Linear cross-entropy benchmarking with Clifford circuits

Jianxin Chen,[1] Dawei Ding,[2] Cupjin Huang,[1] and Linghang Kong[3]
[1]*Alibaba Quantum Laboratory, Alibaba Group USA, Bellevue, Washington 98004, USA*
[2]*Alibaba Quantum Laboratory, Alibaba Group USA, Sunnyvale, California 94085, USA*
[3]*Alibaba Quantum Laboratory, Alibaba Group, Hangzhou, Zhejiang 311121, People's Republic of China*

(Received 16 May 2023; revised 25 August 2023; accepted 1 November 2023; published 20 November 2023)

With the advent of quantum processors exceeding 100 qubits and the high engineering complexities involved, there is a need for holistically benchmarking the processor to have quality assurance. Linear cross-entropy benchmarking (XEB) has been used extensively for systems with 50 or more qubits but is fundamentally limited in scale due to the exponentially large computational resources required for classical simulation. In this work we propose conducting linear XEB with random Clifford circuits of constant to logarithmic depth, a scheme we call *Clifford XEB*. Since Clifford circuits can be simulated in polynomial time, Clifford XEB can be scaled to much larger systems. To validate this claim, we run numerical simulations for the classes of Clifford circuits we propose with noise and observe exponential decays. When noise levels are low, the decay rates are well correlated with the noise of each cycle assuming a multiplicative error accumulation, i.e., where the fidelities of individual gates multiply. We perform simulations of systems up to 1225 qubits, where the classical processing task can be easily dealt with by a workstation. Furthermore, using the theoretical guarantees in Chen *et al.* [PRX Quantum **3**, 030320 (2022)], we prove that Clifford XEB with our proposed Clifford circuits must yield exponential decays under a general error model for sufficiently low errors. Our theoretical results explain some of the phenomena observed in the simulations and shed light on the behavior of general linear XEB experiments.

DOI: 10.1103/PhysRevA.108.052613

I. INTRODUCTION

Quantum computers are no longer theoretical constructs or even small-scale, proof-of-concept devices. There are already multiple processors with over 50 qubits [1,2], and recently a processor with more than 100 qubits has been built [3,4]. With this number of qubits comes a litany of engineering and operating challenges, including the overwhelming number of control lines and pulse generators required to perform gates and measurements. There are also challenges in processor design, including significant crosstalk between qubits and poor two-qubit gate connectivity. Given all these challenges, there is a demand for benchmarking these processors to have assurance that the algorithms we run will give accurate results. In addition, the figure of merit provided by such a benchmarking scheme can serve as an essential guide for hardware iteration to build ever higher-quality processors. Although common methods such as single- and two-qubit randomized benchmarking (RB) can provide useful information [5–18], these methods are not sensitive to aggregate effects such as the difference between calibrating qubits on a processor individually or simultaneously [13]. To capture such system-level effects, holistic benchmarks which should involve all or a large number of qubits on the processor are needed. Such benchmarks can also provide information about whether there are correlated errors in the system by comparing the results to the multiplicative error accumulation (MEA) model, where the fidelity of each circuit component multiplies.[1] This is crucial for realizing fault-tolerant quantum computation.

To holistically benchmark modern quantum processors with tens of qubits, various schemes have been developed, most notably linear cross-entropy benchmarking (linear XEB). Linear XEB was originally proposed for the "quantum supremacy" experiment [1], where it was used to characterize increasingly larger quantum circuits so as to extrapolate the error of the 20-cycle Sycamore circuit. To implement linear XEB, we run quantum circuits with random layers of gates and perform a computational basis measurement. We next classically compute the probability of the bit string we measure. We then repeat, taking an overall average. This average, up to constants, is the linear XEB measure. It has been experimentally and numerically observed that this measure exponentially decays with the number of cycles for a noisy circuit, and this decay exponent is proposed as a measure of gate quality [1,2,19]. Although originally conceived to support the "quantum supremacy" claim, linear XEB has become a benchmarking scheme in its own right [1,2,20,21]. Linear XEB has the advantage of requiring only a shallow circuit, which is easy to implement on current processors. However, if we want to benchmark larger devices, linear XEB has an obvious Achilles' heel. Namely, the random gate set is chosen such that the classical processing needed for linear XEB, that is, simulating the resulting random quantum circuits, scales poorly with the number of qubits. Indeed, by the quantum

[1]This is referred to as the digital error model in [1].

supremacy claim itself, linear XEB *by design* cannot benchmark a processor with more than a few tens of qubits [1].[2] Therefore, if we want to take advantage of linear XEB's low overhead to benchmark devices but keep apace with current hardware development, we need an alternative benchmarking scheme that can scale to larger qubit numbers.

In this work, we propose a scalable benchmarking scheme which replaces the random circuits in linear XEB with random Clifford circuits of constant to logarithmic depth for which we can also theoretically guarantee an exponential decay. The scalability naturally follows from the fact that Clifford gates can be efficiently classically simulated [22]. We call our scheme *Clifford XEB*. Of course, by replacing general gates with Clifford gates, we lose the direct benchmarking of non-Clifford gates that linear XEB provides. However, in most quantum computing platforms, gate errors do not discriminate between Clifford and non-Clifford gates. As a result, the benchmarking results on Clifford circuits should in principle indicate the performance of non-Clifford gates as well. Compared to Clifford RB [17,18,23] which requires $\Omega(n/\log n)$ depth to compile an n-qubit Clifford gate [24], Clifford XEB retains the convenience of general linear XEB in requiring only shallow circuits. And due to the ease of classical simulation, we can benchmark much larger devices. Moreover, given the prevalence of Clifford RB and linear XEB, it is a negligible overhead to adopt Clifford XEB as a new benchmarking scheme.

To demonstrate the scalability of our scheme, in Sec. II we perform numerical simulations of the random Clifford circuits we propose on various topologies with depolarizing noise. We simulate circuits with up to 225 qubits. For noise levels comparable to what is possible in current hardware, we can see clear exponential decays whose rates are consistent with MEA. We also conduct simulations on the two-dimensional (2D) grid topology with 1225 qubits, which is almost three times the number of qubits of the largest device as of this writing [4], to further showcase the scalability of Clifford XEB. On top of the promising numerical results, in Sec. III we also prove that Clifford XEB for the classes of circuits we propose yields an exponential decay under a general error model for sufficiently low error. This is necessary since we *a priori* do not know if we would even measure an exponential decay for every experimental setting we encounter (although this can be shown for special error models [19,21]). In addition to providing this assurance, our proof also explains some of the phenomena we observe in the simulations, such as the linear XEB measure rapidly decaying for low cycle numbers before entering a smooth exponential decay. Through our theoretical results we provide concrete guidelines for how experiments, both Clifford XEB *and* conventional linear XEB, should be interpreted so as to correctly extract information about gate errors. In Sec. IV, we conclude with a discussion.

[2]Readers who are familiar with the "quantum supremacy" paper may point out that the paper's method of using patch circuits or elided circuits can push the qubit number further. However, using patch circuits is effectively benchmarking two separate processors. As for elided circuits, the complexity of classical simulation still scales poorly.

II. THE CLIFFORD XEB SCHEME

In this section we give an explicit description of the Clifford XEB scheme and perform numerical simulations showing its viability for benchmarking quantum processors with more than a thousand qubits.

A. Description of the scheme

Although we focus on Clifford XEB, our experimental protocols and performance guarantees equally apply to general linear XEB. Here we first briefly introduce linear XEB before investigating Clifford XEB in more detail. Consider a quantum computer with n qubits. Let $S \subset \mathrm{SU}(2^n)$ be a subset of the special unitary group and let μ be a probability distribution on S. The implementation map $\phi : S \to \mathcal{C}(2^n)$, where $\mathcal{C}(2^n)$ is the set of quantum channels on n qubits, defines the noisy implementation of gates on the system. This in particular means that the noise for each element is independent of the previous elements, that is, the noise is Markovian. Assume the initial state is ρ_0 and final measurement positive operator-valued measure (POVM) is $M = \{M_x\}_{x\in\{0,1\}^n}$, which are noisy realizations of some desired initial state and measurement. Then, a linear XEB experiment is defined as follows.

Definition 1 (Linear XEB). A linear XEB scheme with parameters $(S, \mu, \phi, M, \rho_0)$ is given by the following procedure:

(1) For a given number of cycles m, sample m i.i.d. quantum gates $g_1, \ldots, g_m$ from S according to the distribution μ.

(2) Initialize the system to ρ_0 and apply $\phi(g_1), \phi(g_2), \ldots, \phi(g_m)$ to the system.

(3) Measure the system under a POVM $\{M_x\}_{x\in\{0,1\}^n}$ and get a binary string x. Calculate the probability of getting x for the ideal circuit: $p_x(m) = |\langle x|g_m \ldots g_2 g_1|0^n\rangle|^2$.

(4) Repeat steps 2 to 3 multiple times for this sampled circuit and calculate the average of $p_x(m)$ over sampled strings x.

(5) Repeat steps 1 to 4 multiple times and let $\hat{p}(m)$ be the average of $\mathbb{E}_x[p_x(m)]$ over different circuits. Let $\hat{q}(m) = -1 + 2^n\hat{p}(m)$.

(6) Repeat steps 1 to 5 for number of cycles $m_1, \ldots, m_k$. The returned results are $\hat{q}(m_1), \ldots, \hat{q}(m_k)$.

We can see that $\hat{q}(m)$ is an unbiased estimator of the following quantity:

$$q_R(m) = -1 + 2^n \mathbb{E}_{g_1,\ldots,g_m\sim\mu} \times \left[\sum_{x\in\{0,1\}^m} |\langle x|g_m \ldots g_2 g_1|0^n\rangle|^2 \tilde{p}(x)\right], \quad (1)$$

where $\tilde{p}(x)$ is the probability of measuring the binary string x in the noisy circuit:

$$\tilde{p}(x) = \mathrm{tr}[M_x \phi(g_m) \circ \cdots \circ \phi(g_2) \circ \phi(g_1)(\rho_0)]. \quad (2)$$

We next define Clifford XEB, a special case, illustrated in Fig. 1.

Definition 2 (Clifford XEB). A Clifford XEB scheme with parameters $(S, \mu, \phi, M, \rho_0)$ is a linear XEB scheme, where $S \subseteq \mathrm{Cl}(2^n)$ is a subset of the Clifford group.

That is, this Clifford XEB only differs from conventional linear XEB in that we require S to be Clifford. Again, we

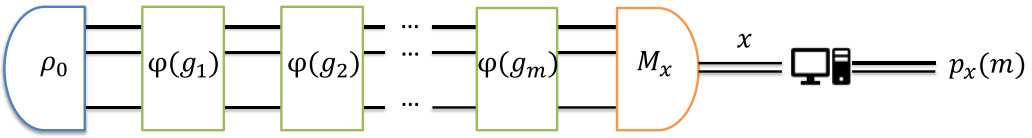

FIG. 1. Illustration of one iteration of a Clifford XEB experiment. We start with an initial state, apply m cycles of noisy implementations of Clifford elements, and finally perform a measurement. The measured bit string's noiseless probability is then classically computed. This value is averaged over multiple measurements and random circuits.

do this since the noiseless distributions resulted by Clifford circuits can be efficiently computed classically [22]. Also note that our definition of cycles is slightly different from that of [1]. There they allow heterogeneous circuit geometries for every cycle, while we require that the gates in each cycle is sampled from the same distribution.

B. Numerical simulations

We numerically study the behavior of Clifford XEB for various connectivity topologies to demonstrate its wide applicability. The topologies we consider include a one-dimensional (1D) chain, a 2D grid, and a star topology, i.e., there is a center qubit that is connected to all the other qubits, to implement the construction in [25], which we will refer to as the *Clifford approximate twirl*. We describe the particular distribution over Cliffords for each topology, illustrated in Fig. 2.

(1) For the 1D chain, each cycle consists of two layers of two-qubit gates each preceded by a layer of single-qubit gates, in a brickwork manner. The single-qubit gate layer involves independent and identically distributed (i.i.d.) random single-qubit Clifford gates on each qubit. The two-qubit layers are controlled NOT (CNOT) gates: the first connecting qubits $(0, 1), (2, 3), \ldots, (2k, 2k+1), \ldots$ and the second connecting $(1, 2), (3, 4), \ldots, (2k+1, 2k+2), \ldots$.

(2) For the 2D grid, we take motivation from the random circuits on the Sycamore processor in [1]. Each cycle consists of a single-qubit gate layer consisting of i.i.d. single-qubit Cliffords on every qubit, followed by a two-qubit gate layer consisting of CNOT gates. The structure of the CNOT's is chosen from the configurations A, B, C, and D in [[1], Fig. 3] with equal probability.[3]

(3) We implement a specific family of random Clifford circuits called Clifford approximate twirls [25] on a star topology. Each cycle in this construction consists of two repeated circuits, each circuit involving three layers of two-qubit gates, interleaved by four layers of single-qubit gates. See Appendix A for the detailed construction.

