

Large-scale quantum device benchmarking via LXEB with particle-number-conserving random quantum circuits

Takumi Kaneda,^{1,*} Keisuke Fujii,^{1,2,3} and Hiroshi Ueda^{2,4}

¹Graduate School of Engineering Science, *The University of Osaka*, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

²Center for Quantum Information and Quantum Biology, *The University of Osaka*, 1-2 Machikaneyama, Toyonaka 560-0043, Japan

³*RIKEN Center for Quantum Computing (RQC)*, Hirosawa 2-1, Wako, Saitama 351-0198, Japan

⁴*RIKEN Center for Computational Science (R-CCS)*, Kobe, Hyogo 650-0047, Japan



(Received 25 July 2025; accepted 23 December 2025; published 26 January 2026)

Linear cross-entropy benchmarking (LXEB) with random quantum circuits is a standard method for evaluating quantum computers. However, LXEB requires classically simulating the ideal output distribution of a given quantum circuit with high numerical precision, which becomes infeasible beyond approximately 50 qubits, even on state-of-the-art supercomputers. As a result, LXEB cannot be directly applied to evaluate large-scale quantum devices, which now exceed 100 qubits and continue to grow rapidly in size. To address this limitation, we introduce a constraint known as particle-number conservation into the random quantum circuits used for benchmarking. This restriction significantly reduces the size of the Hilbert space for a fixed particle number, enabling classical simulations of circuits with over 100 qubits when the particle number is $O(1)$. Furthermore, we propose a modified version of LXEB, called MLXEB, which enables fidelity estimation under particle-number-conserving dynamics. Through numerical simulations, we investigate the conditions under which MLXEB provides accurate fidelity estimates.

DOI: [10.1103/ssrm-kddv](https://doi.org/10.1103/ssrm-kddv)

I. INTRODUCTION

Quantum computers are emerging as next-generation computational devices with the potential to solve problems that are intractable for conventional classical computers [1]. In recent years, experimental research in quantum computing has advanced rapidly, leading to several notable breakthroughs. For example, in 2019, Google reported the first demonstration of so-called quantum supremacy [2] by showing that a quantum processor could sample from the output distribution of a random quantum circuit faster than the most powerful classical supercomputers at that time. Subsequent studies have extended these demonstrations to even larger superconducting devices: Google's 67-qubit processor has been shown to achieve quantum supremacy in sampling tasks [3], and USTC's Zuchongzhi 3.0 platform—with 83 qubits—has similarly pushed the boundary of classically intractable sampling [4]. In the context of random-circuit sampling, these results indicate that quantum processors are venturing into regimes that exceed the capabilities of the most powerful classical simulators. Moreover, even quantum hardware other than superconducting architectures has reached comparable scales: Quantinuum's 56-qubit trapped-ion system has recently

executed complex circuits with high fidelity, underscoring the growing stability and scalability of ion-trap platforms [5].

However, the practical use of current quantum computers—commonly referred to as noisy intermediate-scale quantum (NISQ) devices—continues to face numerous challenges. One major issue is that quantum computers are inherently more susceptible to noise than classical computers, which makes it difficult to obtain error-free and accurate computational results. To mitigate the effects of such noise and to ensure reliable quantum computation, quantum error correction (QEC) techniques are essential. Notably, significant progress has been made in quantum error correction. In 2024, it was experimentally demonstrated for the first time that the lifetime of a logical qubit could exceed that of a physical qubit by employing surface codes [6]. Despite this milestone, current QEC methods require between 10^6 and 10^8 physical qubits to realize a fault-tolerant quantum computer (FTQC) [7]. In contrast, today's quantum devices are limited to approximately 156 qubits at most [8], highlighting a significant gap between the current state of hardware and the requirements for realizing FTQC.

Although the realization of FTQC remains a major technological challenge, a crucial intermediate step toward this goal is to develop a deep understanding of the properties of current NISQ devices [9,10]. In particular, many recent QEC schemes, such as those based on surface codes, are built on the assumption that errors affecting different qubits and gates are statistically independent [11,12]. However, in real devices, errors often exhibit spatial and temporal correlations—such as those arising from gate crosstalk—which can violate these assumptions and potentially degrade the performance of QEC

*Contact author: u007243b@ecs.osaka-u.ac.jp

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

protocols [13,14]. Therefore, developing benchmarking techniques that can quantitatively capture such correlated errors is fundamentally important for the design of future hardware and the realization of FTQC [15–18]. To address this issue, various benchmarking methods have been proposed. Traditional approaches, such as standard randomized benchmarking, can characterize systems up to a few qubits [18–21]. More scalable techniques, including cycle benchmarking and its variants, have been developed to assess the performance of larger quantum systems [16,17,22–25].

However, benchmarking protocols mentioned above are all built upon Clifford gates, which limits their applicability to more general circuits that incorporate non-Clifford operations. To address this limitation, we focus on linear cross-entropy benchmarking (LXEB), a protocol based on random circuit sampling, which was employed in Google’s demonstration of quantum supremacy [2,26]. LXEB offers a significant advantage over Clifford-based methods by enabling the estimation of overall circuit fidelity even for random quantum circuits that include non-Clifford gates [2,27]. Nonetheless, LXEB also faces several significant challenges. In particular, it requires the ideal output probability distribution obtained from a noiseless simulation of the circuit, which becomes exponentially more difficult to compute as the number of qubits increases. As a result, even with state-of-the-art supercomputers, classical simulations are practically limited to circuits with at most around 50 qubits [2]. Moreover, Google’s original work [2] reported that the circuit-level fidelity estimated by LXEB matched that predicted from gate-level fidelities, suggesting that correlated errors could be neglected. However, the theoretical justification for this observation remains unclear. Given these limitations, several recent studies have proposed extensions to LXEB aimed at improving its accuracy and scalability [28–30]. For instance, Liu *et al.* [28] demonstrated that LXEB does not overestimate the effects of correlated noise and supports layer-wise fidelity estimation. They further demonstrated that when correlations between errors are weak, the fidelity predicted from gate-level metrics agrees with that obtained using LXEB. Chen *et al.* [29] addressed the scalability issue of classical simulation by restricting circuits to Clifford gates, thereby enabling efficient and scalable benchmarking while retaining certain features of LXEB. Additionally, Cheng *et al.* [30] introduced the concept of ergodicity to strengthen the theoretical foundation of LXEB by analyzing its validity under the assumption of unitary $2t$ -designs, thereby providing new insights into the behavior of noisy random circuits and the design of quantum benchmarking protocols. While each of these previous works represents a significant advancement, there remains no benchmarking method that simultaneously achieves both scalability and compatibility with random quantum circuits containing non-Clifford gates.

To overcome the limitations of existing methods and enable benchmarking of quantum circuits with non-Clifford gates on devices with more than 100 qubits, we propose an approach based on particle-number-conserving random quantum circuits and a modified version of LXEB, which we call MLXEB. Specifically, the particle-number-conserving random quantum circuits considered in this study are those in which the number of excitations (i.e., qubits in the $|1\rangle$ state)

remains constant throughout the computation. Such conservation laws naturally arise in physical quantum many-body systems with symmetries, such as fermionic systems, and the corresponding circuits are built using gates that preserve particle number—for example, two-qubit gates with $U(1)$ symmetry. Physical symmetries such as $U(1)$ symmetry have been effectively utilized in quantum computing. In particular, in variational quantum algorithms such as the variational quantum eigensolver, explicitly incorporating symmetries can drastically reduce the dimensionality of the Hilbert space to be explored. This, in turn, helps mitigate optimization challenges such as the barren plateau problem [31–37]. However, to the best of our knowledge, no prior work has explored the use of such structured circuits for benchmarking quantum hardware.

Unlike general random quantum circuits, imposing a particle-number conservation constraint restricts the entanglement structure of the output state. In particular, when the particle number is $O(1)$, computing the ideal output probability distribution becomes significantly easier via classical simulation. This provides a major advantage in terms of scalability. MLXEB leverages this feature by benchmarking quantum devices using the cross entropy between the bit-string distribution obtained from the actual hardware and the ideal distribution computed classically under the particle-number-preserving constraint. By doing so, MLXEB enables experimental fidelity estimation for large-scale quantum systems, which would be infeasible using conventional LXEB due to the prohibitive cost of classical simulation.

