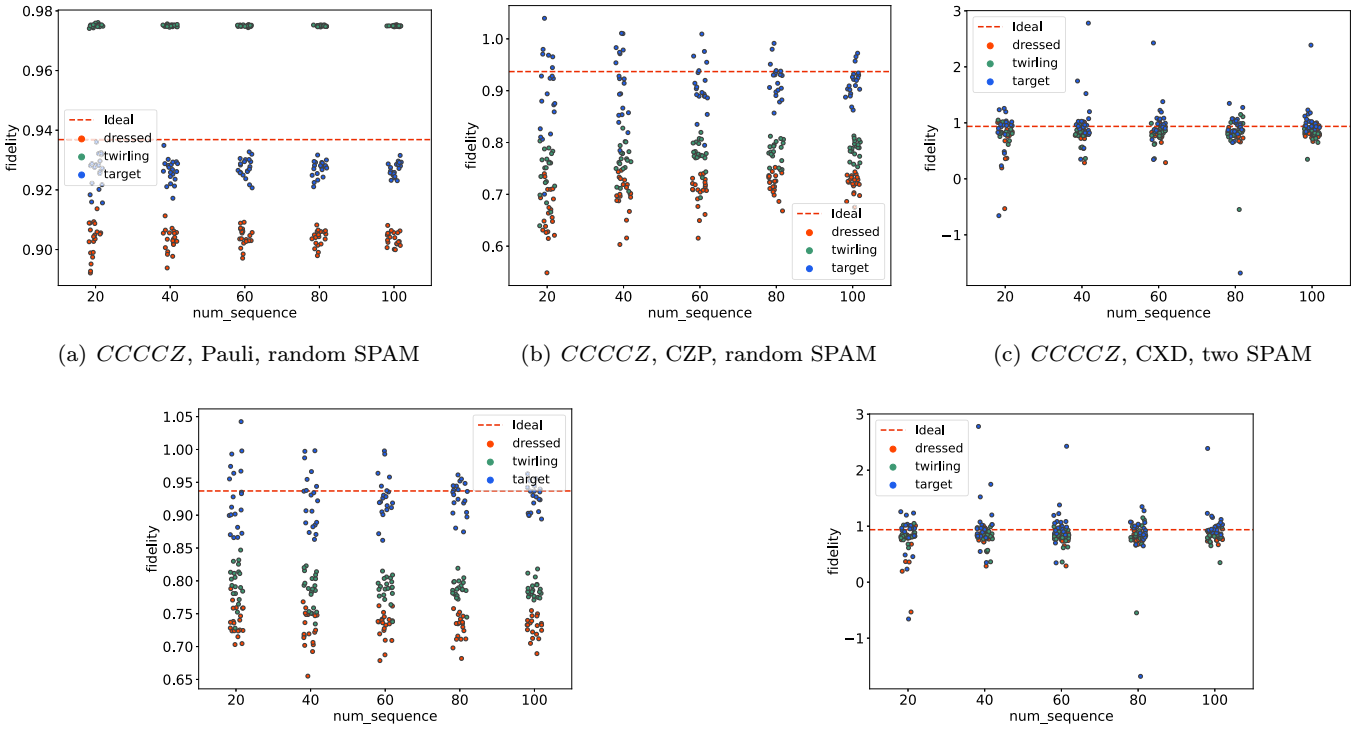


(d) $CCCZ$, CZD, ZX-SPAM



(e) $CCCZ$, CXD, ZX-SPAM

FIG. 15. The results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the $CCCZ$ gate with five benchmarking methods. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity.



(a) $CCCCZ$, Pauli, random SPAM (b) $CCCCZ$, CZP, random SPAM (c) $CCCCZ$, CXD, two SPAM

(d) $CCCCZ$, CZD, ZX-SPAM (e) $CCCCZ$, CXD, ZX-SPAM

FIG. 16. The results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the $CCCCZ$ gate with five benchmarking methods. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity.