The numerical experiments are run using the stabilizer-based Clifford circuit simulator Stim [26] on a 40-core Intel Xeon Platinum machine with 96 GB memory, with source code slightly modified[4] to accelerate the calculation of the noiseless probability distribution. For each topology, we run experiments with different qubit numbers and noise levels. We pick qubit numbers 25, 100, and 225, which are respectively 5×5, 10×10, and 15×15 on the 2D grid. As for noise levels, we assume depolarizing noise for single- and two-qubit gates of magnitudes $(10^{-5}, 10^{-4})$, $(10^{-4}, 10^{-3})$, and $(10^{-3}, 10^{-2})$, respectively. These error rates are similar to what is achievable in current hardware. We also simulate the noiseless circuit. Each experiment is run for 2 to 50 cycles, where for each cycle number 3000 random circuits are generated, each sampled 1×10^6 times. The results are plotted in Fig. 3. For illustration purposes, we additionally run experiments on a 4×4 2D grid, for single- and two-qubit gates with noise levels $(10^{-3}, 10^{-2})$, $(5 \times 10^{-4}, 5 \times 10^{-3})$, $(2 \times 10^{-4}, 2 \times 10^{-3})$, $(10^{-4}, 10^{-3})$. A gate with noise level p undergoes a depolarizing channel with probability p. Each experiment is run for 1 to 100 cycles, where for each cycle number 3×10^6 random circuits are generated, each sampled 100 000 times. The result is plotted in Fig. 4. All of the experiments added together only take a total of several hours on the aforementioned workstation. In comparison, conventional linear XEB is already infeasible for just tens of qubits. Indeed, because of this ease of simulation, Clifford XEB provides a rich playground for studying the behavior of linear XEB in larger systems. Our findings can be used as general guidelines for using linear XEB as a benchmarking scheme, *including scenarios in which a naïve exponential curve fitting would fail to reflect the correct fidelity*. We summarize our findings with two main observations:

a. Convergence and two-phase behavior. We see a universal behavior of linear XEB with increasing cycle number: the noiseless value always converges to 1,[5] while the noisy value converges to 0.[6] This is the same as the experimental findings in [1]. Intuitively, for sufficiently many cycles, the random Clifford circuit will converge to a uniform distribution over the Clifford group [27], while a noisy implementation will eventually lose all information of the input. This explains the final convergence values. A more rigorous proof is given in Appendixes B 1 and B 2.

[3]In the Sycamore random circuit sampling experiments, the two-qubit gates in different cycles follow a fixed pattern ABCDCDAB. We do not do this because we require every cycle to have the same distribution.

[4]Explicitly, it gives a different canonical presentation of the stabilizer groups.

[5]Technically, the noiseless linear XEB value converges to $\frac{D-1}{D+1}$ as shown in Appendix B 1, D being the dimension of the quantum system. In the rest of the paper we take the approximation $\frac{D-1}{D+1} \approx 1$ unless explicitly stated.

[6]This is not visually apparent in some of the plots due to the low noise levels. However, the numerical data support this conclusion.

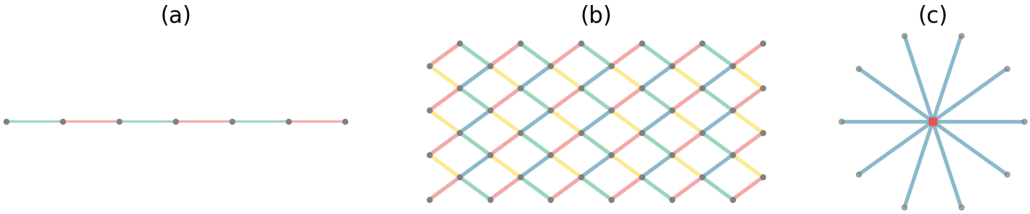

FIG. 2. Illustration of different quantum device topologies we consider in our paper: (a) 1D chain, (b) 2D grid, and (c) star topology. In (a) and (b), qubit connections with different colors indicate different layers of parallel two-qubit gates to be applied. The Clifford approximate twirl can be naturally applied on a star topology depicted in (c)

The curves also exhibit a two-phase behavior: In the first phase, which we call the *scrambling phase* in reference to the connection between random circuits and quantum scrambling [27–34], both the noiseless and noisy linear XEB values decay quickly from being exponentially large to a constant value. The second phase, which we call the *decaying phase*, the noiseless linear XEB value converges to 1, while the noisy values undergo single exponential decays (which would be a linear decay on the semilog plot). We show in Fig. 4 that exponential fits of data in the second phase agree well with the experimental results. Intuitively, only the data in the second phase are of relevance to benchmarking, as the first phase is characterized by the *scrambling* of quantum information, rather than the *loss* of it.

One way of interpreting the decay rate is to use it as a proxy for the fidelity of a cycle, which in turn can be estimated by assuming MEA, that is, the fidelity of the circuit is the product of the fidelities of its constituent gates. To see if this applies to the numerical experiments, we extract the decay rates by linearly fitting the curve in the second phase, manually choosing the range of the number of cycles. We then compare them with the predicted fidelity according to MEA. We compare the two values under different settings in Fig. 5, and we give the detailed data processing procedure in Appendix C. We see that the MEA model fidelity is well correlated to the linear XEB decay rate when the noise level is low, but not when the noise level is too high. This is because even in the decaying phase, the distribution of the ideal circuit converges to the uniform distribution on all Cliffords at a certain rate, resulting in a mixing term. This mixing term is plotted as a red dashed-dotted line in Fig. 4. As a result, when the gate error exceeds a certain threshold, the decay of the linear XEB value no longer represents the fidelity but is dominated by this slower mixing of the ideal circuit.

To verify this intuition, for each circuit configuration (topology and number of qubits) we sampled 300 000 random circuits to better evaluate the convergence rate of the noiseless linear XEB value towards 1, which we call the *noiseless scrambling rate* (NSR). We do this because of the quick decay of the noiseless value and the high variance of Clifford XEB (see Appendix B 1). We fit the differences of linear XEB values and 1 with an exponential fit to extract the NSRs, and plot them as dashed lines in Fig. 5. It can be observed that the MEA infidelity in general is not correlated with the extracted linear XEB error rate when the former is greater than the NSR. This indicates that the Clifford XEB, and in general linear XEB, *only provides useful information when the noise level is sufficiently small.* Our simulations suggest we can even roughly quantify what that noise level is. As a rule of thumb, we can estimate the convergence rate of the noiseless linear XEB values prior to running linear XEB experiments to ensure that the extracted decay rate from the experiments reflects gate fidelities. To characterize large errors, we could either change the twirling distribution μ for faster mixing (for example, the one given in Appendix A gives a theoretically provable NSR) or benchmark a subset of the qubits on the device to lower the error of each cycle.

b. Fluctuations. Occasionally (e.g., in the middle plot of Sec. II B), there are noticeable bumps in the curves. Moreover, we observe that the red line, corresponding to a two-qubit gate noise level of 10^{-2}, fluctuates heavily when the linear XEB value is small. We attribute such instabilities to three main causes:

(i) During the scrambling phase, both the expectation and the variance of the noiseless and noisy linear XEB value are large, resulting in significant fluctuations for the noiseless and noisy curves. The middle plot in Sec. II B has such a bump at $m = 20$.

(ii) Unlike non-Clifford scrambling circuits that almost surely result in a measurement distribution obeying Porter-Thomas statistics, the measurement distribution of a Clifford circuit is always uniform on the support. As a result, the noiseless linear XEB value has constant variance with respect to the random circuits chosen, even after the scrambling phase. This is different from doing linear XEB with Haar random gates because the Clifford group forms a unitary three-design but not a four-design. We derive the variance in Appendix B 1. Since the variance is constant, taking a large but constant number of random circuits smooths out the fluctuation. It can be observed in our plot that taking 3000 random circuits per data points effectively resolves this issue.

(iii) The linear XEB values of the noisy curves converge to 0 exponentially with respect to the cycle number. Given a fixed number of samples per random circuit, the absolute error of the linear XEB value is a constant, but the relative error quickly grows as the expectation approaches 0. Similar to [1], sufficiently many random samples need to be taken from each random circuit to ensure an accurate evaluation of the expectation value for linear XEB. This can be observed in the rightmost figure in Sec. II B.

Understanding these sources of instability can help guide Clifford XEB and general linear XEB experiments. The second source only applies to Clifford XEB, but a constant number of random circuits suffices to reduce the variance to a small number. The third source applies for both Clifford XEB and linear XEB, in fact, for any RB-based scheme. The

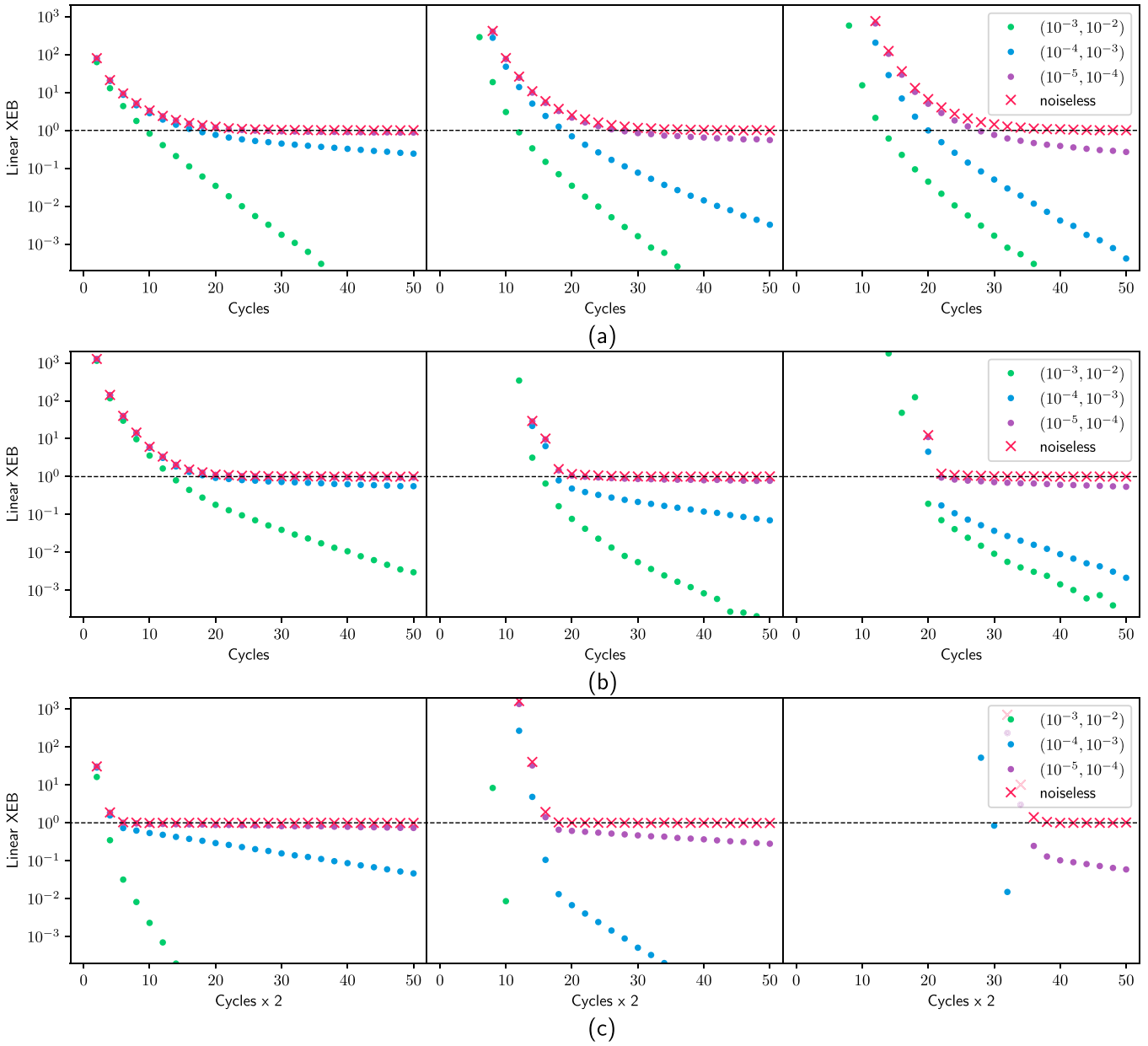

FIG. 3. Linear XEB as a function of the number of cycles on (a) 1D chain, (b) 2D grid, and (c) Clifford approximate twirl (star topology) (the x axis is twice the cycle number since we need to apply the circuit construction twice to get a γ-approximate twirl, see Sec. III for details), with different qubit numbers and noise levels. Each row corresponds to the labeled topology. The number of qubits simulated on each topology is, from left to right, 25, 100, and 225. The horizontal line is linear XEB $= 1$ for reference. The curves show some deviations from an exponential decay. This can be attributed to rapidly decaying transients and statistical fluctuations. The former can be addressed by a sufficient number of cycles, while the latter can be addressed by increasing the number of samples. See main text for further details.

number of samples taken must be sufficiently large such that the exponential decay curves can be confidently recovered. However, note that in our simulations we can often already isolate a clean single exponential decay at moderately small linear XEB values for which we do not need as many samples.