In our numerical experiments, we executed noiseless, particle-number-conserving random quantum circuits on a square lattice for systems ranging from $N = 36$ to 196 qubits. We confirmed that, with sufficient circuit depth, the output distributions converge to the Porter-Thomas distribution, and the fidelity estimated by MLXEB, F_{MLXEB} , asymptotically approaches the ideal value $F = 1$. We evaluated MLXEB under two circuit configurations: one in which the two-qubit gate rotation angles were chosen randomly, and another in which they were fixed to specific values. In both cases, F_{MLXEB} exhibited behavior consistent with theoretical expectations. Notably, although it is unclear whether circuits with fixed gate angles form a unitary 2-design, MLXEB still performed effectively, suggesting that the method remains valid even for circuits with limited randomness. Furthermore, under a depolarizing noise model, we found that F_{MLXEB} matched the true fidelity when the noise rate was sufficiently low ($\sim 10^{-3}$). These results collectively demonstrate that introducing physical symmetries—such as particle-number conservation—can significantly extend the range of quantum circuits that are classically simulatable. Moreover, under this constraint, a benchmarking metric similar to LXEB can still reliably estimate circuit fidelity. In particular, we have verified that noiseless simulations remain feasible for systems with up to 1000 qubits, suggesting that the applicability of MLXEB may extend well beyond the scale studied in this study. Leveraging this classical simulability, and considering typical experimental gate fidelities around 99.9% (corresponding to a noise rate of $\sim 10^{-3}$) and associated gate-count overhead, MLXEB can practically benchmark devices with at least approximately 70 qubits, thereby accessing regimes beyond the reach of

conventional LXEB. By bridging the gap between classical simulability and experimental feasibility, MLXEB offers a practical and theoretically grounded approach to benchmarking near-term quantum devices—enabling systematic fidelity assessments for complex circuits that incorporate non-Clifford operations.

The remainder of this paper is organized as follows. In Sec. II, we review the types of random quantum circuits used in LXEB and the fidelity-related quantities that are evaluated. Section III introduces the particle-number-conserving random quantum circuits required for MLXEB, and defines the evaluation metric for fidelity estimation under this constraint. In Sec. IV, we describe the specific circuit setup and present numerical validation of MLXEB. Finally, Sec. V provides a summary of our findings and discusses future directions. Technical details of the MLXEB estimator and its relation to state fidelity are given in Appendix A, practical rules for choosing the particle number n are compiled in Appendix B, and a practitioner-oriented comparison with randomized benchmarking variants and elided-circuit LXEB is provided in Appendix C.

II. LXEB USING RANDOM QUANTUM CIRCUITS

A. Properties required for random quantum circuits

Let $\hat{U}$ be a unitary operator representing a quantum circuit acting on an N -qubit system with Hilbert space dimension $D = 2^N$ and computational basis $\{|0\rangle, |1\rangle\}^{\otimes N}$. Applying $\hat{U}$ to the initial quantum state $|\psi_0\rangle = |0\rangle^{\otimes N}$ yields the final state $|\psi\rangle = \hat{U}|\psi_0\rangle$. Upon measuring all qubits of $|\psi\rangle$ in the computational basis, the system collapses to a product state $|x\rangle = |b_N\rangle \otimes \cdots \otimes |b_1\rangle$, where $b_\ell \in \{0, 1\}$ denotes the measurement outcome of the ℓ th qubit for $\ell \in [1, N]$. We define x as the integer corresponding to the bit string $(b_N b_{N-1} \cdots b_1)$. In the absence of noise or other imperfections, the probability of obtaining a particular outcome x is given by

$$p(x) = |\langle x | \psi \rangle|^2. \quad (1)$$

Under the assumption of a sufficiently large Hilbert space, if the unitary operator $\hat{U}$ is sampled uniformly at random from the set of all unitary operators defined on that space, the resulting distribution $\mathcal{P}(p)$ of the probability p converges to the Porter-Thomas (PT) distribution [27]:

$$\mathcal{P}_T(p) = D e^{-Dp}. \quad (2)$$

In order for random quantum circuits to be used for fidelity estimation in LXEB, the ideal output probability distribution must follow $\mathcal{P}_T$. Therefore, it is essential to verify, within the regime where the classical simulation of the quantum circuit is tractable, that the user-specified circuit structure indeed produces this expected distribution.

B. LXEB

By executing a random quantum circuit on a real quantum device and performing N_s repeated measurements in the computational basis, one obtains a set of bit strings represented as integers, denoted by $\{x_i\}_{1 \leq i \leq N_s}$. LXEB evaluates the fidelity

using these outcomes according to the following formula:

$$F_{\text{LXEB}} = \frac{2^N}{N_s} \left(\sum_{i=1}^{N_s} p_{\text{sim}}(x_i) \right) - 1, \quad (3)$$

as originally proposed in Ref. [2]. Here, $p_{\text{sim}}(x_i)$ denotes the probability of observing the bit string x_i from the ideal output state of the random quantum circuit. This quantity must be computed in advance using a classical simulator of the quantum circuit.

An important property of F_{LXEB} is that it coincides with the fidelity

$$F = \left(\text{tr} \sqrt{\sqrt{\hat{\rho}} |\psi\rangle\langle\psi| \sqrt{\hat{\rho}}} \right)^2 \quad (4)$$

in the limit of a large number of samples $N_s \gg 1$, provided that the noise rate is sufficiently low and the output distribution $p(x)$ can be modeled as a linear combination of the Porter-Thomas distribution and the uniform distribution induced by depolarizing noise. Here, $\hat{\rho}$ denotes the density matrix representing the mixed state resulting from the effect of noise, and $|\psi\rangle$ is the ideal output state in the absence of noise. Another well-known property of F_{LXEB} is that, in the absence of noise ($p_{\text{noise}} = 0$), sampling from the output distribution yields $F_{\text{LXEB}} = 1$. As p_{noise} increases, the Porter-Thomas distribution gradually breaks down, and the output distribution approaches the uniform distribution. In this regime, F_{LXEB} asymptotically approaches 0, and no longer coincides with the fidelity, which instead saturates at the lower bound $1/2^N$ due to the completely mixed output state. For a side-by-side summary of what each protocol measures and under which assumptions—covering randomized benchmarking variants as well as elided-circuit LXEB—see Appendix C.

III. MLXEB USING PARTICLE-CONSERVING RANDOM QUANTUM CIRCUITS

A. Particle-conserving random quantum circuits

As a quantity characterizing the computational basis state $|x\rangle$, we define the particle number n as the expectation value

$$n = \langle x | (N - \hat{Z}_{\text{tot}}) / 2 | x \rangle, \quad (5)$$

where $\hat{Z}_{\text{tot}} = \sum_{\ell=1}^N \hat{Z}_\ell$ and $\hat{Z}_\ell$ is the Pauli-Z operator acting on the ℓ -th qubit. Then, we also define the set

$$G_n = \{x | \langle x | (N - \hat{Z}_{\text{tot}}) / 2 | x \rangle = n\} \quad (6)$$

to specify the subspace that conserves the particle number. The number of elements of G_n is given by $D_n = \binom{N}{n}$. A unitary operator $\hat{U}'$ that preserves particle number can be written as

$$\hat{U}' = \sum_{n=0}^N \left(\sum_{x \in G_n} \sum_{x' \in G_n} U_{x,x'} |x\rangle\langle x'| \right), \quad (7)$$

and satisfies unitarity within each particle-number n subspace:

$$\sum_{x' \in G_n} U_{x,x'} U_{x'',x'}^* = \delta_{x,x''}, \quad \text{for all } x, x'' \in G_n. \quad (8)$$

An advantage of considering $\hat{U}'$ is that, when it acts on a quantum state expressed as a linear combination of

computational basis states $\{|x\rangle\}_{x \in G_n}$ with fixed particle number n , the resulting state remains within the same subspace and can also be represented as a linear combination of $\{|x\rangle\}_{x \in G_n}$. Therefore, if such a state is chosen as the initial state, the simulation of the corresponding quantum circuit can be performed by tracking a complex vector of dimension D_n . When n remains relatively small, this enables the simulation of quantum circuits with more than 100 qubits on modern classical computers.