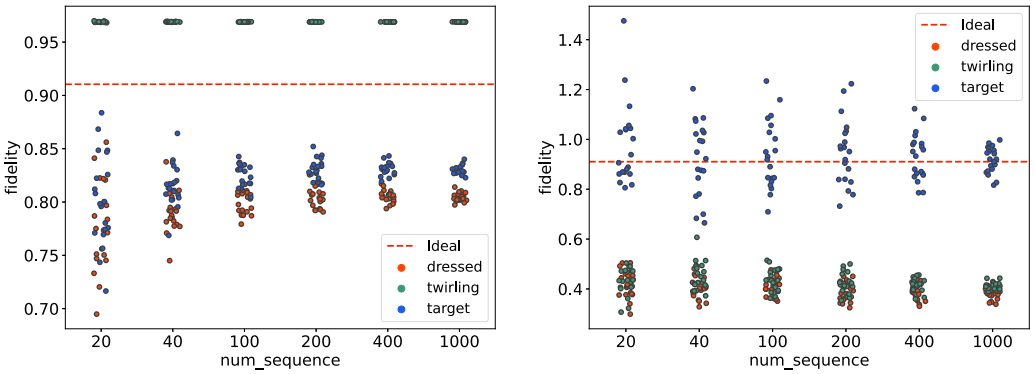
benchmarking results of the four gates with five methods in Figs. 13–16, respectively. We can see that the results of the Pauli group have the minimum fluctuation, thanks to the low error rate of the twirling group. The results of CZP or CZD are effective in a small-scale system below 5 qubits. The CNOT dihedral group fails in benchmarking when the qubit number is no less than 4. Thus, in the main text, we do not present associated results.

4. Simulation results of C^5Z

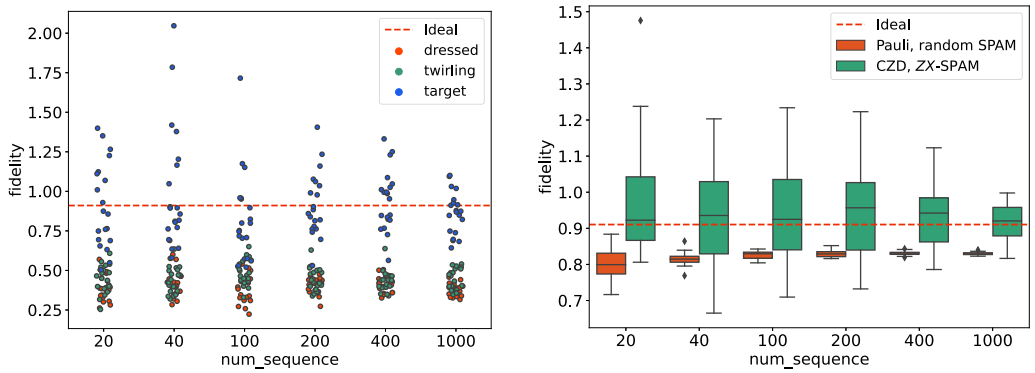
In this part, we present the benchmarking results of C^5Z, with the Pauli group, $\langle C^4Z, \cdots, CZ, X, S\rangle$, and $\langle C^4Z, \cdots, CZ, X, Z\rangle$. The latter two are labeled with CZD and CZP, respectively. The results are shown in Fig. 17. It can be seen that due to a large gate-dependent noise, the results of the latter two methods fluctuate a lot. Nonetheless, from Fig. 17(d), we can still observe that though the variance of the results associated with the Pauli group is lower, the bias associated with CZD is smaller. Due to a large fluctuation in the results, we increased the maximum number of sampled sequences from 100 to 1000 compared to previous simulations. One can see that when the number of random sequences increase from 20 to 1000, the fluctuation of results corresponding to CZD can decrease. In contrast, the bias of the results associated with the Pauli group does not change with the number of sampled sequences increase.

5. Simulation results with noiseless twirling groups

To further understand how the choice of the twirling group influences the final benchmarking results, we also simulate the benchmarking procedure with the noiseless twirling group. The noise model of the target gate does not change and is still given in Appendix C 1. In Figs. 18 and 19, we present the results of benchmarking the CS and CCZ gates with five benchmarking methods, which are introduced in Appendix C 3. It can be seen that the methods "CXD, two SPAM," "CZD, ZX-SPAM," and "CXD, ZX-SPAM" can both estimate fidelity unbiasedly. The results of the methods "Pauli, random SPAM" and "CZD (CZP), random SPAM" deviate from the ideal value, and the results of "Pauli, random SPAM" deviate more. The deviation of "Pauli, random SPAM" and "CZD (CZP), random SPAM" can be explained by the effect of random SPAM settings. The randomly sampled SPAM introduces a bias to the fidelity estimation. Meanwhile, the weak twirling 'ability' of the Pauli group, or the less symmetric of the twirled noise channel associated with the Pauli group, introduces another bias effect to the fidelity estimation. Thus, the results of the Pauli group deviate the most in the case of the noiseless twirling group. In practice, the Pauli group is the least noisy. Hence, when considering a noisy twirling group, especially in a larger system, benchmarking with the Pauli group can provide relatively more effective fidelity estimation than other groups.