To accurately extract decay rates, one needs to perform Clifford XEB experiments in the decaying phase rather than the scrambling phase. One natural question to ask is at what depths a Clifford XEB scheme enters the decaying phase. Using the results from Sec. III and Appendix B 2, we can give a lower bound on the necessary number of cycles that scales linearly with respect to number of qubits but we observe from our numerical experiments that this only applies to the Clifford approximate twirl. Mixing happens much faster for the 1D chain and 2D grid experiments and seems only weakly dependent on the number of qubits. Although a finer theoretical analysis of the mixing behavior of the noiseless experiment is needed to fully explain this phenomenon, we propose *a general numerical approach to extracting accurate decay rates*:

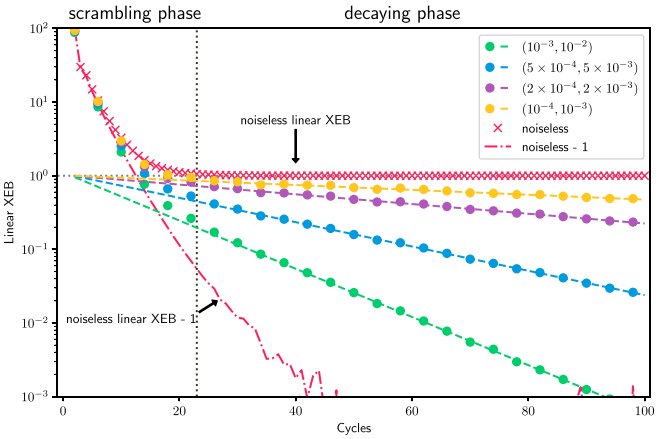

FIG. 4. Illustration of a typical linear XEB experiment on a 4×4 grid. Decay of the noisy Clifford XEB curves can be roughly divided into a scrambling phase and a decaying phase with the threshold being about $m = 22$. We perform a linear fitting for the noisy Clifford XEB curves for $m \in [40, 100]$ and plot the fit as dashed lines. The red crosses illustrate the convergence of the noiseless linear XEB towards 1. In the decaying phase, it can be seen that the noiseless linear XEB value converges to 1 at a certain rate we call the noiseless scrambling rate (NSR) which can computed from the red dashed-dotted line on the bottom. More discussions of the NSR can been found in Sec. II B.

(1) Classically simulate the noiseless experiment to determine both the scrambling phase length and the NSR in the decaying phase. Note that we can perform this simulation offline, prior to any physical experiment.

(2) Check that the linear XEB decay will not be dominated by the mixing term. We can do this by estimating the total infidelity from one cycle of the circuit, for example, by assuming MEA.

(3) Check that we have sufficiently many cycles in the physical experiment to extract accurate decay rates. We can do this by comparing with the scrambling phase length we find in approach 1.

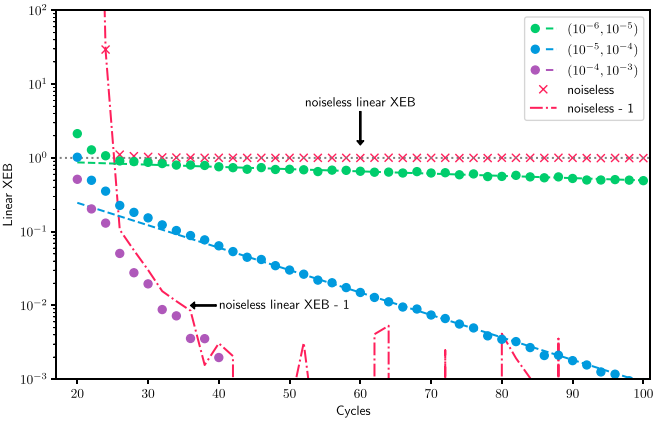

FIG. 6. Clifford XEB on a 2D grid of $35 \times 35 = 1225$ qubits. Data points for curves with noise are evaluated over 3000 random circuits with 100 000 samples each. Data points for the noiseless curve are evaluated over 300 000 random circuits. Predictions assuming MEA are plotted in dashed lines for comparison. The prediction assuming MEA with noise level $(10^{-4}, 10^{-3})$ is beyond the range of the plot. The red dashed-dotted line is used to illustrate the noiseless scrambling rate (NSR). It again suffers from fluctuations with large depths.

Finally, we present the results of Clifford XEB on the 2D grid with $35 \times 35 = 1225$ qubits, the results shown in Fig. 6. The experiments are run with two-qubit noise levels ranging from 10^{-5} to 10^{-3} due to the larger number of qubits. We plot linear XEB for cycle numbers from 20 to 100 because of longer mixing time. We see that, while the 10^{-3} error decay curve is dominated by the noiseless scrambling, Clifford XEB is consistent with MEA when the noise level is below the NSR. This shows that the Clifford XEB scheme can be readily applied to quantum devices with more than 1000 qubits. On the aforementioned workstation, all the noisy experiments took approximately 12 hours to run, and two days and a half to run the noiseless experiments in order to reliably extract the NSR.

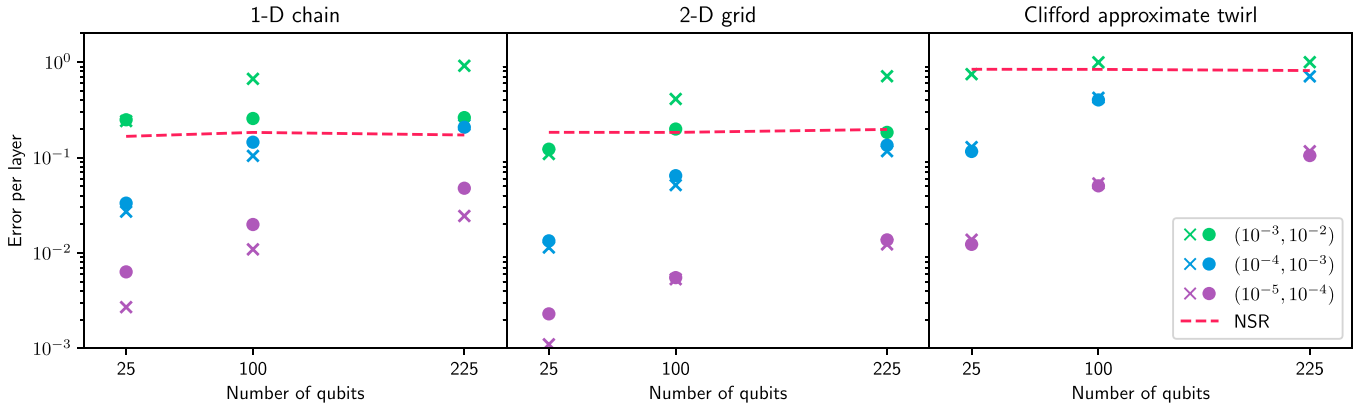

FIG. 5. Comparison between the extracted error per layer from the linear XEB experiments, the error per layer predicted by MEA, and the NSR, under different topologies and number of qubits. The error per layer is defined as $1 - u$, where u is the extracted decay rate for linear XEB experiments, and the predicted fidelity assuming MEA. For each noise level, the crosses ($\times$) are predictions given by MEA, and the dots ($\bullet$) are extracted from the numerical experiments. A few data points are missing due to failure of fitting. As we will discuss below, the MEA infidelity matches the fitted error rate when they are below the NSR.

III. THEORY OF CLIFFORD XEB

In this section we mathematically prove that Clifford XEB with the 1D chain distribution, 2D grid distribution, or the Clifford approximate twirl yields a single exponential decay under a general error model for sufficiently small errors. For the Clifford approximate twirl, this decay is theoretically guaranteed for a qubit number that scales inversely with the infidelity. Our results also support our explanations of phenomena observed in Sec. II B. These technical proofs can be skipped with little effect on the reading of the rest of the paper.

In [35], a generalized framework for RB called universal randomized benchmarking (URB) is proposed, where they prove that for a particular class of schemes known as twirling schemes, an experiment would yield a single exponential decay. Here we show that Clifford XEB with our specific distributions are twirling schemes.

For any $g \in \mathrm{SU}(2^n)$, define $\omega(g)$ as the corresponding quantum channel $\omega(g) : \rho \to g\rho g^\dagger$. We first define a γ-approximate twirl, which is one possible definition of an approximate unitary 2-design [25,35]:

Definition 3. Let μ be a measure on the unitary group and $\mathcal{C}$ be a linear operator on the space of Hermitian matrices. The twirling map

$$\Lambda(\mu) : \mathcal{C} \mapsto \int_{g\sim\mu} dg\, \omega(g^\dagger) \circ \mathcal{C} \circ \omega(g) \tag{3}$$

is a γ-approximate twirl if it satisfies

$$|||\Lambda(\mu) - \Lambda(\mu_H)|||_\diamond \leqslant \gamma, \tag{4}$$

where μ_H is the Haar measure and $|||\cdot|||_\diamond$ is the induced diamond norm:

$$|||\Lambda|||_\diamond := \max_{\mathcal{C}} \frac{\|\Lambda(\mathcal{C})\|_\diamond}{\|\mathcal{C}\|_\diamond}. \tag{5}$$

Alternatively, $\Lambda(\mu)$ is a γ-approximate twirl in spectral norm if it satisfies

$$|||\Lambda(\mu) - \Lambda(\mu_H)|||_2 \leqslant \gamma, \tag{6}$$

where $|||\cdot|||_2$ is the spectral norm of the twirling map treated as a linear operator on superoperators.

A. Clifford approximate twirl

1. Proof of exponential decay

For Clifford approximate twirls, we use the following theorem in [35]:

Theorem 1 (Theorem 8 from [35] and Appendix B 2). Let $(S, \mu, \phi, M, \rho_0)$ be a linear XEB scheme on n qubits. Suppose that the twirling map corresponding to μ is a γ-approximate twirl and that the implementation map satisfies

$$\mathbb{E}_{g\sim\mu}\|\phi(g) - \omega(g)\|_\diamond \leqslant \delta. \tag{7}$$

If $\delta \leqslant \frac{1-\gamma}{11}$, then there exists $A, B \in \mathbb{R}$ and $p \in [1-2\delta, 1]$, such that

$$|q_R(m) - (A + Bp^m)| \leqslant 16 \times 2^n(\gamma + 6\delta)^m. \tag{8}$$

Further, when $p < 1$ we have $A = 0$ and $B = O(1)$.

Note that we give here a general result for linear XEB, that is, S need not be Clifford. The main difference from the statement in [35] is the factor of 2^n in Eq. (8). This arises from the factor of 2^n in the definition of the XEB quantity in Eq. (1) which is used for normalization [1]. However, for large n this is problematic since this would make the error term on the right-hand side in Eq. (8) unacceptably large, considering that $B = O(1)$. To make the error term smaller than the exponential decay $A + Bp^m$ we want to extract, we need the number of cycles m to scale linearly with n: $m = \Omega(n)$.[7] Indeed, we do observe that the mixing time (end of the scrambling phase) is approximately a linear function of the number of qubits in Sec. II B.

Now, we obtained the Clifford approximate twirl construction from [25], where they proved that it constitutes a γ-approximate twirl. Hence, we can directly apply Theorem 1 to Clifford XEB with the Clifford approximate twirl to theoretically guarantee an exponential decay. This guarantee holds under a general error model (we only assume Markovianity), giving much more confidence to what we can expect from running benchmarking experiments. *A priori*, for schemes where a particular error model is assumed or where there is no theoretical guarantee whatsoever, we might run an experiment and not even see an exponential decay, rendering the scheme completely inapplicable for benchmarking the device in question.

The Clifford approximate twirl is a theoretical construction with excellent circuit size and depth scaling, involving a probabilistic circuit of size $O(n \log 1/\gamma)$ and depth $O(\log n \log 1/\gamma)$. We prove in Appendix D that this is actually optimal in n by showing any γ-approximate twirl requires a circuit of size $\Omega[(1-\gamma)n]$ and depth $\Omega[(1-\gamma)\log n]$. However, to achieve this scaling, we would need a complete graph topology. Fortunately, we can easily adapt it to other topologies that are more relevant to hardware at the cost of suboptimal scaling. For a detailed analysis, see Appendix A 2.

2. Qubit scaling with error

Assuming a fixed gate error, any benchmarking scheme has a limit on the number of qubits it can be applied to. One reason for this simply follows from the size of the quantum benchmarking circuit required. To obtain a meaningful result from the circuit, the total error of the circuit should be less than unity.

We consider the error of one cycle of a RB scheme. Now, we are working under a general error model where gate errors within a cycle can be arbitrarily correlated, so we need to consider worst-case additive error accumulation (AEA),[8] where errors of individual gates add. Then, we must require

$$s\epsilon \lesssim 1, \tag{9}$$

where s is the benchmarking circuit size and ϵ is the individual gate error (assumed to be the same for all gates). Since s is a

[7] Note that Eq. (8) holds for arbitrarily large m. Intuitively, the signal does not disappear at such large m because we assume the noise between each cycle is Markovian, although error correlations are allowed to exist within a cycle.

[8] If errors are uncorrelated, we expect a multiplicative accumulation of error, i.e., MEA.

function of n, this gives an upper bound on n. In particular, for Clifford RB or any scheme where at least one uniformly random n-qubit Clifford element needs to be generated, $s = \Omega(n^2)$, and so

$$n \lesssim \frac{1}{\sqrt{\epsilon}}. \tag{10}$$

This is somewhat unsatisfactory; for instance, we would have to lower gate errors by a factor of 100 to be able to benchmark 10 times more qubits.

We show in this section that Clifford XEB with the Clifford approximate twirl can do quadratically better: an inverse linear scaling $n \sim \epsilon^{-1}$. That is, by halving the gate error, we can benchmark twice as many qubits. This is actually optimal for twirling schemes which must implement γ-approximate twirls given the circuit lower bound we prove in Appendix D. In fact, as long as we have local gates (acts on a constant number of qubits), any benchmarking circuit in which all of the qubits are involved in a gate must have linear size. Thus, assuming a worst-case AEA model, the best we can hope for in such a scenario is inverse linear scaling.

We now show achievability. The Clifford approximate twirl is able to form a γ-approximate twirl for $\gamma \approx \frac{1}{2}$ with a linear size circuit [25], so the bound on n from that would be $n \lesssim \epsilon^{-1}$. However, we have an additional constraint. In Theorem 1, we require

$$\delta \leqslant \frac{1-\gamma}{11}. \tag{11}$$

This is a similar requirement to that of the main result in [17]. Intuitively, γ is the distance of our noiseless twirling map from the Haar twirling map, which is known to give an exponential decay [25]. Hence, the higher γ is, the lower the tolerance the exponential decay has to the error δ, which is effectively a perturbation on the noiseless twirling map. We analyze what Eq. (11) implies about how the number of qubits we can benchmark scales with gate error for our scheme.