To simplify the discussion, we adopt a specific product state $|o_n\rangle \in G_n$ as the initial state, defined as

$$|o_n\rangle = \prod_{k=1}^n \hat{X}_{i_k} |\psi_0\rangle, \quad (9)$$

where $\{i_k\}$ is a set of integers satisfying $1 \leq i_1 < i_2 < \dots < i_n \leq N$. Applying the particle-number-conserving unitary operator $\hat{U}'$ to the initial state $|o_n\rangle$ results in the final state $|\psi_n\rangle = \hat{U}'|o_n\rangle$. The corresponding noise-free probability distribution is defined as $p_n(x) = |\langle x|\psi_n\rangle|^2$. Note that $p_n(x)$ is identically zero for $x \notin G_n$, and it depends not only on the circuit $\hat{U}'$ but also on the choice of the initial configuration o_n . Following the discussion in Sec. II A, when the dimension D_n of the particle-number- n subspace becomes sufficiently large (e.g., in systems with large N), and $\hat{U}'$ is sampled uniformly at random from the space of particle-number-conserving unitaries, the probability distribution p_n converges to the particle-number-rescaled Porter-Thomas distribution:

$$\mathcal{P}_n(p_n) = D_n e^{-D_n p_n}, \quad (10)$$

as shown in Ref. [38,39].

B. MLXEB

In this section, we introduce a phenomenological modification of the fidelity estimator F_{LXEB} , discussed in Sec. II B, to enable its application to random quantum circuits that conserve particle number. Specifically, we define the modified linear cross-entropy benchmarking (MLXEB) for a given particle number n as

$$F_{\text{MLXEB},n} = \frac{n_s}{N_s} \left[\frac{D_n}{n_s} \left(\sum_{j=1}^{n_s} p_n(x'_j) \right) - 1 \right], \quad (11)$$

where n_s denotes the number of measurement outcomes (out of N_s total samples) that belong to the particle-number- n subspace and $\{x'_j\}_{1 \leq j \leq n_s}$ is the set of corresponding bit-string indices. Each $p_n(x'_j)$ represents the ideal (noise-free) probability of observing x'_j , as defined in the particle-number-conserving scenario.

When $p_{\text{noise}} = 0$, $F_{\text{MLXEB},n}$ converges to 1 in the limit of a sufficiently large number of samples $n_s \gg 1$, just as F_{LXEB} does in the noiseless case. This is because, in the absence of noise, $n_s = N_s$, and the definition of $F_{\text{MLXEB},n}$ becomes equivalent to that of F_{LXEB} in Eq. (3), with the total Hilbert space dimension D replaced by the subspace dimension D_n . Similarly, the Porter-Thomas distribution $\mathcal{P}$ is replaced by its particle-number-rescaled counterpart $\mathcal{P}_n$ under this constraint. In the opposite limit of depolarizing noise, the measured

bitstrings are sampled from the uniform distribution $p(x) = D^{-1}$. Assuming $n_s \gg 1$, the summation $\sum_{j=1}^{n_s} p_n(x'_j)$ converges to N_s/D , while the fraction n_s/N_s converges to D_n/D . Substituting these into the definition of $F_{\text{MLXEB},n}$ shows that it asymptotically approaches 0 in this high-noise regime.

As noted above, $F_{\text{MLXEB},n}$ reproduces the limiting behavior of F_{LXEB} in both the noiseless and fully depolarizing regimes. In Appendix A, we further show that, under in-sector depolarizing noise and Porter-Thomas (unitary 2-design) statistics within the particle-number- n subspace, the expectation of $F_{\text{MLXEB},n}$ coincides with

$$\mathbb{E}[F_{\text{MLXEB},n}] = (1 - L_n) F_n, \quad (12)$$

where we decompose the noisy output state as $\rho_U = (1 - L_n)\rho_n + L_n\rho_{\text{leak}}$ with $L_n \in [0, 1]$ the leakage probability out of $\mathcal{H}_n$, ρ_n supported on $\mathcal{H}_n$, and define the in-sector state fidelity $F_n := \langle \psi_{\text{ideal},n} | \rho_n | \psi_{\text{ideal},n} \rangle$. Thus, $(1 - L_n)F_n$ is the circuit-level fidelity to the ideal target within the sector, and $F_{\text{MLXEB},n}$ is an unbiased estimator of this quantity under the stated conditions (see Appendix A for details).

For practical use on finite-depth, noisy devices, however, it remains important to characterize (1) finite-shot effects and deviations from the Porter-Thomas assumption, and (2) the rate at which different circuit constructions approach the rescaled Porter-Thomas distribution required for MLXEB. In particular, for a given hardware layout and native gate set, there is no simple analytic criterion for how the choice of two-qubit gate parameters or the initial particle position governs the scrambling rate within the sector. To address these implementation-level questions, the following sections present numerical simulations that mimic realistic experimental settings and examine representative design choices (e.g., fixed versus random two-qubit angles and bulk versus boundary seeding of the initial particle), in order to quantify the onset of the Porter-Thomas regime and the range of noise rates over which $F_{\text{MLXEB},n}$ provides a reliable estimate of circuit fidelity. Practical guidance on selecting the initial particle number n —balancing sector mixing depth, statistical resolvability, and classical simulability—is provided in Appendix B. For “when to use what” guidance that positions MLXEB relative to randomized benchmarking variants and elided-circuit LXEB on real hardware, see Appendix C.

MLXEB should therefore be interpreted as a scalable proxy for the circuit-level fidelity of the device in a $U(1)$ -symmetric operating regime (particle-number-conserving gates). Under the noise model of Appendix A, the MLXEB estimator coincides with the circuit fidelity $(1 - L_n)F_n$ within the particle-number- n subspace. More generally, when the dominant noise mechanisms are not strongly sensitive to small symmetry-breaking components of the entangling gate, and when agreement between MLXEB and standard XEB is verified on classically simulable subgraphs, it is reasonable to regard MLXEB as an informative proxy for the fidelity of nearby nonconserving circuits at comparable depth. We emphasize that MLXEB does not, by itself, certify the fidelity of arbitrary, fully generic supremacy-type circuits that explore the entire Hilbert space; rather, it provides a scalable and practically relevant benchmark for symmetry-conserving workloads and an experimentally accessible proxy for device

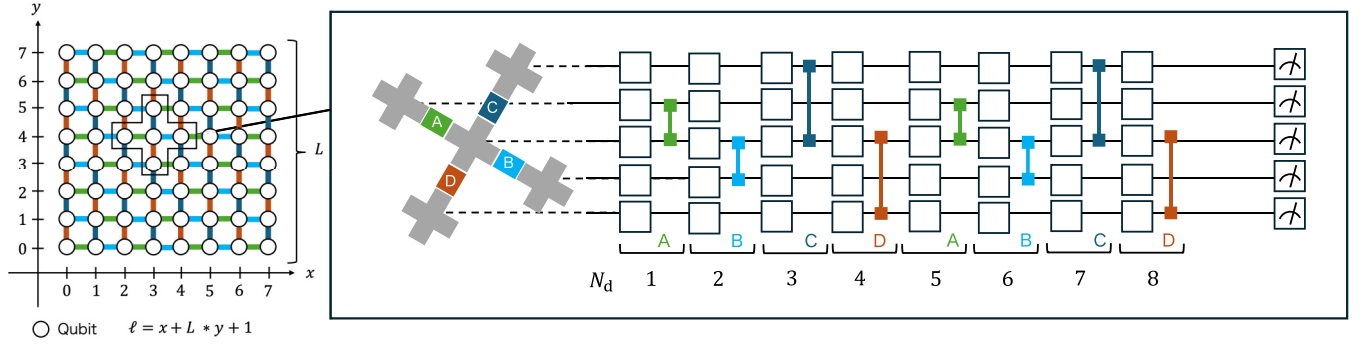


FIG. 1. Unit structure of a particle-number-conserving random quantum circuit on a square lattice. A single cycle consists of four layers labeled A, B, C, and D, which together define one unit of circuit depth.

performance in the vicinity of a $U(1)$ -symmetric operating point.