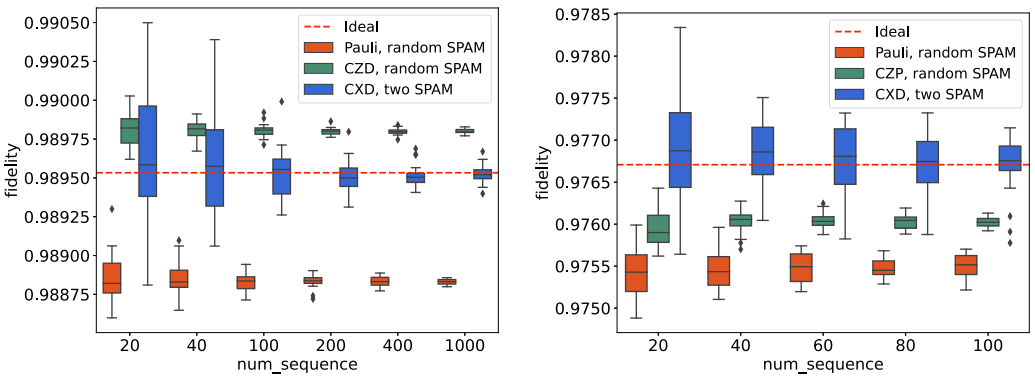


(a) $CCCCCZ$, Pauli, random SPAM (b) $CCCCCZ$, CZD, ZX-SPAM



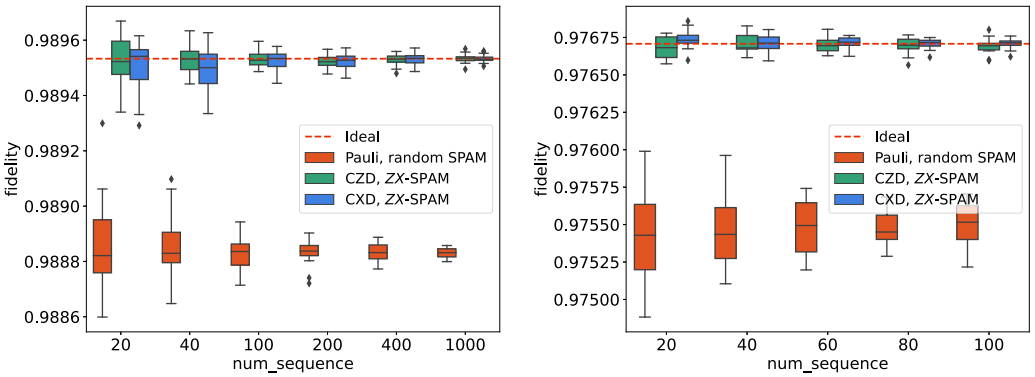
(c) $CCCCCZ$, CZP, random SPAM (d) $CCCCCZ$, Pauli vs CZD

FIG. 17. Panels (a), (b), and (c) show the results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the $CCCCCZ$ gate with three benchmarking methods. The three methods are the Pauli group with a random SPAM, $\langle C^4Z, \cdots, CZ, X, S\rangle$ with the ZX-SPAM, and $\langle C^4Z, \cdots, CZ, X, Z\rangle$ with a random SPAM. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity. The maximum number of sampled sequences is 1000. Panel (d) compares the target gate fidelities associated with the Pauli group and $\langle C^4Z, \cdots, CZ, X, S\rangle$. It can be seen that the variance of the results corresponding to the Pauli group is lower and the bias associated with $\langle C^4Z, \cdots, CZ, X, S\rangle$ is smaller.



(a) CS (b) CCZ

FIG. 18. Benchmarking results for the CS and CCZ gates in panels (a) and (b), respectively, with the optimal twirling group, the Pauli group, and the CNOT dihedral group in the case of noiseless twirling group. The optimal twirling group is the CZ dihedral group $\langle CZ, Z, S\rangle$ for CS, and the CZ Pauli group $\langle CZ, Z, P\rangle$ for CCZ. The red dashed line is the theoretical value of the noise channel fidelity. Each box plot contains 20 fidelities, and each fidelity is estimated with circuit depths $\{2, 4, 6, 8, 10\}$, and the total number of different gate sequences for each depth is specified by the horizontal axis. Here, for the Pauli group and the optimal group, we randomly sample and estimate 20 different diagonal terms of the twirled noise channel. We mark this setting with "random SPAM" on the label. For the CNOT dihedral group, we estimate only two different diagonal terms of the twirled noise channel. We mark this setting with 'two SPAM' on the label. Each SPAM setting prepares an eigenstate of a Pauli observable with eigenvalue 1 and measures this Pauli observable. In "two SPAM," the two Pauli observables are chosen as $Z^{\otimes N}$ and $X^{\otimes N}$.