Recall δ is the error of an entire γ-approximate twirl, not just individual gate errors. We assume a simplified[9] error model where single-qubit gates are noiseless and two-qubit gates have the same noise level. Under AEA, δ is of the form

$$\delta = s\epsilon, \tag{12}$$

where ϵ is the two-qubit gate error and s is the size of the circuit. Now, for the Clifford approximate twirl, $s = cnk$, where c is a constant that appears in the circuit construction, and $k = O[\log(1/\gamma)] \in \mathbb{N}$ is the number of times the circuit is repeated to achieve a γ-approximate twirl [25]. Explicitly, with k repetitions, we achieve a γ-approximate twirl with

$$\gamma = \frac{1}{2^{k-1}}\left(1+\frac{1}{4^n}\right)^k. \tag{13}$$

Hence, Eq. (11) becomes

$$cnk\epsilon \leqslant \frac{1-\frac{1}{2^{k-1}}\left(1+\frac{1}{4^n}\right)^k}{11}. \tag{14}$$

We want to look at the asymptotic case where n is large, so we make a simplifying approximation $\frac{1}{4^n} \approx 0$. Then, Eq. (14) becomes an upper bound on n:

$$n \leqslant \frac{1-2^{-(k-1)}}{11ck\epsilon}. \tag{15}$$

We see that for fixed ϵ, to maximize the upper bound, we set $k = 2$ or 3 (they give the same answer) to obtain

$$n \leqslant \frac{1}{44c\epsilon}. \tag{16}$$

Thus, the qubit number scales inversely with the infidelity.

To be extremely concrete, we can fill in some numbers. State-of-the-art ion trap and superconducting circuit devices can achieve a two-qubit gate infidelity of around 10^{-3} [36,37]. Tracking the constants in [25], $c = 3$ for a star topology circuit. We therefore have a theoretical guarantee of an exponential decay for up to $n = 7$ qubits[10] under a general error model. Note that there was little attempt in [35] to optimize the constant in Eq. (11) or in [25] to optimize the circuit size constants, so the largest n for which we have a theoretical guarantee may be substantially higher. We leave this for future work.

We compare our theoretical predictions with the findings in our numerical simulations. For the Clifford approximate twirl simulations, we only have a theoretical guarantee for the case with 25 qubits and two-qubit gate noise level 10^{-4} (yellow line in the leftmost plot in Sec. II B) if we fill in all the numbers and the constants. Nevertheless, most of the experiments give rise to single exponential decays that closely reflect the overall noise level of the circuits, which indicates that Clifford XEB is applicable to a wider variety of settings than might be suggested by theory. One reason for this disparity is that Theorem 1 applies to *general* error models within each cycle, which can be highly correlated and, in the worst case, even adversarial. In contrast, in our numerical experiments we have gate-independent depolarizing errors. Since our scheme is optimal in scaling for arbitrary errors, it is unlikely that a theoretical guarantee for a general error model can be extended to over 100 qubits given the two-qubit noise levels on current devices. To extend Theorem 1 to larger quantum devices, we either have to reduce gate errors or make assumptions on the error models. For example, it is proven in [38] that the number of qubits can scale with ϵ^{-2}, provided that the noise level is sufficiently weak and incoherent.[11] In general, we should seek a balance of theoretical guarantee and applicability, i.e., make a minimal number of realistic assumptions about errors to obtain a theoretical guarantee.

B. 1D chain and 2D grid

We would hope that the theoretical analysis for the Clifford approximate twirl would extend to the other topologies we consider. However, for the 1D chain and 2D grid schemes, the depth of each circuit cycle is constant. As shown in

[9]This is for the sake of numerical convenience and *not* a limiting assumption on what errors we allow.

[10]That is, $n = 7, k = 2, c = 3, \epsilon = 10^{-3}$ satisfies Eq. (14).

[11]Note, however, they considered a setting different from our quest for exponential decays.

Appendix D, logarithmic depth is necessary to get an approximate twirl with γ bounded away from 1 by a constant. For these schemes, we instead argue that the measures on the unitary group they induce are γ-approximate twirls with respect to the $|||\cdot|||_2$ norm. In this case we can use the following theorem proved in [35]:

Theorem 2 (Corollary 16 of [35]). Let $(S, \mu, \phi, M, \rho_0)$ be a linear XEB scheme on n qubits. Suppose that the twirling map corresponding to μ is a γ-approximate twirl with respect to the $|||\cdot|||_2$ norm and that the implementation map satisfies

$$\mathbb{E}_{g\sim\mu}\|\phi(g)-\omega(g)\|_\diamond \leqslant \delta. \tag{17}$$

Let $\delta' := 2^{n/2}\delta$. If $\delta' \leqslant \frac{1-\gamma}{11}$, then there exists $A, B \in \mathbb{R}$ and $p \in [1-2\delta', 1]$, such that

$$|q_R(m) - (A + Bp^m)| \leqslant 16 \times 2^{\frac{5}{2}n}(\gamma + 6\delta')^m. \tag{18}$$

Although the 2^n factors can be very large, this still proves that for sufficiently small errors, we can prove a single exponential decay under a general error model. In the case that the error channels are mixtures of unitaries, we have another version of Theorem 2 which relaxes the noise-level restriction to essentially that of Theorem 1. We have the following.

Theorem 3. Let $(S, \mu, \phi, M, \rho_0)$ be a linear XEB scheme on n qubits. Suppose that the twirling map corresponding to μ is a γ-approximate twirl with respect to the $|||\cdot|||_2$ norm, and that ϕ maps to probabilistic mixtures of unitaries:

$$\phi(g) = \int dv_g u$$

for probabilistic measures v_g on $\mathrm{SU}(d)$. Additionally, suppose that the implementation map satisfies

$$\mathbb{E}_{g\sim\mu}\mathbb{E}_{u\sim v_g}\|\phi(g)-u\|_\diamond \leqslant \delta. \tag{19}$$

If $\delta \leqslant \frac{1-\gamma}{11}$, then there exists $A, B \in \mathbb{R}$ and $p \in [1-2\delta, 1]$, such that

$$|q_R(m) - (A + Bp^m)| \leqslant 16 \times 2^{\frac{5}{2}n}(\gamma + 6\delta)^m. \tag{20}$$

Compared to Theorem 2, Theorem 3 lifts the exponential prefactor $2^{n/2}$ on the noise-level tolerance, with a slightly tighter restriction on the definition of the noise level itself. However, the noise level δ can be well approximated in the presence of error model that can be readily decomposed as mixtures of unitaries. We present the proof of Theorem 3 in Appendix E 1.

In general, it is much easier to prove a spectral gap than a gap with respect to the induced diamond norm. We have the following.

Definition 4. Let μ be a probabilistic measure on a finite group G, and $A := \mathrm{Supp}(\mu) \subseteq G$ be its support. We say that μ is *weakly mixing* if either of following holds:

(i) $\langle A^{-1}A\rangle = G$ or $\langle AA^{-1}\rangle = G$;

(ii) $\forall\, a \in G$, $\langle aA\rangle = G$, or $\forall\, a \in G$, $\langle Aa\rangle = G$.

Theorem 4. Assume that μ is weakly mixing over the n-qubit Clifford group. Then,

$$|||\Lambda(\mu) - \Lambda(\mu_H)|||_2 < 1, \tag{21}$$

where μ_H is Haar measure on $\mathrm{SU}(2^n)$.

The proof of Theorem 4 can be found in Appendix E 2. That the 1D chain and 2D grid Clifford XEB schemes satisfy Theorem 4 is straightforward and left as an exercise for the reader.[12]

IV. DISCUSSION

In this work we propose Clifford XEB, a benchmarking scheme for large-scale quantum devices using Clifford circuits. Clifford XEB is both efficient in quantum circuit size and necessary classical processing, and it has great promise for benchmarking quantum devices much larger than heretofore possible. We numerically show that this scheme is feasible for more than 1000 qubits and theoretically prove that the scheme yields a single exponential decay for Clifford approximate twirls.

It is worthwhile to first take a step back and ask why we would need to holistically benchmark a quantum processor with tens or hundreds of qubits in the first place. One may think that it is sufficient to only perform single- or two-qubit benchmarking to characterize the native gate set. Even some effects of crosstalk can be captured via simultaneous single- or two-qubit RB [12]. The authors of [13] study this topic in depth by performing Clifford RB on a three-qubit processor. Interestingly, they find that the three-qubit RB result varies depending on the calibration procedure in a way that is not captured by simultaneous single- or two-qubit RB. This suggests there is additional information to be gained from running multiqubit benchmarks. This is especially relevant as the number of qubits on processors continues to increase and the calibration process becomes increasingly convoluted. Furthermore, there are a variety of obstacles to overcome to be able to scale quantum processors to achieve fault tolerance, including the size of dilution fridges, and any workaround could introduce errors that can only be detected by a multiqubit benchmark. Lastly, using a multiqubit benchmark helps verify that the quantum processor does not have large correlated errors [1] which are detrimental to quantum error correction.

There have been other proposals for holistically benchmarking quantum computers. The quantum volume experiment [39] aims to compare different quantum computers in terms of their computational capability rather than gate quality, but it is not scalable with respect to the effective number of qubits. Many proposals for estimating the global Pauli errors [40–43] provide more detailed characterizations of the error, and are probably more suitable after a holistic, single-number metric is extracted first. There have been variants of RB for holistically benchmarking quantum computers in a scalable way, such as match-gate benchmarking [44], direct randomized benchmarking [45], and mirror randomized benchmarking [46]. Part of the reason why linear XEB has been used in experiments such as [1,2] is due to its convenience of hardware implementation. It just requires a certain random gate sequence followed by a measurement. Our scheme boasts the exact same convenience but is also manifestly scalable. Another distinguishing feature is that for certain Clifford circuits such as the ones we considered, we can rigorously prove

[12]Technically, the proof for the 1D chain requires an additional layer of random single-qubit gates at the end. This poses a very minor modification to the random circuit generation.

the scheme yields an exponential decay under a general error model. Another possible alternative scheme is to use other gate sets that are classically simulable. If we can construct γ-approximate twirls using other gate sets, e.g., match gates [47], then we can formulate XEB with such gate sets that might be more relevant in other experimental settings.

Now, there is also work showing that we can realize exact unitary 2-designs using almost linear-sized circuits (up to logarithm factors) [48]. Specifically, they require $O(n \log n \log\log n)$ size circuits. In contrast, the Clifford circuits we consider in our work are exactly linear size [25]. In fact, these logarithm factors can still be quite large in practice (for example, if $n = 50$, $\log_2 n \log_2 \log_2 n \approx 14$, a nontrivial overhead on the required noise levels). Finally, in [21], there is a mention of conducting linear XEB with Clifford circuits. However, they do not pursue this line of thought nor further elaborate on its theory. Here we study Clifford XEB in depth from both a numerical and theoretical standpoint.

Similar to mirror RB [46], instead of classical simulation, we could alternatively sequentially reverse the Clifford layers on the quantum circuit and observe the exponential decay of the population of the initial state with respect to the number of cycles. The proof of exponential decay in [35] can be readily applied to this scheme as well. Compared to the Clifford XEB, the mirror RB does not rely on the quantum information be sufficiently scrambled in order to extract fidelity information. Consequently, the mirror RB would probably need a smaller number of cycles. However, each cycle now consists of twice the number of gates compared to the Clifford XEB, hence, the error tolerance of the protocol is halved compared to that of the Clifford XEB. Furthermore, as it shares the same noiseless twirling map as the Clifford XEB, it could also potentially suffer from a small spectral gap of the twirling map, making large gate error rates unable to be extracted. We leave a more detailed comparison for future work.

We give some directions for future research. Numerical experiments in Sec. II show that Clifford XEB can be used to benchmark large quantum processors up to 1000 qubits. However, there are currently only a few quantum processors on the order of 100 qubits. On the other hand, it would still be interesting to test Clifford XEB on real quantum devices with tens of qubits. Given that the noise in experiments is not adversarial and mostly Markovian, we expect that Clifford XEB can probe the overall performance of a quantum processor in a relatively short time.

One experimentally relevant question is to build upon Clifford XEB and propose a comprehensive benchmarking workflow. Since it benchmarks circuits of linear size, the Clifford XEB can serve as an overall figure of merit of the entire processor, and serve as a guideline for subsequent, more detailed, benchmarking schemes. Possible candidates of such subsequent experiments include cycle benchmarking [40] or individual RB experiments. Moreover, one can apply Clifford XEB to smaller regions on a quantum chip to extract more local and detailed information regarding heterogeneous performance over the quantum chip.

It would also be interesting to investigate, both from a theoretical and an experimental perspective, the NSR for a Clifford random circuit. Estimating the NSR is an essential step in experiment, as the NSR rate serves as an indicator whether the decay rate extracted from linear XEB faithfully represents the gate error rather than the mixing of the ideal circuit. For the numerical simulation on the 35×35 grid, it took half a day to gather the noisy linear XEB results, but two days and a half to get the NSR. This is because the fluctuation of the noiseless linear XEB value is purely from the randomness of the circuits chosen and is numerically found to be much larger than that of the noisy experiments. A faster algorithm that can extract the decay exponent without having to sample millions of random circuits would be the key to more efficient Clifford XEB benchmarking. Towards this end, there have been studies on the spectral gap from random Clifford circuits from a Markov chain perspective [27,49–52], and we leave it to future work to incorporate such methods to a better estimation of the spectral gap on particular random Clifford circuit ensembles. Of course, it would also be useful to improve the constants in our number of qubits we can benchmark given by Eq. (14). In particular, it would be interesting to realize a γ-approximate twirl in a way that is relatively more natural for topologies that are more relevant for hardware, such as the 2D grid. We could also consider topologies tailored for realizing quantum error correction, such as the honeycomb lattice [53]. Furthermore, it would be interesting if we can make minimal assumptions on the gate errors to obtain a theoretical guarantee for a much higher number of qubits. For instance, most gates in superconducting circuits are limited by decoherence, which implies incoherent errors are dominant [36].