IV. VALIDATION BY NUMERICAL SIMULATIONS

A. Construction of particle-conserving random quantum circuits

This work focuses on particle-number-conserving random quantum circuits arranged on a two-dimensional square lattice. We numerically simulate quantum circuits that implement the unitary operator $\hat{U}'$, which is composed of layers of two-qubit gates that preserve $U(1)$ symmetry, equivalent to particle-number conservation. Even within this restricted class, it has been shown that the output probability distribution still converges to the Porter-Thomas distribution [38,39]. An overview of the layer-wise structure of the circuit used in our numerical experiments is shown in Fig. 1. In this figure, the white square symbols represent single-qubit R_z gates, where each qubit is assigned an independent and uniformly random rotation angle $\theta \in [0, 2\pi)$. The two-qubit gates, illustrated as colored dumbbell shapes, are XXZ-type gates derived from the canonical decomposition [40,41], with a shared XY interaction. The gates are defined by

$$w(\alpha_1, \alpha_2) = \exp(i[\alpha_1(X \otimes X + Y \otimes Y) + \alpha_2 Z \otimes Z]), \quad (13)$$

where i denotes the imaginary unit.

We consider two types of parameter settings for the two-qubit gates: (1) random angles $\alpha_1, \alpha_2 \in [0, 2\pi)$ drawn independently and uniformly, and (2) fixed angles $\alpha_1 = \pi/8$ and $\alpha_2 = 0$. In the random-angle case, it has been shown that the circuit forms a unitary t -design within the fixed particle-number subspace [38,39], thereby ensuring convergence to the Porter-Thomas distribution. Recent work by Suzuki *et al.* [42] reported that fixing the parameters of two-qubit gates can enhance global randomness in a shallow-depth circuit and accelerate convergence to the Porter-Thomas distribution by better approximating a unitary 2-design [20,43]. While these results were established in the context of generic random quantum circuits without symmetry constraints, we are particularly interested in $U(1)$ -symmetric random quantum circuits, where such convergence properties have not yet been fully explored. Motivated by these findings, we hypothesize that introducing fixed-angle two-qubit gates may similarly enhance convergence to the Porter-Thomas distribution even

under $U(1)$ symmetry constraints. To test this hypothesis, we compare random-angle and fixed-angle configurations and examine their impact on the convergence rate. In the context of MLXEB, achieving convergence with shallow circuit depths is particularly valuable, as it enhances the robustness of the benchmark against noise (i.e., higher p_{noise}). In our circuits, two-qubit gates play the central role in directly generating entanglement and superposition between distinct qubits. Therefore, when using fixed-angle configurations, the choice of gate parameters becomes critical. In this study, we adopt $\alpha_1 = \pi/8$ and $\alpha_2 = 0$ to ensure that the two-qubit gates effectively generate entanglement by transforming input states such as $|1\rangle \otimes |0\rangle$ or $|0\rangle \otimes |1\rangle$ into maximally entangled states. This enables the realization of the Porter-Thomas distribution even with shallow-depth circuits. However, under the fixed-angle constraint, whether the output distribution still converges to the rescaled Porter-Thomas form remains nontrivial and must therefore be verified numerically. The scheduling of the XXZ-type two-qubit gates in layers A, B, C, and D follows the same pattern as that used in the original LXEB benchmarking study [2].

In the MLXEB protocol for a given particle number n , the quantum circuit follows the sequence: (1) applying n single-qubit X gates to construct the initial state $|o_n\rangle$; (2) performing N_{cycle} cycles of single- and two-qubit gate layers in the order $A \rightarrow B \rightarrow C \rightarrow D$; and (3) measuring all qubits in the computational basis. Depolarizing noise [44] with probability p_{noise} is independently applied to each gate operation and measurement: (1) after each single-qubit; (2) after each measurement operation; and (3) after each two-qubit gate. Due to the high computational cost of simulating noisy circuits using the density matrix formalism, we instead adopt the stochastic method (also known as the Monte Carlo wave function method) [45]. Specifically, one of the Pauli operators X , Y , or Z is randomly applied with probability p_{noise} after each single-qubit gate and before each measurement operation. Similarly, after each two-qubit gate, a two-qubit Pauli operator randomly selected from the set $\{I, X, Y, Z\}^{\otimes 2} \setminus \{I \otimes I\}$ is applied with the same probability. The noise pattern is independently re-sampled for each of the N_s circuit executions. Under sufficient sampling, this method is known to yield results that are statistically equivalent to those obtained from full density matrix simulation [2]. Similarly, the fidelity defined in Eq. (4) can be statistically estimated by averaging the overlap $|\langle \psi_{\text{noise}} | \psi_n \rangle|^2$

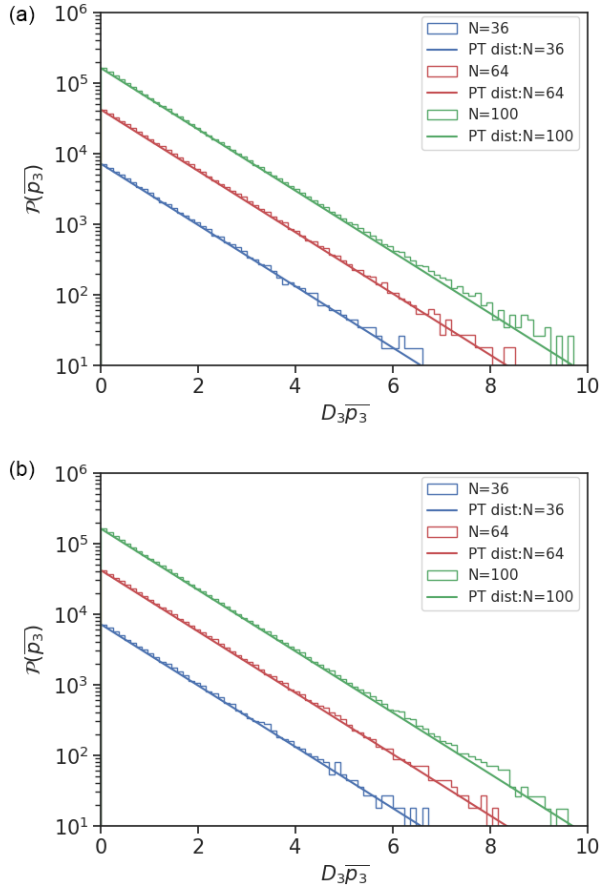


FIG. 2. Probability distributions of output probabilities of $n = 3$ from particle-number-conserving random quantum circuits under two different gate parameter settings: (a) two-qubit gate rotation angles α_1 and α_2 are randomly selected from the interval $[0, 2\pi)$; (b) fixed parameters $\alpha_1 = \pi/8$ and $\alpha_2 = 0$. The bin width for the histogram is set to $\Delta(D_3 \bar{p}_3) = 0.12$. See Sec. IV B for additional computational conditions.

between the noisy final state $|\psi_{\text{noise}}\rangle$ —generated by circuits with inserted Pauli noise gates—and the ideal, noiseless state $|\psi_n\rangle$.

B. Validation of the Porter-Thomas distribution

Figure 2 presents results verifying whether the quantum circuits constructed in Sec. IV A behave as suitable random quantum circuits for MLXEB evaluation under noiseless condition ($p_{\text{noise}} = 0$). We consider system sizes of $N = 6 \times 6$, 8×8 , and 10×10 , each with a fixed particle number of $n = 3$. The depth per cycle is set to 4, and the total circuit depth is $N_d = 160$, which is sufficiently deep relative to the system size. The number of random-circuit instances is set to $N_c = 10$. In Fig. 2, the averaged probability distribution $\bar{p}_3$ is computed as the ensemble average over N_c instances, with the probability vector for each instance sorted in ascending order before averaging. The state-vector simulations for fixed-particle-number subspaces in large-scale systems are performed using a custom Python code built upon internal routines from the quantum spin simulator QS³ [46,47]. This

implementation enables explicit access to the full probability distribution p_3 even for systems as large as 100 qubits.