(a) CS (b) CCZ

FIG. 19. Benchmarking results for the CS and CCZ gates in panels (a) and (b), respectively, with the Pauli group, the CZ dihedral group, and the CNOT dihedral group in the case of the noiseless twirling group. The red dashed line is the theoretical fidelity value. Each box plot contains 20 fidelities. The setting of circuit depths and sampling is the same as in Fig. 18. Nonetheless, for the CZ dihedral group and the CNOT dihedral group, we adopt the SPAM setting and the benchmarking procedure of Box 1. We mark this setting with "ZX-SPAM" on the label.

[1] J. Emerson, R. Alicki, and K. Zyczkowski, Scalable noise estimation with random unitary operators, J. Opt. B **7**, S347 (2005).

[2] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Symmetrized characterization of noisy quantum processes, Science **317**, 1893 (2007).

[3] L. M. K. Vandersypen and I. L. Chuang, Nmr techniques for quantum control and computation, Rev. Mod. Phys. **76**, 1037 (2005).

[4] S. Chu, Cold atoms and quantum control, Nature (London) **416**, 206 (2002).

[5] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Phys. Rev. A **77**, 012307 (2008).

[6] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011).

[7] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A **85**, 042311 (2012).

[8] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, Nat. Phys. **16**, 1184 (2020).

[9] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).

[10] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, ACM Trans. Quantum Comput. **1**, 1 (2020).

[11] R. Harper, W. Yu, and S. T. Flammia, Fast estimation of sparse quantum noise, PRX Quantum **2**, 010322 (2021).

[12] S. T. Flammia and R. O'Donnell, Pauli error estimation via Population Recovery, Quantum **5**, 549 (2021).

[13] C. Rouzé and D. S. Frana, Efficient learning of the structure and parameters of local Pauli noise channels, arXiv:2307.02959.

[14] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, Nat. Commun. **10**, 5347 (2019).

[15] Y. Zhang, W. Yu, P. Zeng, G. Liu, and X. Ma, Scalable fast benchmarking for individual quantum gates with local twirling, Photon. Res. **11**, 81 (2023).

[16] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, Scalable randomised benchmarking of non-Clifford gates, npj Quantum Inf. **2**, 16012 (2016).

[17] J. Claes, E. Rieffel, and Z. Wang, Character randomized benchmarking for non-multiplicity-free groups with applications to subspace, leakage, and matchgate randomized benchmarking, PRX Quantum **2**, 010351 (2021).

[18] J. Helsen, S. Nezami, M. Reagor, and M. Walter, Matchgate benchmarking: Scalable benchmarking of a continuous family of many-qubit gates, Quantum **6**, 657 (2022).

[19] S. Garion, N. Kanazawa, H. Landa, D. C. McKay, S. Sheldon, A. W. Cross, and C. J. Wood, Experimental implementation of non-Clifford interleaved randomized benchmarking with a controlled-s gate, Phys. Rev. Res. **3**, 013204 (2021).

[20] L. B. Nguyen, Y. Kim, A. Hashim, N. Goss, B. Marinelli, B. Bhandari, D. Das, R. K. Naik, J. M. Kreikebaum, A. N. Jordan, D. I. Santiago, and I. Siddiqi, Programmable Heisenberg interactions between floquet qubits, Nat. Phys. **20**, 240 (2024).

[21] Y. Kim, A. Morvan, L. B. Nguyen, R. K. Naik, C. Jünger, L. Chen, J. M. Kreikebaum, D. I. Santiago, and I. Siddiqi, High-fidelity three-qubit Toffoli gate for fixed-frequency superconducting qubits, Nat. Phys. **18**, 783 (2022).

[22] P. Shor, Algorithms for quantum computation: Discrete logarithms and factoring, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE, New York, 1994), pp. 124–134.

[23] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the 28th Annual ACM Symposium on Theory of Computing* (ACM, New York, 1996), pp. 212–219.

[24] A. M. Childs and N. Wiebe, Hamiltonian simulation using linear combinations of unitary operations, Quantum Inf. Comput. **12**, 901 (2012).

[25] M. Rossi, M. Huber, D. Bruß, and C. Macchiavello, Quantum hypergraph states, New J. Phys. **15**, 113022 (2013).

[26] A. Fedorov, L. Steffen, M. Baur, M. P. da Silva, and A. Wallraff, Implementation of a Toffoli gate with superconducting circuits, Nature (London) **481**, 170 (2012).