Note added. Recently, another paper was posted on the arXiv [54] which looks into a generalization of linear XEB called filtered RB, giving general guarantees of exponential decays as well as sufficient circuit depth and sample complexity bounds.

ACKNOWLEDGMENT

We would like to thank Y. Shi and M. Heinrich for helpful discussions and comments.

APPENDIX A: CLIFFORD APPROXIMATE TWIRL CONSTRUCTION

For completeness, we describe explicitly the sequence of gates to construct the Clifford approximate twirl given in [[25], Fig. 1].

1. Description

We first define the following procedures. Suppose we have n qubits in a star topology, where we denote the center qubit as qubit 1.

(i) *Perform a $\mathcal{P}_n$ twirl*: Apply a random Pauli gate on each qubit.

(ii) *Perform a $\mathcal{C}_1/\mathcal{P}_1$ twirl*: Apply a gate uniformly chosen from the set $\{I, S, H, SH, (SH)^2, SHS\}$, with S being the single-qubit phase gate and H the Hadamard gate. Alternatively, this could be implemented by applying $(SH)^j$ with j uniformly chosen from $\{0, 1, 2\}$, according to [25].

(iii) *Conjugate qubit 1 by a random XOR*: For each qubit $2, \ldots, n$, apply a CNOT gate between this qubit and qubit 1 with probability $3/4$.

We now describe how to implement the Clifford approximate twirl:

(1) Perform a $\mathcal{P}_n$ twirl.
(2) Perform $\mathcal{C}_1/\mathcal{P}_1$ twirl on all of the qubits.
(3) Conjugate qubit 1 by a random XOR.
(4) Apply H to qubit 1 and $\mathcal{C}_1/\mathcal{P}_1$ twirl the other qubits.
(5) Conjugate qubit 1 by a random XOR.
(6) Apply H to qubit 1 and $\mathcal{C}_1/\mathcal{P}_1$ twirl the other qubits.
(7) Apply S to the first qubit with probability $1/2$.
(8) Conjugate qubit 1 by a random XOR.
(9) $\mathcal{C}_1/\mathcal{P}_1$ twirl the first qubit.
(10) To obtain a γ-approximate twirl, repeat steps 2 to 9 for $O[\log(1/\gamma)]$ times.

2. Different topologies

The only two-qubit gate operation involved in this construction is conjugating qubit 1 by a random XOR. This requires a star topology circuit. Furthermore, a direct implementation of this procedure requires a circuit depth of $\Omega(n)$. It is possible to implement it in depth $O(\log n)$ on a fully connected graph, as was shown in [[25], Fig. 3], but realistic quantum processors have limited connectivity. To remedy this, we describe how to implement the random XOR circuit on a general connectivity graph. As a side result we also show that we can achieve depth $O(\log n)$ on a binary tree connectivity graph.

It is easy to see that in the computational basis, the conjugation by XOR operation effectively applies an X gate to qubit 1 conditioned on the XOR of the random set. We therefore give a construction on a general graph by finding a spanning tree and iteratively computing the XOR of the random set through each level of the tree. Given the graph is connected, we can choose an arbitrary qubit to be the root (qubit 1), and form a spanning tree rooted from qubit 1. Each qubit v apart from qubit 1 is then assigned a *depth* $d(v)$ as its distance from qubit 1 on the spanning tree, and a *degree* $\delta(v)$. Moreover, denote $D = \max_v d(v)$ and $\Delta = \max_v \delta(v)$ be the depth and degree of the spanning tree. Now, for each of the qubits other than qubit 1, add it to a set S with probability $3/4$. We call a qubit v *active*, if at least one of the qubits in the subtree rooted at v (including v) is in S. We then perform the following operations.

(1) Do the following sequentially for each l from $D-1$ to 1:
  (a) Do the following in parallel for each active qubit x at depth l:
    (i) If x is not in S, it must have an active child y. Choose an arbitrary such y and apply a CNOT gate from x to y.
    (ii) Apply CNOT gates from each active child of x to x sequentially.
(2) Apply CNOT gates from all active qubits at depth 1 to the root sequentially.
(3) Apply the CNOT gates generated in step 1 again but in reverse order.

To see why this circuit works, we first assume that every qubit is in the computational basis before the circuit, and the general case follows from linearity. For step 1, one can see by induction that all qubits of depth l' will have a value that equals to the XOR of qubits in its subtree that are contained in S when the loop with $l = l'$ finishes. Step 1(a)i ensures that if x is not in S, its value will be canceled after step 1(a)ii. Step 2 ensures that qubit 1 gets the corresponding XOR values from its children, and step 3 "uncomputes" the intermediate results in other qubits.

The total depth of the circuit can be upper bounded by $2 - D\Delta$. To see this, observe that step 3 takes the same depth as step 1, which takes $(D-1)\Delta$ layers for sequentially executing $D-1$ repetitions of step 1(a), each takes at most Δ layers. Therefore, the depth of the circuit greatly depends on the spanning tree, which in turn depends on the underlying connectivity graph. When the spanning tree can be chosen to be a perfect binary tree, the random XOR can be done in $\Theta[\log(n)]$ depth. For hardware relevant graphs such as the 1D chain and the 2D grid, the circuit depth is upper bounded by $O(D)$ as the degree of the spanning tree is upper bounded by the degree of the connectivity graph. In this case, the depth of the spanning tree can be minimized to the radius of the connectivity graph $r(G) = \min_u \max_v d(u, v)$ by rooting at $\arg\min_u \max_v d(u, v)$ and taking the spanning tree to consist of the shortest paths from each node to the root. The radius of the 1D chain and the 2D grid is $\Theta(n)$ and $\Theta(\sqrt{n})$, respectively, giving a circuit depth for the XOR circuit $\Theta(n)$ and $\Theta(\sqrt{n})$, respectively. One can see that this is asymptotically tight as it takes $\Omega[r(G)]$ steps to propagate information from an arbitrary qubit to a fixed one.

APPENDIX B: EXPECTATION AND VARIANCE OF LINEAR XEB

It is observed in Sec. II that the Clifford XEB value converges to 1 and 0 for the noiseless cases and the noisy cases, respectively. Similar behavior has also been observed in linear XEB experiments [1,2]. In this Appendix we analyze the expectation and variance for noiseless and noisy Clifford XEB values when the number of cycles is sufficiently large. We assume that for Clifford XEB and linear XEB, the first four moments of the distribution of the random gates converge to those of the uniform distribution over the Clifford group and the Haar measure over the unitary group, respectively.

1. Noiseless case

For any circuit $C \in \mathrm{SU}(2^n)$, we define

$$\beta_C = 2^n \sum_x |\langle x|C|0^n\rangle|^4, \tag{B1}$$

which is the noiseless linear XEB value plus 1. When C is Haar random in $\mathrm{SU}(2^n)$, $C|0^n\rangle$ is a uniformly random unit vector in $\mathbb{C}^{2^n}$. This vector can be parametrized as

$$(e^{i\phi_1}\cos\theta_1, e^{\phi_2}\sin\theta_1\cos\theta_2, \ldots, e^{\phi_D}\sin\theta_1\sin\theta_2\ldots\sin\theta_{D-1}) \tag{B2}$$

in the computational basis, where $D = 2^n$ is the dimension. To compute the variance we claim we only need to consider the dependence on θ_1 and θ_2. The metric of the integral will contain two factors: the first factor is related to the spherical coordinates θ and is proportional to $\sin^{D-2}\theta_1 \sin^{D-3}\theta_2$, and

the second factor is related to the phase factors and is equal to

$$\cos\theta_1 \times \sin\theta_1 \cos\theta_2 \times \sin\theta_1 \sin\theta_2 \cos\theta_3 \times \cdots \propto \cos\theta_1 \cos\theta_2 \sin^{D-1}\theta_1 \sin^{D-2}\theta_2. \tag{B3}$$

Let $t_1 = \cos^2\theta_1,\ t_2 = \sin^2\theta_1 \cos^2\theta_2$, and the distribution of t_1 and t_2 is given by

$$\begin{aligned}\Pr[t_1, t_2] \propto \int d\theta_1 d\theta_2 \delta(t_1 - \cos^2\theta_1)\delta(t_2 - \sin^2\theta_1 \cos^2\theta_2) \\ \times \sin^{D-2}\theta_1 \sin^{D-3}\theta_2 \cos\theta_1 \cos\theta_2 \\ \times \sin^{D-1}\theta_1 \sin^{D-2}\theta_2 \propto (1 - t_1 - t_2)^{D-3}\end{aligned} \tag{B4}$$

for $0 \leqslant t_1,\ t_2 \leqslant 1,\ t_1 + t_2 \leqslant 1$. One can normalize the distribution and have

$$\Pr[t_1, t_2] = (D-1)(D-2)(1 - t_1 - t_2)^{D-3}. \tag{B5}$$

If we integrate over t_2, the distribution of t_1 is

$$\Pr[t_1] = (D-1)(1 - t_1)^{D-2}. \tag{B6}$$

By symmetry, for any $x \neq y$ the value of $|\langle x|C|0^n\rangle|^2$ and $|\langle y|C|0^n\rangle|^2$ will satisfy this distribution by identifying them with t_1 and t_2, respectively. Then one can calculate the mean and variance of β_C as follows:

$$\begin{aligned}\mathbb{E}_C \beta_C &= D \sum_x |\langle x|C|0^n\rangle|^4 \\ &= D^2 \int_0^1 dt_1 \Pr[t_1] t_1^2 \\ &= \frac{2D}{D+1},\end{aligned} \tag{B7}$$

$$\begin{aligned}\mathbb{E}_C \beta_C^2 &= D^2 \mathbb{E}_C \left(\sum_x |\langle x|C|0^n\rangle|^4 \right)^2 \\ &= D^2 \mathbb{E}_C \sum_{x \neq y} |\langle x|C|0^n\rangle|^4 |\langle y|C|0^n\rangle|^4 \\ &\quad + D^2 \mathbb{E}_C \sum_x |\langle x|C|0^n\rangle|^8 \\ &= D^3(D-1) \int_0^1 dt_1 \int_0^{1-t_1} dt_2 \Pr[t_1, t_2] t_1^2 t_2^2 \\ &\quad + D^3 \int_0^1 dt_1 \Pr[t_1] t_1^4 \\ &= \frac{4D^2(D+5)}{(D+1)(D+2)(D+3)}\end{aligned} \tag{B8}$$

and so the variance is given by

$$\begin{aligned}\mathbb{E}_C \beta_C^2 - (\mathbb{E}_C \beta_C)^2 &= \frac{4D^2(D-1)}{(D+1)^2(D+2)(D+3)} \\ &= \frac{4}{D} + O(D^{-2}).\end{aligned} \tag{B9}$$

From (B7) one can see that the noiseless linear XEB value converges to $\frac{D-1}{D+1}$, which is close to 1 for large D. This also applies to Clifford XEB since the uniform distribution on the Clifford group forms a unitary 2-design.

When C is a uniformly random Clifford gate, $C|0^n\rangle$ will be a uniformly random stabilizer state. We introduce the concept of k neighbor of a stabilizer state $|\psi\rangle$, which is the set of stabilizer states $|\phi\rangle$ such that $|\langle\phi|\psi\rangle| = 2^{-k/2}$. It is known that every n-qubit stabilizer state has the same number of k neighbors [55], which is denoted by $\mathcal{L}_n(k)$. We know that for any stabilizer state $C|0^n\rangle$, the inner product $|\langle x|C|0^n\rangle|$ is $2^{-k/2}$ for 2^k choices of x, where k is an integer, and for the remaining values of x the inner product is 0. It is easy to see that $\beta_C = 2^{n-k}$ in this case.

For each $|x\rangle$, there are $\mathcal{L}_n(k)$ k neighbors, and each such state is in turn a k neighbor of 2^k computational states. This means that there are $2^{n-k}\mathcal{L}_n(k)$ stabilizer states that lead to $\beta_C = 2^{n-k}$. Let $\mathcal{N}(n)$ be the total number of n-qubit stabilizer states. Then

$$\mathbb{E}_C \beta_C = \sum_k 2^{n-k} \frac{2^{n-k}\mathcal{L}_n(k)}{\mathcal{N}(n)} = \sum_t 4^t \frac{\mathcal{L}_n(n-t)}{\mathcal{N}(n)}, \tag{B10}$$

$$\mathbb{E}_c \beta_C^2 = \sum_k 4^{n-k} \frac{2^{n-k}\mathcal{L}_n(k)}{\mathcal{N}(n)} = \sum_t 8^t \frac{\mathcal{L}_n(n-t)}{\mathcal{N}(n)}. \tag{B11}$$

It is known [[55], Theorem 16] that in the large-n limit,

$$c_1 2^{-t(t+5)/2} \leqslant \frac{\mathcal{L}_n(n-t)}{\mathcal{N}(n)} \leqslant c_2 2^{-t(t+3)/2} \tag{B12}$$

for some constants c_1 and c_2. So $\mathbb{E}_c \beta_C$ and $\mathbb{E}_c \beta_C^2$ are both $\Theta(1)$. This means that the variance of β_C is $\Theta(1)$ when C is a uniformly random Clifford circuit.