The simulation results confirm that the probability distribution $\mathcal{P}$ is well described by the rescaled Porter-Thomas distribution $\mathcal{P}_n = D_n e^{-D_n p_n}$ for systems with 36–100 qubits, regardless of whether the two-qubit gate rotation angles are randomly chosen or fixed. This agreement holds under sufficiently large circuit depth N_d , indicating that the output statistics $\mathcal{P}$ converge to the expected Porter-Thomas behavior in both configurations.

C. MLXEB evaluation in the absence of noise and size dependence

Having confirmed in Sec. IV B that the constructed random quantum circuits exhibit the desired statistical properties, we now proceed to evaluate MLXEB. In the noiseless case ($p_{\text{noise}} = 0$), the modified linear cross-entropy benchmarking value $F_{\text{MLXEB},n}$ is expected to asymptotically converge to the ideal fidelity $F = 1$ as the total circuit depth N_d increases. Figure 3 shows the dependence of $F_{\text{MLXEB},n}$ on the circuit depth N_d for particle numbers $n = 1, 2$, and 3 under noiseless conditions.

For all system sizes considered, we observe that $F_{\text{MLXEB},1}$, $F_{\text{MLXEB},2}$, and $F_{\text{MLXEB},3}$ exhibit exponential convergence toward unit fidelity as the circuit depth N_d becomes comparable to the number of qubits N . In the shallow-circuit regime (small N_d), the output quantum state remains insufficiently randomized, and the probability distribution $\mathcal{P}$ deviates from the Porter-Thomas distribution. Consequently, MLXEB fails to accurately characterize the circuit fidelity in this regime.

To quantitatively examine how the circuit depth N_d required for the emergence of asymptotic behavior depends on the system size N and the particle number n , we replot the data using the following finite-size scaling relation (Fig. 4):

$$f(N_d) = \left[\frac{F_{\text{MLXEB},n} - 1}{D_n - 2} \right]^{\frac{1}{n+a}}. \quad (14)$$

Here, the scaling function $f(N_d)$ is defined based on the asymptotic behavior of $F_{\text{MLXEB},n}$:

$$F_{\text{MLXEB},n}|_{N_d=0} = D_n - 1, \quad F_{\text{MLXEB},n}|_{N_d \rightarrow \infty} = 1. \quad (15)$$

From these properties, $f(N_d)$ must asymptotically approach 1 as $N_d \rightarrow 0$, and 0 as $N_d \rightarrow \infty$. The derivation of the scaling function $f(N_d)$, which is independent of the system size N and the particle number n , is based on the following three steps:

(1) Subtract the trivial asymptotic value 1 from $F_{\text{MLXEB},n}$, to which it converges as $N_d \rightarrow \infty$.

(2) Normalize $F_{\text{MLXEB},n} - 1$ by dividing by $D_n - 2$, so that the result becomes 1 at $N_d = 0$, regardless of N and n .

(3) Under the condition $n \ll N$, the quantum state in the basis $\{|x\rangle \in G_n\}$ expands to approximately $O(2^n)$ states under the action of a single layer (A, B, C, or D) of the quantum circuit. To eliminate the n dependence of the convergence rate of $F_{\text{MLXEB},n}$ toward 1, the exponent is divided by a linear function of n , specifically $n + a$, where a is a fitting parameter.

Through the above three steps, we expect the behavior of the rescaled function near both limits, $N_d = 0$ and $N_d \rightarrow \infty$, to be independent of N and n . Whether the scaling function $f(N_d)$ accurately captures the behavior across the full range of

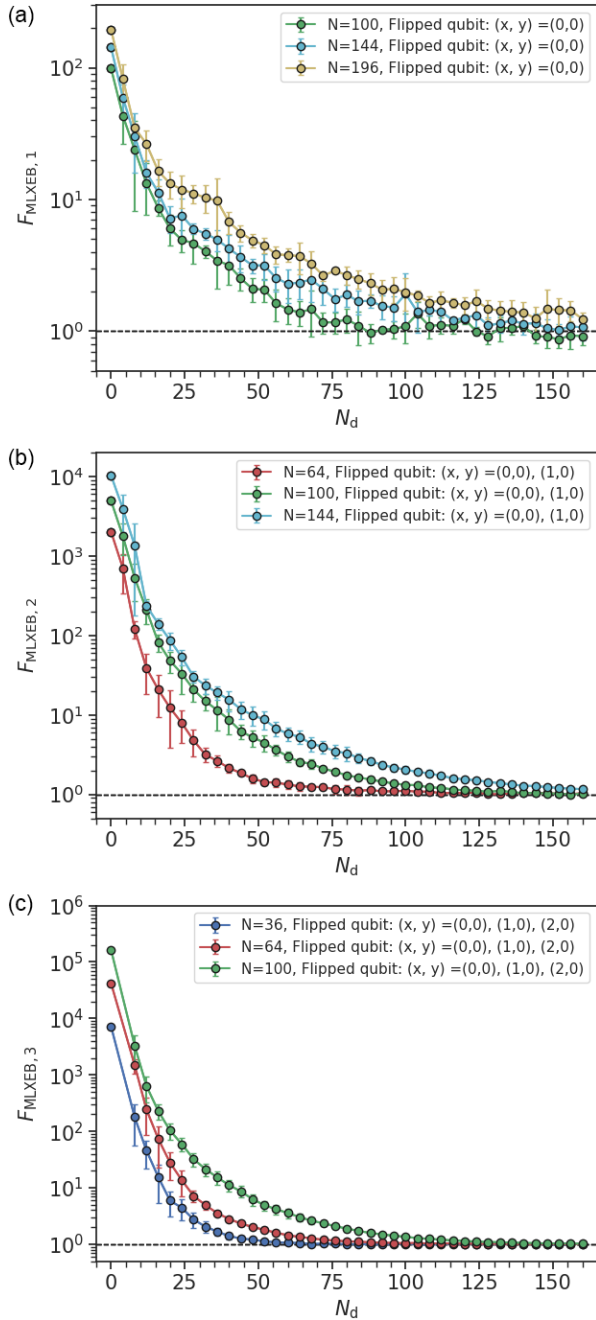


FIG. 3. Circuit depth dependence of (a) $F_{\text{MLXEB},1}$, (b) $F_{\text{MLXEB},2}$, and (c) $F_{\text{MLXEB},3}$ under noiseless conditions ($p_{\text{noise}} = 0$). The initial state is prepared by applying X gates to qubits at positions $(0, 0), \dots, (n-1, 0)$ of the vacuum state $|\psi_0\rangle$. The rotation angles of the two-qubit gates are randomly chosen from the interval $\alpha_1, \alpha_2 \in [0, 2\pi)$. The number of random-circuit instances and the number of samples per instance are set to $(N_c, N_s) = (5, 100\,800)$. The dashed line indicates the ideal fidelity $F = 1$, which is trivially achieved in the absence of noise.

N_d must be verified by examining the actual data, particularly in the intermediate-depth regime.