[27] S. Li, A. D. Castellano, S. Wang, Y. Wu, M. Gong, Z. Yan, H. Rong, H. Deng, C. Zha, C. Guo, L. Sun, C. Peng, X. Zhu, and J.-W. Pan, Realisation of high-fidelity nonadiabatic CZ gates with superconducting qubits, npj Quantum Inf. **5**, 84 (2019).

[28] T. Monz, K. Kim, W. Hänsel, M. Riebe, A. S. Villar, P. Schindler, M. Chwalla, M. Hennrich, and R. Blatt, Realization of the quantum Toffoli gate with trapped ions, Phys. Rev. Lett. **102**, 040501 (2009).

[29] H. Levine, A. Keesling, G. Semeghini, A. Omran, T. T. Wang, S. Ebadi, H. Bernien, M. Greiner, V. Vuletić, H. Pichler, and M. D. Lukin, Parallel implementation of high-fidelity multiqubit gates with neutral atoms, Phys. Rev. Lett. **123**, 170503 (2019).

[30] A. Paetznick and B. W. Reichardt, Universal fault-tolerant quantum computation with only transversal gates and error correction, Phys. Rev. Lett. **111**, 090505 (2013).

[31] T. Jochym-O'Connor and T. J. Yoder, Four-dimensional toric code with non-Clifford transversal gates, Phys. Rev. Res. **3**, 013118 (2021).

[32] Y.-F. Wang, Y. Wang, Y.-A. Chen, W. Zhang, T. Zhang, J. Hu, W. Chen, Y. Gu, and Z.-W. Liu, Efficient fault-tolerant implementations of non-Clifford gates with reconfigurable atom arrays, arXiv:2312.09111.

[33] G. Liu, X. Zhang, and X. Ma, Classically replaceable operations, Quantum **6**, 845 (2022).

[34] T. Baumgratz, M. Cramer, and M. B. Plenio, Quantifying coherence, Phys. Rev. Lett. **113**, 140401 (2014).

[35] E. Chitambar and G. Gour, Critical examination of incoherent operations and a physically consistent resource theory of quantum coherence, Phys. Rev. Lett. **117**, 030401 (2016).

[36] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, General framework for randomized benchmarking, PRX Quantum **3**, 020357 (2022).

[37] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, A new class of efficient randomized benchmarking protocols, npj Quantum Inf. **5**, 71 (2019).

[38] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).

[39] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, Nat. Rev. Phys. **2**, 382 (2020).

[40] C. W. Curtis and I. Reiner, *Representation theory of Finite Groups and Associative Algebras*, Vol. 356 (American Mathematical Society, Providence, RI, 1966).

[41] W. Burnside, *Theory of Groups of Finite Order* (Cambridge University Press, Cambridge, UK, 1911).

[42] S. Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, and L. Jiang, The learnability of Pauli noise, Nat. Commun. **14**, 52 (2023).

[43] J. Chen, D. Ding, and C. Huang, Randomized benchmarking beyond groups, PRX Quantum **3**, 030320 (2022).

[44] M. Heinrich, M. Kliesch, and I. Roth, Randomized benchmarking with random quantum circuits, arXiv:2212.06181.

[45] Y. Gu, W.-F. Zhuang, X. Chai, and D. E. Liu, Benchmarking universal quantum gates via channel spectrum, Nat. Commun. **14**, 5880 (2023).

[46] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[47] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A. Werner, J. Eisert, and I. Roth, Estimating gate-set properties from random sequences, Nat. Commun. **14**, 5039 (2023).

[48] Y. Wang, G. Liu, Z. Liu, Y. Tang, X. Ma, and H. Dai, Robust estimation of nonlinear properties of quantum processes, Phys. Rev. A **110**, 032415 (2024).

[49] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A **60**, 1888 (1999).

[50] B. Steinberg, *Representation Theory of Finite Groups: An Introductory Approach* (Springer, Berlin, 2012).

[51] W. Fulton and J. Harris, *Representation Theory: A First Course* (Springer, New York, NY, 2004).

[52] A. Winter and D. Yang, Operational resource theory of coherence, Phys. Rev. Lett. **116**, 120404 (2016).

[53] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Robust extraction of tomographic information via randomized benchmarking, Phys. Rev. X **4**, 011050 (2014).

[54] T. G. d. Brugière, M. Baboulin, B. Valiron, S. Martiel, and C. Allouche, Reducing the depth of linear reversible quantum circuits, IEEE Trans. Quantum Eng. **2**, 1 (2021).