We have discussed above the variance of XEB with respect to the sampled circuit but, in practice, the sampling of bit strings can introduce additional uncertainty. We consider the general experiment where N_1 circuits are sampled, and for each circuit C_i, N_2 bit strings are sampled. We denote these bit strings by $x_{j|i}$ where $1 \leqslant i \leqslant N_1,\ 1 \leqslant j \leqslant N_2$. We define

$$q(x_{1|i}, x_{2|i}, \ldots, x_{N_2|i}|C_i) = \frac{1}{N_2} \sum_{j=1}^{N_2} |\langle x_{j|i}|C_i|0^n\rangle|^2. \tag{B13}$$

Then the estimated XEB from this experiment would be

$$\hat{p} = \frac{D}{N_1} \sum_{i=1}^{N_1} q(x_{1|i}, x_{2|i}, \ldots, x_{N_2|i}|C_i) - 1. \tag{B14}$$

$\hat{p}$ is dependent on the circuits $C_1, \ldots, C_{N_1}$, and according to the law of total variance,

$$\begin{aligned}\mathrm{var}(\hat{p}) = \mathbb{E}_{C_1,\ldots,C_{N_1}}[\mathrm{var}(\hat{p}|C_1, \ldots, C_{N_1})] \\ + \mathrm{var}_{C_1,\ldots,C_{N_1}}[\mathbb{E}(\hat{p}|C_1, \ldots, C_{N_1})].\end{aligned} \tag{B15}$$

Note that the samplings of the circuits are independent, we have

$$\mathbb{E}(\hat{p}|C_1, \ldots, C_{N_1}) = \frac{D}{N_1} \sum_{i=1}^{N_1} \mathbb{E}[q(x_{1|i}, x_{2|i}, \ldots, x_{N_2|i}|C_i)] - 1, \tag{B16}$$

$$\mathrm{var}(\hat{p}|C_1, \ldots, C_{N_1}) = \frac{D^2}{N_1^2} \sum_{i=1}^{N_1} \mathrm{var}[q(x_{1|i}, x_{2|i}, \ldots, x_{N_2|i}|C_i)], \tag{B17}$$

and therefore

$$\begin{aligned}
&\mathrm{var}_{C_1,\ldots,C_{N_1}}[\mathbb{E}(\hat{p}|C_1,\ldots,C_{N_1})] \\
&= \frac{D^2}{N_1^2}\sum_{i=1}^{N_1}\mathrm{var}_{C_i}\{\mathbb{E}[q(x_{1|i},x_{2|i},\ldots,x_{N_2|i}|C_i)]\} \\
&= \frac{1}{N_1^2}\sum_{i=1}^{N_1}\mathrm{var}_{C_i}\,\beta_{C_i} \\
&= \frac{1}{N_1}[\mathbb{E}_C\beta_C^2-(\mathbb{E}_C\beta_C)^2] \qquad \text{(B18)} \\
&\quad\times\mathbb{E}_{C_1,\ldots,C_{N_1}}[\mathrm{var}(\hat{p}|C_1,\ldots,C_{N_1})] \\
&= \frac{D^2}{N_1^2}\sum_{i=1}^{N_1}\mathbb{E}_{C_i}\{\mathrm{var}[q(x_{1|i},x_{2|i},\ldots,x_{N_2|i}|C_i)]\} \\
&= \frac{D^2}{N_1^2}\sum_{i=1}^{N_1}\mathbb{E}_{C_i}\left\{\frac{1}{N_2^2}\sum_{j=1}^{N_2}\mathrm{var}_{x_{j|i}}[\langle x_{j|i}||C_i|0^n\rangle]\right\} \\
&= \frac{D^2}{N_1N_2}\mathbb{E}_C\left[\sum_x|\langle x|C|0^n\rangle|^6-\left(\sum_x|\langle x|C|0^n\rangle|^4\right)^2\right] \\
&= \frac{1}{N_1N_2}\mathbb{E}_C\left[D^2\sum_x|\langle x|C|0^n\rangle|^6-\beta_C^2\right]. \qquad \text{(B19)}
\end{aligned}$$

So we have

$$\begin{aligned}
\mathrm{var}(\hat{p}) = \frac{1}{N_1}\Bigg[&\frac{1}{N_2}\left(D^2\mathbb{E}_C\sum_x|\langle x|C|0^n\rangle|^6-\mathbb{E}_C\beta_C^2\right) \\
&+\mathbb{E}_C\beta_C^2-(\mathbb{E}_C\beta_C)^2\Bigg]. \qquad \text{(B20)}
\end{aligned}$$

The term $|\langle x|C|0^n\rangle|^6$ involves the third moment of C and should be the same for the uniform distribution over the Clifford group and the Haar measure on SU(D). Using the notation in Eq. (B6),

$$\mathbb{E}_C\sum_x|\langle x|C|0^n\rangle|^6 = D\int_0^1 t_1^3\,\Pr[t_1]dt_1 = \frac{6}{(D+1)(D+2)}. \qquad \text{(B21)}$$

Also note that for any distribution on the Clifford group, the term involving N_2 should vanish, as the measurement result of a stabilizer state on the computational basis is a uniform distribution over a set of bit strings, and the corresponding variance is 0.

2. Noisy case

In this section, we investigate the single exponential decay behavior in the noisy case. In particular, we give estimates for the coefficients A and B in Theorem 1, showing that $A=0$ and $B=O(1)$. Following the universal randomized benchmarking (URB) framework [35], we can write the Clifford XEB experiment results as

$$q_R(m) = -1+2^n\mathrm{tr}\big[|0^n\rangle\langle 0^n|\big(\Lambda_R^m(\mathcal{D}_M)(\rho_0)\big)\big],$$

where the *noisy twirling map* Λ_R is defined as a map on quantum channels:

$$\Lambda_R:\mathcal{C}\mapsto\mathbb{E}_{g\sim u}[g^\dagger\circ\mathcal{C}\circ\phi(g)].$$

Now, using the notations in the proof of [[35], Theorem 8], Λ_R is regarded as a perturbed version of Λ_R^* and has a gapped spectrum leading to the single exponential decay in an appropriate parameter range. Specifically, define

$$X_1 = \frac{\langle\cdot,\mathcal{D}\rangle_{\mathrm{SO}}}{\langle\mathcal{D},\mathcal{D}\rangle_{\mathrm{SO}}}\mathcal{D}+\frac{\langle\cdot,\mathcal{I}-\mathcal{D}\rangle_{\mathrm{SO}}}{\langle\mathcal{I}-\mathcal{D},\mathcal{I}-\mathcal{D}\rangle_{\mathrm{SO}}}(\mathcal{I}-\mathcal{D}),\ X_2=\mathcal{I}-X_1.$$

One can verify that $X_i^2=X_i$ for $i=1,2$ and

$$X_i\Lambda_R^*X_j=0$$

when $i\neq j$. The matrix perturbation result states that when $|||\Lambda_R^*-\Lambda_{\mathrm{Haar}}|||_\diamond<\gamma$, $|||\Lambda_R^*-\Lambda_R|||_\diamond\leqslant\delta$, and $\delta<\frac{1-\gamma}{11}$, there exists twirling maps L,R such that

(i) $L=R^{-1}$, $|||L|||_\diamond\leqslant 4$, $|||R|||_\diamond\leqslant 4$;
(ii) $X_iR\Lambda_RLX_j=0$ when $i\neq j$;
(iii) $|||X_2R\Lambda_RLX_2|||_\diamond\leqslant\gamma+6\delta$;
(iv) $|||X_1R\Lambda_RLX_1-X_1|||_\diamond\leqslant 2\delta$.

Let $A_i'=X_iR\Lambda_RLX_i$. It can be further proven that A_1' has one eigenvalue 1 and another real eigenvalue lying in $[1-2\delta,1]$. Then

$$\Lambda_R^m = L(A_1')^mR+L(A_2')^mR$$

and

$$\begin{aligned}
q_R(m) &= -1+2^n\mathrm{tr}\big[|0^n\rangle\langle 0^n|\big(\Lambda_R^m(\mathcal{D}_M)(\rho_0)\big)\big] \qquad \text{(B22)} \\
&= -1+2^n[\mathrm{tr}[|0^n\rangle\langle 0^n|(L(A_1')^mR(\mathcal{D}_M)(\rho_0))] \\
&\quad+\mathrm{tr}[|0^n\rangle\langle 0^n|(L(A_2')^mR(\mathcal{D}_M)(\rho_0))]]. \qquad \text{(B23)}
\end{aligned}$$

Since

$$|||A_2'|||_\diamond\leqslant\gamma+6\delta,$$

the second term vanishes quickly, and the single exponential decay is given by the first term.

We assumed $p<1$, so A is well defined and is the contribution corresponding to the eigenchannel of Λ_R with eigenvalue 1. It is easy to see that this eigenchannel is the depolarization channel $\mathcal{D}(\rho)=\frac{I}{D}$:

$$\begin{aligned}
\Lambda_R(\mathcal{D})(\rho) &= \int d\mu\,\omega^\dagger(g)\circ\mathcal{D}[\phi(g)(\rho)] \\
&= \int d\mu\,\omega^\dagger(g)\left(\frac{I}{D}\right)=\frac{I}{D}
\end{aligned}$$

since the inverting maps $\omega^\dagger(g)$ are all unital. We then have

$$A = 2^n\mathrm{tr}[|0\rangle\langle 0|^{\otimes n}\mathcal{D}(\rho)]-1=0.$$

We thereby obtain the first conclusion $A=0$. Note that this holds in general when Λ_R only has only one eigenvalue with modulus 1: this ensures that $\Lambda_R^m(\mathcal{D}_M)\to\mathcal{D}$ when $m\to\infty$. This criterion does not pose any requirement on the state preparation or the measurement.

Both Λ_R and Λ_R^* are twirling maps; they map channels to channels. Consequently, they map differences of channels to differences of channels. It is thus more convenient to restrict

them on the space spanned by the differences of channels. In this case one can define X_1' to be

$$X_1' = \frac{\langle \cdot, \mathcal{I} - \mathcal{D}\rangle_{\mathrm{SO}}}{\langle \mathcal{I} - \mathcal{D}, \mathcal{I} - \mathcal{D}\rangle_{\mathrm{SO}}}(\mathcal{I} - \mathcal{D}).$$

The perturbation result tells us that

$$\begin{aligned}|B| &= |2^n \mathrm{tr}[|0\rangle\langle 0|^{\otimes n} L X_1' R(\mathcal{D}_M)(\rho)]| \\ &\leqslant 2^n \|L X_1'(R(\mathcal{D}_M))\|_\diamond \\ &\leqslant 2^n \left\| \frac{\langle R(\mathcal{D}_M), \mathcal{I} - \mathcal{D}\rangle_{\mathrm{SO}} L(\mathcal{I} - \mathcal{D})}{\langle \mathcal{I} - \mathcal{D}, \mathcal{I} - \mathcal{D}\rangle_{\mathrm{SO}}} \right\|_\diamond \\ &= \frac{2^n}{\langle \mathcal{I} - \mathcal{D}, \mathcal{I} - \mathcal{D}\rangle_{\mathrm{SO}}} \|L(\mathcal{I} - \mathcal{D})\|_\diamond |\langle \mathcal{D}_M, R^\dagger(\mathcal{I} - \mathcal{D})\rangle_{\mathrm{SO}}|.\end{aligned}$$

For sake of simplicity we denote $\mathcal{N} := R^\dagger(\mathcal{I} - \mathcal{D})$. We now bound the quantity $|\langle \mathcal{D}_M, \mathcal{N}\rangle_{\mathrm{SO}}|$.

Let $\mathcal{P} : \rho \to \sum_i \langle i|\rho|i\rangle |i\rangle\langle i|$ be the dephasing channel. One can check that it is self-adjoint with respect to the superoperator inner product. Therefore,

$$\langle \mathcal{D}_M, \mathcal{N}\rangle_{\mathrm{SO}} = \langle \mathcal{P} \circ \mathcal{D}_M, \mathcal{N}\rangle_{\mathrm{SO}} = \langle \mathcal{D}_M, \mathcal{P} \circ \mathcal{N}\rangle_{\mathrm{SO}}.$$

As $\|\mathcal{P} \circ \mathcal{N}\|_\diamond \leqslant \|\mathcal{N}\|_\diamond$ due to data processing inequality, we can assume without generality that $\mathcal{N}$ is a quantum-classical superoperator: that is, there exists a set of operators $\{N_i\}_i$ such that

$$\mathcal{N} : \rho \to \sum_i \mathrm{tr}[N_i \rho]|i\rangle\langle i|.$$

Then

$$\begin{aligned}\langle \mathcal{D}_M, \mathcal{N}\rangle_{\mathrm{SO}} &= \sum_{P_i \in P^{\otimes n}} \mathrm{tr}[\mathcal{D}(P_i)\mathcal{N}(P_i)] && \text{(B24)} \\ &= \sum_{P_i \in P^{\otimes n}} \sum_{j \in \{0,1\}^n} \mathrm{tr}[M_j P_i][N_j P_i] && \text{(B25)} \\ &= 2^n \sum_j \mathrm{tr}[M_j N_j] && \text{(B26)} \\ &= 4^n \sum_j \frac{\mathrm{tr}[M_j]}{2^n} \mathrm{tr}[\bar{M}_j N_j], && \text{(B27)}\end{aligned}$$

where $\bar{M}_i = \frac{M_i}{\mathrm{tr}[M_i]}$. Since $\{M_i\}_i$ is a POVM, we have $M_i \succcurlyeq 0$ and $\sum_i M_i = I$, and therefore $\bar{M}_i$ is a quantum state. We argue that $|\langle \mathcal{D}_M, \mathcal{N}\rangle_{\mathrm{SO}}| \leqslant 4^n \|\mathcal{N}\|_\diamond$. Assume otherwise, then there exists i such that

$$|\mathrm{tr}[\bar{M}_i N_i]| > \|\mathcal{N}\|_\diamond.$$

However, this is not possible as

$$\|\mathcal{N}\|_\diamond \geqslant |\mathcal{N}(\bar{M}_i)|_{\mathrm{tr}} = \sum_j |\mathrm{tr}[\bar{M}_i N_j]| \geqslant |\mathrm{tr}[\bar{M}_i N_i]|.$$