As shown in Fig. 4, when $a = 0.3$ was chosen, the behavior of $F_{\text{MLXEB},n}$ for various combinations of N and n is consistently captured by the single scaling function $f(N_d)$ across

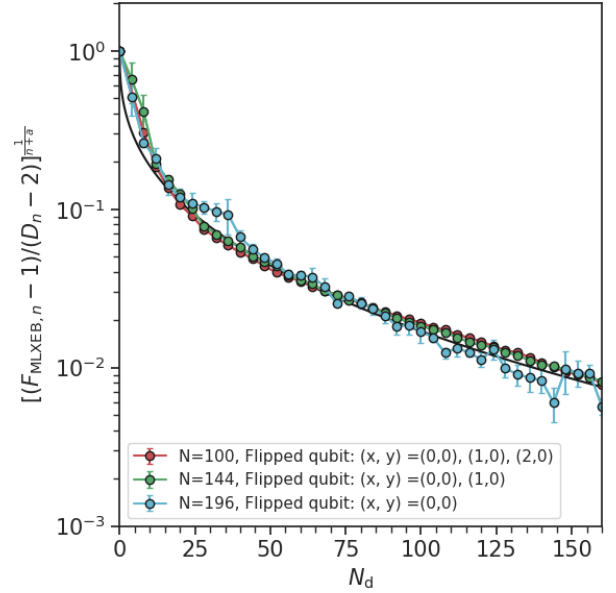


FIG. 4. Rescaled plot of the data from Fig. 3, based on the finite-size scaling relation given in Eq. (14). The resulting data collapse indicates a universal scaling behavior of $F_{\text{MLXEB},n}$ across different system sizes and particle numbers.

the entire range of circuit depths N_d . This confirms that the proposed scaling relation accurately captures the fidelity convergence behavior of MLXEB. Moreover, the overall shape of $f(N_d)$ suggests that it follows a stretched-exponential form. To support this, we fit the data for $(N, n) = (100, 3)$, $(144, 2)$, and $(196, 1)$ over the range $N_d = 0$ –160 using the following relation:

$$\left[\frac{F_{\text{MLXEB},n} - 1}{D_n - 2} \right]^{\frac{1}{n+a}} = \exp \left[- \left(\frac{N_d}{\tau} \right)^\beta \right]. \quad (16)$$

This yields the fitting parameters $\tau = 2.64 \pm 0.24$ and $\beta = 0.387 \pm 0.010$. This scaling relation is particularly useful for estimating the circuit depth N_d required to achieve a desired level of fidelity in MLXEB for larger quantum systems not considered in this study. It also provides a practical guideline for designing shallow yet reliable benchmarking circuits in large-scale quantum devices.

Next, we compare the behavior of $F_{\text{MLXEB},1}$ for a system with $N = 64$ and $n = 1$ under four distinct circuit configurations. Specifically, we consider the following cases: (1) fixed two-qubit gate parameters $\alpha_1 = \pi/8$, $\alpha_2 = 0$; (2) randomly chosen gate parameters $\alpha_1, \alpha_2 \in [0, 2\pi)$; and for each, we compare the effect of placing the initial particle either near the center of the lattice at position $(\sqrt{N}/2 - 1, \sqrt{N}/2 - 1)$ or at the corner $(0, 0)$. The corresponding results are presented in Fig. 5. Regardless of whether the two-qubit gate parameters are fixed or randomized, placing the initial particle near the center of the lattice leads to faster convergence of $F_{\text{MLXEB},1}$ to unity at smaller circuit depths N_d . This observation is consistent with a qualitative picture based on circuit light cones: the circuit depth required to produce a sufficiently random quantum state—i.e., one in which the state effectively explores the entire Hilbert space—must be at least comparable to the depth needed for the causal light cone, originating from

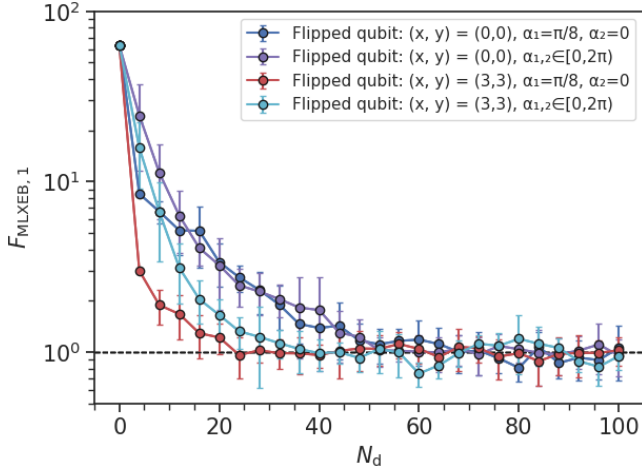


FIG. 5. Circuit depth dependence of $F_{\text{MLXEB},1}$ under noiseless conditions ($p_{\text{noise}} = 0$) for a system of $N = 64$ qubits. Four configurations are compared based on two binary factors: (1) whether the two-qubit gate parameters are fixed ($\alpha_1 = \pi/8, \alpha_2 = 0$) or randomly chosen, and (2) whether the initial particle is placed near the center or at the corner of the square lattice.

the initial particle position, to cover the entire system. In this view, initializing the particle near the center allows its light cone to reach all parts of the system more quickly than in the case where the particle starts at the corner, thereby achieving convergence at shallower depths. In addition, we confirm that fixing the two-qubit gate parameters to $\alpha_1 = \pi/8$ and $\alpha_2 = 0$ accelerates convergence, regardless of the initial particle position. This trend is consistent with findings reported by Suzuki *et al.* [42], where enhanced global randomness was observed in $U(1)$ -symmetric random circuits by fixing gate parameters.

D. Noise dependence of MLXEB

We now present the numerical results under noisy conditions for the quantum circuits introduced in Sec. IV A. As mentioned in Sec. IV A, each single-qubit and two-qubit gate is followed by depolarizing noise, and each qubit experiences depolarizing noise just before measurement. Here, we introduce the error probabilities e_g and e_q , which denote the depolarizing noise rates for gate operations and measurements, respectively. We evaluate the behavior of MLXEB in the perturbative regime, where $e_g, e_q \ll 1$. The probability that a state with particle number n transitions to another state with a particle number differing by δn due to noise is trivially of order $O(p_{\text{noise}}^{\delta n})$. Therefore, to capture the effects of noise up to order $(\delta n + 1)$, it is sufficient to retain only the subspaces with particle numbers ranging from $n_{\min} = \max[0, n - \delta n]$ to $n_{\max} = \min[n + \delta n, N]$. In this work, we always set $n_{\min} = 0$ to improve numerical accuracy. This is justified because, under the condition $n = O(1) \ll N$, the inequality $\sum_{k=0}^n D_k \ll D_{n+1}$ clearly holds, making the inclusion of lower particle-number sectors computationally inexpensive. To verify the validity of this truncation strategy, we investigated the noise dependence of the fidelity in a system with $N = 4 \times 4$ qubits and $n = 3$, where the initial particle positions are set to $(0, 0)$, $(1, 0)$, and $(2, 0)$. We compared three different truncation thresholds: $n_{\max} = 4, 5$, and 16 (i.e., full Hilbert space). The

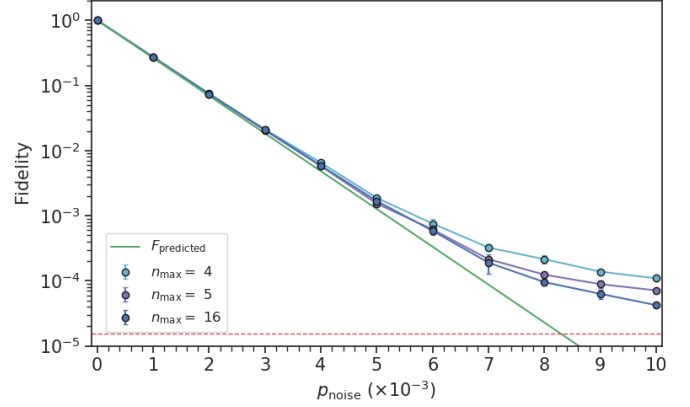


FIG. 6. Noise-rate dependence of the fidelity F for a 16-qubit system with circuit depth $N_d = 60$. The number of circuit instances and samples per instance are $(N_c, N_s) = (5, 100\,800)$. The red dashed line indicates the reference value $1/2^N$.

results are shown in Fig. 6. The two-qubit gate parameters $\alpha_1 \in [0, 2\pi)$ and $\alpha_2 \in [0, 2\pi)$ were independently and uniformly sampled at random. The green curve represents the lower bound of the predicted circuit fidelity, calculated from the noise rates of all gates and qubits using the following expression:

$$F_{\text{predicted}} = \prod_{g \in g_1} (1 - e_g) \prod_{g \in g_2} (1 - e_g) \prod_{q \in Q} \left(1 - \frac{2}{3} e_q\right), \quad (17)$$

where g_1 , g_2 , and Q denote the sets of one-qubit gates, two-qubit gates, and all qubits, respectively. Note that the depolarizing noise applied before the measurement induces a bit-flip error with probability $2e_q/3$, which affects the measurement fidelity. In the numerical experiments, for simplicity, we set all gate error rates e_g and measurement error rates e_q to the same value p_{noise} . As shown in the figure, the results of truncated simulations with $n_{\max} = 4$ and 5 closely match those from full-vector simulations ($n_{\max} = 16$) in the region $p_{\text{noise}} \leq 5.0 \times 10^{-3}$. On the other hand, for $p_{\text{noise}} \geq 5.0 \times 10^{-3}$, noticeable deviations arise between $F_{\text{predicted}}$ and the fidelities obtained from full-vector simulation. In particular, for small N , the fidelity tends to approach the lower bound $1/2^N$ (shown as the red dashed line) likely due to finite-size effects.