We then calculate $\langle \mathcal{I} - \mathcal{D}, \mathcal{I} - \mathcal{D}\rangle_{\mathrm{SO}}$. By linearity we have

$$\begin{aligned}&\langle \mathcal{I} - \mathcal{D}, \mathcal{I} - \mathcal{D}\rangle_{\mathrm{SO}} && \text{(B28)} \\ &= \langle \mathcal{I}, \mathcal{I}\rangle_{\mathrm{SO}} - 2\langle \mathcal{D}, \mathcal{I}\rangle_{\mathrm{SO}} + \langle \mathcal{D}, \mathcal{D}\rangle_{\mathrm{SO}}. && \text{(B29)}\end{aligned}$$

For any quantum channel $\mathcal{C}$,

$$\begin{aligned}\langle \mathcal{D}, \mathcal{C}\rangle_{\mathrm{SO}} &= \sum_{P_i \in P^{\otimes n}} \mathrm{tr}[\mathcal{D}(P_i)\mathcal{C}(P_i)] && \text{(B30)} \\ &= \sum_{P_i \in P^{\otimes n}} \mathrm{tr}[2^{-n} I \,\mathrm{tr}[P_i]\mathcal{C}(P_i)] && \text{(B31)} \\ &= \mathrm{tr}[2^{-n} 2^n \mathcal{C}(I)] = 2^n. && \text{(B32)}\end{aligned}$$

On the other hand,

$$\begin{aligned}\langle \mathcal{I}, \mathcal{I}\rangle_{\mathrm{SO}} &= \sum_{P_i \in P^{\otimes n}} \mathrm{tr}[\mathcal{I}(P_i)\mathcal{I}(P_i)] && \text{(B33)} \\ &= \sum_{P_i \in P^{\otimes n}} \mathrm{tr}[I] = 4^n 2^n = 8^n. && \text{(B34)}\end{aligned}$$

Therefore,

$$\langle \mathcal{I} - \mathcal{D}, \mathcal{I} - \mathcal{D}\rangle_{\mathrm{SO}} = 8^n - 2^n.$$

Plugging everything in we have

$$B \leqslant \|\mathcal{N}\|_\diamond \|L(\mathcal{I} - \mathcal{D})\|_\diamond \frac{4^n}{4^n - 1}.$$

Since $\mathcal{N} = R^\dagger(\mathcal{I} - \mathcal{D})$, we have

$$|B| \leqslant |||L|||_\diamond |||R|||_\diamond \|\mathcal{I} - \mathcal{D}\|_\diamond^2 \frac{4^n}{4^n - 1} \leqslant 16 \times 4 \times 2 \leqslant 128,$$

proving that $B = O(1)$.

APPENDIX C: FITTED DECAY CURVES

In this Appendix we fit the exponential decays observed in Fig. 3 by doing a linear fit to the logarithm plots. As the linear XEB exhibits a two-phase behavior, we would like to perform the linear fit only on the decaying phase. However, the data for higher cycle numbers can become unstable due to random fluctuation. We thus pick ranges of data that visually appear to be a single exponential decay, and we list the chosen ranges below. We compare the extracted decay exponents to the fidelity we would expect if we assume MEA. The data for the 1D chain, 2D grid, and Clifford approximate twirl schemes are shown in Table I.

We see that the decay exponents and fidelities assuming MEA agree within a few percentage points except for the entries where the MEA fidelities are around 50% or less. For the Clifford approximate twirl, an exponential decay is not even identifiable for such cases. This is sensible since the perturbation theory used in proving Theorem 1 has an upper bound on the level of error that can be tolerated, and MEA fidelities for 50% or less probably exceed that bound. Even for the 1D chain and 2D grid schemes where we can always find an exponential decay, there are fluctuations for higher cycle numbers that may make the fit somewhat questionable.

APPENDIX D: CIRCUIT LOWER BOUNDS FOR γ-APPROXIMATE TWIRL

In this Appendix we prove lower bounds on the circuit required to implement γ-approximate twirls. We state our result in the form of a theorem:

TABLE I. Fitting data of the Clifford XEB experiments, corresponding to Figs. 3 and 5. Top part: Comparison between the fitted decay exponents, the expected fidelity calculated from multiplicative error accumulation (MEA), and the fitted noiseless scrambling rate (NSR). The exponential fits are performed with data selected from the cycle ranges below to eliminate data from the scrambling phase, and data where the random fluctuation is too large to recover the signal. Bottom part: Cycle ranges used to do the exponential fits to extract the decay exponents and the NSR. For Clifford approximate twirls, two cycles of the approximate twirl account to one cycle for the Clifford XEB experiment. N/A in both parts indicates failure to obtain an exponential fit.

(Decay exponent) [MEA]	Qubit number	Gate noise $(10^{-5}, 10^{-4})$	Gate noise $(10^{-4}, 10^{-3})$	Gate noise $(10^{-3}, 10^{-2})$	NSR Noiseless
1D chain	25	(99.36%) [99.73%]	(96.67%) [97.29%]	(75.10%) [75.86%]	83.32%
	100	(98.01%) [98.91%]	(85.51%) [89.58%]	(74.33%) [33.12%]	81.65%
	225	(95.23%) [97.56%]	(79.21%) [78.07%]	(73.76%) [8.32%]	82.76%
2D grid	25	(99.77%) [99.89%]	(98.66%) [98.86%]	(87.76%) [89.10%]	81.61%
	100	(99.44%) [99.47%]	(93.54%) [94.86%]	(80.10%) [58.88%]	81.64%
	225	(98.62%) [98.77%]	(86.53%) [88.33%]	(81.64%) [28.79%]	80.28%
2× Clifford approximate twirl	25	(98.77%) [98.63%]	(88.41%) [87.15%]	(N/A) [25.14%]	15.75%
	100	(94.95%) [94.65%]	(59.81%) [57.68%]	(N/A) [0.40%]	15.95%
	225	(89.50%) [88.36%]	(N/A) [29.00%]	(N/A) [0.0004%]	18.30%
Cycle range	Qubit number	Gate noise $(10^{-5}, 10^{-4})$	Gate noise $(10^{-4}, 10^{-3})$	Gate noise $(10^{-3}, 10^{-2})$	NSR Noiseless
1D chain	25	[22,50]	[22,50]	[22,36]	[20,30]
	100	[30,50]	[30,50]	[22,36]	[30,50]
	225	[30, 50]	[30,50]	[22,36]	[30,50]
2D grid	25	[30,50]	[30,50]	[30,50]	[20,30]
	100	[30,50]	[18,50]	[20,40]	[24,34]
	225	[30, 50]	[24,50]	[24,40]	[24,34]
2× Clifford approximate twirl	25	[5,13]	[16,50]	N/A	[2,6]
	100	[10,50]	[20,30]	N/A	[10,16]
	225	[10, 50]	N/A	N/A	[32,36]

Theorem 5. Suppose that the twirling map corresponding to a measure μ on $\mathrm{SU}(2^n)$ is a γ-approximate twirl. Suppose we implement unitaries $g \in \mathrm{SU}(2^n)$ sampled from μ by compiling them using single- and two-qubit gates, and let $s(g)$ and $t(g)$ be the number of gates and the depth for the circuit that implements g. Then

$$\mathbb{E}_{g\sim\mu} s(g) = \Omega[(1-\gamma)n], \mathbb{E}_{g\sim\mu} t(g) = \Omega[(1-\gamma)\log n]. \tag{D1}$$

Proof. This proof is adapted from [48]. We set the channel $\mathcal{N}$ so that $\mathcal{N}(\rho) = Z_1\rho Z_1$ where Z_1 is the Pauli Z gate applied to the first qubit. Then we know that the twirling map Λ_{Haar} maps $\mathcal{N}$ to the following channel:

$$\Lambda_{\mathrm{Haar}}(\mathcal{N}) : \rho \mapsto \frac{1}{4^n-1} \sum_{p\in\{0,1,2,3\}^n, p\neq 0^n} \sigma_p\rho\sigma_p = \frac{4^n}{4^n-1} \times \frac{I}{2^n}\mathrm{tr}[\rho] - \frac{1}{4^n-1}\rho. \tag{D2}$$

Let $\mathcal{D}$ be the completely depolarizing channel on n qubits

$$\mathcal{D}(\rho) = \frac{I}{2^n}\mathrm{tr}[\rho], \tag{D3}$$

then one can see that

$$\|\mathcal{D} - \Lambda_{\mathrm{Haar}}(\mathcal{N})\|_\diamond = O(4^{-n}). \tag{D4}$$

By definition

$$\Lambda_\mu(\mathcal{N}) : \rho \to \mathbb{E}_{g\sim\mu} g^\dagger Z_1 g\rho g^\dagger Z_1 g. \tag{D5}$$

Let $f(g)$ be the size of the support of $g^\dagger Z_1 g$, i.e., the number of qubits that $g^\dagger Z_1 g$ acts nontrivially on. For any $1 \leqslant j \leqslant n$, define p_j as the probability that qubit j is in the support of $g^\dagger Z_1 g$ for g sampled according to μ. Assuming $p_j \leqslant 1/2$, we have

$$\begin{aligned}
\gamma &\geqslant \|\Lambda_\mu(\mathcal{N}) - \Lambda_{\mathrm{Haar}}(\mathcal{N})\|_\diamond \\
&\geqslant \|\Lambda_\mu(\mathcal{N}) - \mathcal{D}\|_\diamond - O(4^{-n}) \\
&\geqslant \|[\Lambda_\mu(\mathcal{N}) - \mathcal{D}](|0^n\rangle\langle 0^n|)\|_1 - O(4^{-n}) \\
&\geqslant \left\|\mathrm{tr}_{\bar{j}}[(\Lambda_\mu(\mathcal{N}) - \mathcal{D})(|0^n\rangle\langle 0^n|)]\right\|_1 - O(4^{-n}) \\
&= \left\|p_j\rho_j + (1-p_j)|0\rangle\langle 0| - \frac{I}{2}\right\|_1 - O(4^{-n}) \qquad \text{(D6)} \\
&\geqslant \left\|p_j|1\rangle\langle 1| + (1-p_j)|0\rangle\langle 0| - \frac{I}{2}\right\|_1 - O(4^{-n}) \\
&= 1 - 2p_j - O(4^{-n}), \qquad \text{(D7)}
\end{aligned}$$

where $\mathrm{tr}_{\bar{j}}[\cdot]$ means tracing over all qubits other than qubit j, and in Eq. (D6) ρ_j refers to the state of qubit j conditioned that some unitary is applied to j in channel $\Lambda_\mu(\mathcal{N})$. This implies

$$p_j \geqslant \frac{1-\gamma}{2} - O(4^{-n}). \tag{D8}$$

Obviously Eq. (D8) is also satisfied when $p_j > 1/2$, so it always holds. Then

$$\mathbb{E}_{g\sim\mu} f(g) = \sum_j p_j \geqslant \frac{1-\gamma}{2} n - O(n4^{-n}). \tag{D9}$$

Note that for any g,

$$s(g) \geqslant f(g) - 1, t(g) \geqslant \log f(g) \geqslant \frac{f(g)-1}{n-1} \log n, \tag{D10}$$

so we have

$$\mathbb{E}_{g\sim\mu} s(g) = \Omega[(1-\gamma)n], \ \mathbb{E}_{g\sim\mu} t(g) = \Omega[(1-\gamma)\log n]. \tag{D11}$$

■

APPENDIX E: BOUNDS ON THE SPECTRAL NORMS

1. Proof of Theorem 3

In this Appendix we give a proof of Theorem 3. Specifically, following the proof of Theorem 16 in [35], it suffices to prove the following.

Proposition 1. Let $\Lambda^* = \int d\mu\, \omega^\dagger(g) \circ \cdot \circ \omega(g)$ be the noiseless twirling map associated to the probabilistic measure μ, and $\Lambda = \int d\mu\, \omega^\dagger(g) \circ \cdot \circ \phi(g)$ be its noisy implementation. Moreover, suppose that ϕ maps to mixtures of unitaries, that is, for each g there exists a probabilistic measure ν_g on $\mathrm{SU}(d)$ such that

$$\phi(g) = \int d\nu_g u.$$

Then

$$|||\Lambda^* - \Lambda|||_2 \leqslant \mathbb{E}_{g\sim\mu} \mathbb{E}_{u\sim\nu_g} \|\omega(g) - u\|_\diamond.$$

Proof. It can be verified that

$$\begin{aligned} |||\Lambda^* - \Lambda|||_2 &\leqslant \int d\mu \|\omega(g) - \phi(g)\|_2 \\ &\leqslant \int d\mu \int d\nu_g \|\omega(g) - u\|_2. \end{aligned}$$

It then suffices to prove that $\|\omega(g) - u\|_2 \leqslant \|\omega(g) - u\|_{\mathrm{tr}}$ for unitary channels $\omega(g)$ and u since it is known that $\|\omega(g) - u\|_{\mathrm{tr}} \leqslant \|\omega(g) - u\|_\diamond$.

Assume without loss of generality that $\omega(g) = \mathrm{Id}$ and $u = UU^\dagger$. Assume further that U is diagonalized over the computational basis: $U = \sum_j e^{i\lambda_j} |j\rangle\langle j|$. Then the trace norm and the 2-norm can then be computed explicitly as

$$\|\mathrm{Id} - u\|_{\mathrm{tr}} = \|\mathrm{Id} - u\|_2 = 2 \max_{i,j} \sin \frac{\lambda_i - \lambda_j}{2}.$$

■

2. Spectrally gapped schemes

In this section we provide a proof to Theorem 4. Our main tool is random walks on finite groups [56]. We first make the following definition:

Definition 5. Let G be a finite group. We say a measure μ over G is *mixing* over G if μ generates G, that is, every element of G can expressed as a finite product of elements in the support of μ, and μ is not contained in a coset of a proper normal subgroup of G.

This is interesting because from [56], we can show that this measure converges to the uniform distribution under iteration, where iteration is defined as

$$(\mu_1 * \mu_2)(g) := \int_{g_2\sim\mu_2} dg_2 \mu_1\left(g g_2^{-1}\right) \mu_2(g_2). \tag{E1}$$

We state this fact as a proposition:

Proposition 2 (Proposition 2.3 and Theorem 2.1 of [56]). Let G be a finite group and let μ be a measure that is mixing over G. Then, μ converges to the uniform distribution μ_U over G under iteration.

We can now prove the following theorem.

Theorem 6. Let μ be an inverse symmetric measure [that is, $\mu(g) = \mu(g^{-1})$ for all g] that is mixing over the n-qubit Clifford group $\mathcal{C}(n)$. Then,

$$|||\Lambda(\mu) - \Lambda(\mu_H)|||_2 < 1, \tag{E2}$$

where μ_H is Haar measure on $\mathrm{SU}(2^n)$.

Proof. Treating the twirling maps as linear maps on real superoperators, $\Lambda(\mu_H)$ is a projector onto the space spanned by the identity channel id and the completely depolarizing channel dep. These two channels are also eigenvectors of $\Lambda(\mu)$ with eigenvalue 1. Since the singular values of $\Lambda(\mu)$ are at most 1, $|||\Lambda(\mu) - \Lambda(\mu_H)|||_2$ is therefore the third largest singular value of $\Lambda(\mu)$.

Now, by the argument in the proof of Theorem 8 in [35], since μ is inverse symmetric, $\Lambda(\mu)$ is self-adjoint with respect to the SO inner product on real superoperators. Hence, it is diagonalizable by the spectral theorem. This implies the third largest singular value of $\Lambda(\mu)$ is third largest element of the elementwise absolute value of spec$[\Lambda(\mu)]$.

Next, it is easy to see that

$$\Lambda(\mu^{*k}) = \Lambda(\mu)^k. \tag{E3}$$

Next, since $\Lambda(\mu)$ is diagonalizable, the eigenvalues of $\Lambda(\mu)^k$ are the kth powers of the eigenvalues of $\Lambda(\mu)$. Two eigenvalues of $\Lambda(\mu)$ are 1, and every other eigenvalue has to be of magnitude at most 1.

Assume for contradiction that there is a third eigenvalue of $\Lambda(\mu)$ with magnitude 1. Then, for any k, $\Lambda(\mu)^k$ always has at least three eigenvalues with magnitude 1. However, by Proposition 2, for all $g \in \mathcal{C}(n)$,

$$\lim_{k\to\infty} \mu^{*k}(g) - \mu_U(g) = 0. \tag{E4}$$

Thus, for arbitrarily large k, the difference is arbitrarily small. Furthermore,

$$|||\Lambda(\mu^{*k}) - \Lambda(\mu_U)|||_2 \leqslant \max_{g\in\mathcal{C}(n)} [\mu^{*k}(g) - \mu_U(g)]|\mathcal{C}(n)|. \tag{E5}$$

Thus, for arbitrarily large k, the $|||\cdot|||_2$ norm is also arbitrarily small. Now, since groups are closed under inverses, μ_U is symmetric under inverse. Thus, $\Lambda(\mu_U)$ is self-adjoint. Furthermore, since $\Lambda(\mu)$ is self-adjoint, $\Lambda(\mu^{*k}) = \Lambda(\mu)^k$ is also self-adjoint. By Weyl's inequality on matrix perturbation, the eigenvalues of $\Lambda(\mu^{*k})$ and $\Lambda(\mu_U)$ can be paired up with a difference of at most $|||\Lambda(\mu^{*k}) - \Lambda(\mu_U)|||_2$. However, this

can be arbitrarily small as $k \to \infty$, and since $\mathcal{C}(n)$ is a unitary 2-design, two eigenvalues of $\Lambda(\mu_U) = \Lambda(\mu_H)$ are 1 and the rest are 0. This clearly contradicts $\Lambda(\mu)^k$ having at least three eigenvalues with magnitude 1. Hence, we conclude $\Lambda(\mu)$ has only two eigenvalues with magnitude 1.

This means the third largest element of the elementwise absolute value of spec$[\Lambda(\mu)]$ is strictly less than 1, which is the desired conclusion. ■

We can actually leverage this theorem to generic weakly mixing measures μ that are not inverse symmetric. Define the inverse measure

$$\mu^{-1}(g) := \mu(g^{-1}). \tag{E6}$$

We first show the following.

Proposition 3. Let μ be a probabilistic measure on a finite group G. Let $A := \mathrm{Supp}(\mu)$. Then the following are equivalent:

(1) A does not lie in any left coset of a proper subgroup of G;

(2) $\langle A^{-1}A\rangle = G$;

(3) $\mu^{-1} * \mu$ is mixing over G.

Proof. We follow the directions $1 \Leftrightarrow 2 \Leftrightarrow 3$.

(i) $1 \Rightarrow 2$. Assume otherwise that $\langle A^{-1}A\rangle = H < G$. Then for any $a, b \in A$, we have

$$a^{-1}b \in H \Rightarrow b \in aH \Rightarrow A \subseteq aH,$$

contradicting with 1.

(ii) $2 \Rightarrow 1$. Assume otherwise that $A \subseteq aH$ for some $a \in G, H < G$. Then $\langle A^{-1}A\rangle \subseteq H < G$, contradicting with 2.

(iii) $2 \Rightarrow 3$. It is easy to verify that $\mathrm{Supp}(\mu^{-1} * \mu) = A^{-1}A$. It suffices to prove that $\langle A^{-1}A\rangle$ does not lie in any coset of a normal subgroup aH of G. Assume otherwise then either that $1 \in aH = H$ and $\langle A^{-1}A\rangle \subseteq H < G$ contradicts with 2, or that $1 \notin aH$ contradicts with that $1 \in A^{-1}A$.

(iv) $3 \Rightarrow 2$. Easy to prove. ■

One can prove similarly a version with right coset, AA^{-1} and $\mu * \mu^{-1}$. We finally proceed to the proof of Theorem 4.

Proof of Theorem 4. Since μ is weakly gapped, then by Proposition 3 we know that either $\mu^{-1} * \mu$ or $\mu * \mu^{-1}$ is mixing. Hence, by Theorem 6,

$$|||\Lambda(\mu^{-1} * \mu) - \Lambda(\mu_H)|||_2 < 1. \tag{E7}$$

That is, the third largest magnitude of the spectrum of $\Lambda(\mu^{-1} * \mu)$ is less than 1. Since $\Lambda(\mu^{-1} * \mu) = \Lambda(\mu)^\dagger \Lambda(\mu)$, we conclude the third largest singular value of $\Lambda(\mu)$ is less than 1, which is the desired conclusion. ■

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[2] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan *et al.*, Strong quantum computational advantage using a superconducting quantum processor, Phys. Rev. Lett. **127**, 180501 (2021).

[3] Y. Kim, A. Eddins, S. Anand *et al.*, Evidence for the utility of quantum computing before fault tolerance, Nature **618**, 500 (2023)..

[4] H. Collins and C. Nay, Ibm unveils 400 qubit-plus quantum processor and next-generation ibm quantum system two, IBM Newsroom (unpublished).

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

[6] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, Characterizing universal gate sets via dihedral benchmarking, Phys. Rev. A **92**, 060302(R) (2015).

[7] 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).

[8] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, New J. Phys. **17**, 113020 (2015).

[9] J. J. Wallman, M. Barnhill, and J. Emerson, Robust characterization of leakage errors, New J. Phys. **18**, 043021 (2016).

[10] 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).

[11] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried *et al.*, Randomized benchmarking of multiqubit gates, Phys. Rev. Lett. **108**, 260503 (2012).

[12] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki *et al.*, Characterization of addressability by simultaneous randomized benchmarking, Phys. Rev. Lett. **109**, 240504 (2012).

[13] D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta, Three-qubit randomized benchmarking, Phys. Rev. Lett. **122**, 200502 (2019).

[14] 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).

[15] J. J. Wallman, Randomized benchmarking with gate-dependent noise, Quantum **2**, 47 (2018).

[16] S. T. Merkel, E. J. Pritchett, and B. H. Fong, Randomized benchmarking as convolution: Fourier analysis of gate dependent errors, Quantum **5**, 581 (2021).

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

[18] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B: Quantum Semiclassical Opt. **7**, S347 (2005).

[19] X. Gao, M. Kalinowski, C.-N. Chou, M. D. Lukin, B. Barak, and S. Choi, Limitations of linear cross-entropy as a measure for quantum advantage, arXiv:2112.01657.

[20] X. Mi, P. Roushan, C. Quintana, S. Mandrà, J. Marshall, C. Neill, F. Arute, K. Arya, J. Atalaya, R. Babbush *et al.*, Information scrambling in quantum circuits, Science **374**, 1479 (2021).

[21] Y. Liu, M. Otten, R. Bassirianjahromi, L. Jiang, and B. Fefferman, Benchmarking near-term quantum computers via random circuit sampling, arXiv:2105.05232.

[22] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, Caltech, 1997, arXiv:quant-ph/9705052.

[23] B. Lévi, C. C. López, J. Emerson, and D. G. Cory, Efficient error characterization in quantum information processing, Phys. Rev. A **75**, 022314 (2007).

[24] J. Jiang, X. Sun, S.-H. Teng, B. Wu, K. Wu, and J. Zhang, Optimal space-depth trade-off of cnot circuits in quantum logic synthesis, in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (SIAM, 2020), pp. 213–229.

[25] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[26] C. Gidney, Stim: a fast stabilizer circuit simulator, Quantum **5**, 497 (2021).

[27] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, Commun. Math. Phys. **291**, 257 (2009).

[28] P. Hayden and J. Preskill, Black holes as mirrors: Quantum information in random subsystems, J. High Energy Phys. 09 (2007) 120.

[29] Y. Sekino and L. Susskind, Fast scramblers, J. High Energy Phys. 10 (2008) 065.

[30] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, Towards the fast scrambling conjecture, J. High Energy Phys. 04 (2013) 022.

[31] W. Brown and O. Fawzi, Scrambling speed of random quantum circuits, arXiv:1210.6644.

[32] S. H. Shenker and D. Stanford, Black holes and the butterfly effect, J. High Energy Phys. 03 (2014) 067.

[33] A. W. Harrow and S. Mehraban, Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates, Commun. Math. Phys. **401**, 1531 (2023).

[34] N. Hunter-Jones, Unitary designs from statistical mechanics in random quantum circuits, arXiv:1905.12053.

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

[36] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Appl. Phys. Rev. **6**, 021318 (2019).

[37] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, Trapped-ion quantum computing: Progress and challenges, Appl. Phys. Rev. **6**, 021314 (2019).

[38] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits transform local noise into global white noise, arXiv:2111.14907.

[39] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, Phys. Rev. A **100**, 032328 (2019).

[40] 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).

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

[42] S. T. Flammia, Averaged circuit eigenvalue sampling, arXiv:2108.05803.

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

[44] 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).

[45] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Direct randomized benchmarking for multiqubit devices, Phys. Rev. Lett. **123**, 030503 (2019).

[46] T. Proctor, S. Seritan, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Scalable randomized benchmarking of quantum computers using mirror circuits, Phys. Rev. Lett. **129**, 150502 (2022).

[47] R. Jozsa and A. Miyake, Matchgates and classical simulation of quantum circuits, Proc. R. Soc. A: Math., Phys. Eng. Sci. **464**, 3089 (2008).

[48] R. Cleve, D. Leung, L. Liu, and C. Wang, Near-linear constructions of exact unitary 2-designs, Quantum Inf. Comput. **16**, 721 (2016).

[49] R. Oliveira, O. C. O. Dahlsten, and M. B. Plenio, Generic entanglement can be generated efficiently, Phys. Rev. Lett. **98**, 130502 (2007).

[50] M. Žnidarič, Exact convergence times for generation of random bipartite entanglement, Phys. Rev. A **78**, 032324 (2008).

[51] W. Brown and O. Fawzi, Decoupling with random quantum circuits, Commun. Math. Phys. **340**, 867 (2015).

[52] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, Commun. Math. Phys. **346**, 397 (2016).

[53] M. B. Hastings and J. Haah, Dynamically generated logical qubits, Quantum **5**, 564 (2021).

[54] M. Heinrich, M. Kliesch, and I. Roth, General guarantees for randomized benchmarking with random quantum circuits, arXiv:2212.06181.

[55] H. J. García, I. L. Markov, and A. W. Cross, On the Geometry of Stabilizer States, Quantum Inf. Comput. **14**, 683 (2014).

[56] L. Saloff-Coste, Random walks on finite groups, *Probability on Discrete Structures* (Springer, Berlin, 2004), pp. 263–346.