In the remainder of this paper, we focus on the case $n = 1$, $n_{\max} = 3$, and $N = 8 \times 8$, which provides a good balance between numerical accuracy and computational cost. Furthermore, to reduce the required circuit depth, we fix the two-qubit gate parameters to $\alpha_1 = \pi/8$ and $\alpha_2 = 0$, and place the initial particle at $(\sqrt{N}/2 - 1, \sqrt{N}/2 - 1)$ of the square lattice.

We begin by evaluating $F_{\text{MLXEB},1}$ under weak depolarizing noise with $p_{\text{noise}} = 10^{-3}$. The evaluation is performed with accuracy up to third order in p_{noise} , and the results are presented in Fig. 7. In this setup, $F_{\text{MLXEB},1}$ begins to converge to the overall circuit fidelity at circuit depths $N_d \gtrsim 20$, consistent with the noiseless simulation shown in Fig. 5. This result demonstrates that the proposed benchmarking scheme remains robust under perturbative noise.

Finally, we investigate up to what level of noise the MLXEB estimate remains consistent with the actual circuit fidelity. The results are shown in Fig. 8, where the

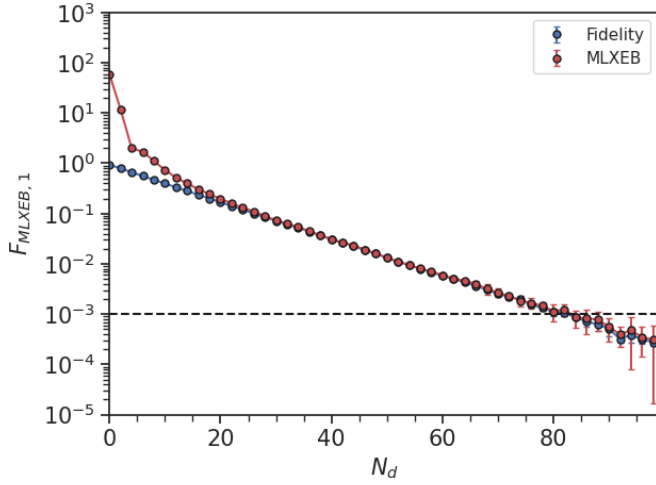


FIG. 7. Circuit depth dependence of $F_{\text{MLXEB},1}$ under weak depolarizing noise ($p_{\text{noise}} = 0.001$). The number of random-circuit instances and the number of samples per instance are $(N_c, N_s) = (10, 100\,800)$, respectively.

circuit depth is fixed at $N_d = 30$, which is sufficient to ensure convergence to the Porter-Thomas distribution. Numerical simulations indicate that for the range $p_{\text{noise}} \leq 2.0 \times 10^{-3}$, the fidelity estimated by MLXEB, $F_{\text{MLXEB},1}$, agrees well with the true circuit fidelity. In contrast, for $p_{\text{noise}} \geq 3.0 \times 10^{-3}$, a discrepancy emerges between the two. A likely reason for this deviation is the sampling error, which scales as $O(1/\sqrt{N_s N_c})$ [2]. This implies that the standard deviation remains non-negligible, and the statistical accuracy is insufficient to reliably resolve the fidelity at higher noise levels. In the low-noise regime, we also find that the predicted fidelity $F_{\text{predicted}}$ provides a good approximation of the overall circuit fidelity, validating the analytical expression used in Eq. (6).

These findings collectively confirm that MLXEB remains robust against realistic levels of depolarizing noise, accurately capturing circuit fidelity up to $p_{\text{noise}} \sim 2 \times 10^{-3}$ within the limits of statistical uncertainty.

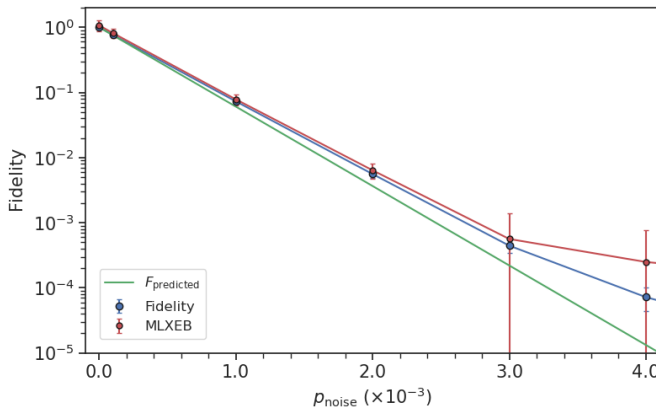


FIG. 8. Comparison of $F_{\text{MLXEB},1}$ with the true fidelity for noisy particle-number-conserving random quantum circuits. The number of circuit instances and the number of samples per instance are $(N_c, N_s) = (10, 100\,800)$, respectively.

V. CONCLUSION

In this work, we addressed the limitation that conventional linear cross-entropy benchmarking (LXEB) is not directly applied to quantum devices with more than 50 qubits due to the high computational cost of exact classical simulation. To overcome this issue, we introduced the particle number n as a conserved quantity and proposed a modified version of LXEB, referred to as MLXEB. This method employs particle-number-conserving random quantum circuits and enables estimation of the overall circuit fidelity within a fixed- n subspace. We numerically validated MLXEB for systems of up to $N = 196$ qubits, assuming a square-lattice interaction topology representative of near-term quantum devices. We considered two types of random circuits: one with randomly chosen two-qubit gate angles and another with fixed angles. In both cases, we demonstrated that with sufficient circuit depth, the output probability distribution converges to the rescaled Porter-Thomas distribution defined on the particle-number- n Hilbert subspace. For the noiseless MLXEB with randomly sampled two-qubit gate parameters, we numerically confirmed that the MLXEB value F_{MLXEB} follows the scaling relation given by Eq. (14). This result allows us to predict, in advance, the circuit depth N_d required for accurate fidelity estimation using MLXEB, even for quantum devices larger than the current scale.

On the other hand, when the two-qubit gate parameters are fixed to $\alpha_1 = \pi/8$ and $\alpha_2 = 0$, and the initial particle is placed near the center of the two-dimensional lattice, the output distribution is observed to converge more rapidly to the rescaled Porter-Thomas distribution compared to the case where the gate parameters are randomly selected and the particle is initialized at the lattice corner. This property is particularly significant when evaluating MLXEB on real quantum hardware, where the number of quantum gates available in a single execution is limited. However, under this fixed-angle condition, the scaling relation in Eq. (14) does not appear to hold consistently across the range of N considered. Identifying an appropriate scaling relation for MLXEB under fixed-angle conditions remains an open question for future research. Furthermore, it remains a theoretical open question to determine which class of unitary t -designs these particle-number-conserving random circuits with fixed two-qubit gate parameters belongs to, particularly within the context of $U(1)$ -symmetric ensembles.

In addition, for a system with $N = 64$ qubits, we evaluated MLXEB under perturbative noise using numerically validated simulations. The results confirmed that MLXEB can estimate the circuit fidelity accurately within the level of statistical uncertainty.

Future work also includes developing advanced MPI-based parallelization schemes and algorithmic improvements to support simulations of larger systems, as well as investigating circuit architectures that are compatible with experimental implementation on real quantum hardware, including constraints such as qubit connectivity, native gate sets, and gate- or qubit-dependent noise characteristics. By addressing these challenges, MLXEB is expected to become a practical benchmarking method for future quantum devices with 100–1000 qubits.

This work demonstrates that, under typical experimental noise levels ($p_{\text{noise}} \sim 10^{-3}$), MLXEB can provide reliable fidelity estimates for quantum circuits with at least 64 qubits, and potentially more, depending on the hardware characteristics.

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions and support by Ryo Watanabe. This work was partially supported by KAKENHI Grants No. JP21H04446, No. JP21H05182, and No. JP21H05191 from JSPS of Japan. We also acknowledge support from MEXT Q-LEAP Grant No. JPMXS0120319794, and from JST COI-NEXT No. JPMJPF2014, and CREST No. JPMJCR24I1 and No. JPMJCR24I3. H.U. was supported by the COE research grant in computational science from Hyogo Prefecture and Kobe City through Foundation for Computational Science

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: MLXEB THEORY: EXPECTATION VALUE AND RELATION TO STATE FIDELITY

Let $\mathcal{H}$ denote the full Hilbert space of N qubits, and let $\mathcal{H}_n$ denote the particle-number- n subspace with dimension $D_n = \binom{N}{n}$. For a fixed particle number n , the MLXEB estimator based on N_s measurement shots is

$$F_{\text{MLXEB},n} = \frac{n_s}{N_s} \left[\frac{D_n}{n_s} \left(\sum_{i=1}^{n_s} p_n(x'_i) \right) - 1 \right], \quad (\text{A1})$$

where n_s is the number of observed bit strings lying in $\mathcal{H}_n$, $\{x'_i\}_{i=1}^{n_s} \subset \mathcal{H}_n$ are those bit strings, and $p_n(\cdot)$ is the ideal (noise-free) output distribution restricted to $\mathcal{H}_n$. The goal below is to identify conditions under which the expectation of $F_{\text{MLXEB},n}$ matches the state fidelity.

To prepare for taking expectations, first rewrite Eq. (A1) so that the two sums range over the same index set:

$$F_{\text{MLXEB},n} = \frac{D_n}{N_s} \left(\sum_{i=1}^{n_s} p_n(x'_i) \right) - \frac{n_s}{N_s}, \quad (\text{A2})$$

and we introduce the indicator

$$y_j = \begin{cases} 1, & \text{if } |x_j\rangle \in \mathcal{H}_n, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A3})$$

Then, the partial sum over the n_s in-subspace outcomes can be written as a full sum masked by y_j , and the count n_s itself is the masked sum of ones:

$$\sum_{i=1}^{n_s} p_n(x'_i) = \sum_{j=1}^{N_s} y_j p_n(x_j), \quad n_s = \sum_{j=1}^{N_s} y_j. \quad (\text{A4})$$

This converts $F_{\text{MLXEB},n}$ into a linear statistic of i.i.d. samples amenable to taking expectations.

1. Setup with leakage and reduction to conditional expectations

Let U denote a particle-conserving random circuit. Acting on the pure initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$ with particle number n ,

$$|\psi_{\text{ideal},n}\rangle = U|\psi_0\rangle \in \mathcal{H}_n. \quad (\text{A5})$$

In the presence of noise, we decompose the output density operator into a part that remains inside $\mathcal{H}_n$ and a “leaked” part:

$$\rho_U = (1 - L_n)\rho_n + L_n\rho_{\text{leak}}. \quad (\text{A6})$$

Here $L_n \in [0, 1]$ is the leakage probability out of $\mathcal{H}_n$, ρ_n acts on $\mathcal{H}_n$, and ρ_{leak} has support orthogonal to $\mathcal{H}_n$.

Let $q(x) = \Pr(x) = \langle x | \rho_U | x \rangle$ be the observed sampling distribution over computational basis strings, $q_n(x) = \Pr(x | |x\rangle \in \mathcal{H}_n) = \langle x | \rho_n | x \rangle$ the conditional (no-leak) distribution restricted to $\mathcal{H}_n$, and $p_{\text{leak}}(x) = \Pr(x | |x\rangle \notin \mathcal{H}_n) = \langle x | \rho_{\text{leak}} | x \rangle$ the leaked distribution. Writing a single-shot random pair (X, Y) with $X \sim q$ and $Y = \mathbf{1}\{|X\rangle \in \mathcal{H}_n\}$, linearity of expectation gives

$$\begin{aligned} \mathbb{E}_{x \sim q(x)}[F_{\text{MLXEB},n}] &= \mathbb{E}_{x \sim q(x)} \left[\frac{D_n}{N_s} \left(\sum_{j=1}^{N_s} y_j p_n(x_j) \right) - \frac{1}{N_s} \left(\sum_{j=1}^{N_s} y_j \right) \right] \\ &= D_n \mathbb{E}_{x \sim q(x)}[Y p_n(X)] - \mathbb{E}_{x \sim q(x)}[Y]. \end{aligned}$$

Using the law of total expectation with respect to Y and the fact that $\Pr(Y=1) = 1 - L_n$, we obtain the following conditional expression:

$$\begin{aligned} \mathbb{E}_{x \sim q(x)}[Y p_n(X)] &= \Pr(Y=1) \cdot \mathbb{E}_{x \sim q(x)}[p_n(X) | Y=1] \\ &= (1 - L_n) \mathbb{E}_{x \sim q(x)}[p_n(X) | Y=1], \end{aligned} \quad (\text{A7})$$

$$\mathbb{E}_{x \sim q(x)}[Y] = 1 - L_n, \quad (\text{A8})$$

and hence

$$\begin{aligned} \mathbb{E}_{x \sim q(x)}[F_{\text{MLXEB},n}] &= (1 - L_n)[D_n \mathbb{E}_{x \sim q(x)}[p_n(X) | Y=1] - 1]. \end{aligned} \quad (\text{A9})$$

This reduces the problem to evaluating the expected value $\mathbb{E}_{x \sim q(x)}[p_n(X) | Y=1]$.

2. Depolarizing model within $\mathcal{H}_n$ and Porter-Thomas statistics

We now impose two standard assumptions:

(1) Depolarizing mixture within $\mathcal{H}_n$: Conditionally on no leakage,

$$\rho_n = F_n |\psi_{\text{ideal},n}\rangle\langle\psi_{\text{ideal},n}| + (1 - F_n) \frac{I_n}{D_n}, \quad (\text{A10})$$

with $F_n = \langle \psi_{\text{ideal},n} | \rho_n | \psi_{\text{ideal},n} \rangle$ and I_n the identity on $\mathcal{H}_n$. This implies the conditional outcome distribution

$$q_n(x) = F_n p_n(x) + (1 - F_n) \frac{1}{D_n}, \quad (\text{A11})$$

where $p_n(x) = |\langle x | \psi_{\text{ideal},n} \rangle|^2$ [and $p_n(x) = 0$ when $x \notin \mathcal{H}_n$].

(2) Porter-Thomas second moment in $\mathcal{H}_n$ [2]:

$$\sum_x p_n^2(x) \approx \frac{2}{D_n}. \quad (\text{A12})$$

Substituting Eqs. (A11) and (A12) into the conditional average gives

$$\begin{aligned} D_n \mathbb{E}_{x \sim q(x)} [p_n(X) | Y = 1] - 1 \\ &= D_n \sum_x p_n(x) q_n(x) - 1 \\ &= D_n \sum_x p_n(x) \left(F_n p_n(x) + (1 - F_n) \frac{1}{D_n} \right) - 1 \\ &= D_n F_n \sum_x p_n^2(x) + (1 - F_n) \sum_x p_n(x) - 1 \\ &= D_n F_n \times \frac{2}{D_n} + (1 - F_n) \times 1 - 1 \\ &= F_n. \end{aligned} \quad (\text{A13})$$

Therefore, under assumptions (1) and (2), the MLXEB expectation is

$$\mathbb{E}_{x \sim q(x)} [F_{\text{MLXEB},n}] = (1 - L_n) F_n, \quad (\text{A14})$$

i.e., the product of the no-leak probability and the in-subspace depolarizing weight.

3. Relation to the state fidelity

We now relate the above to the state fidelity with the ideal target in $\mathcal{H}_n$,

$$F_{\text{state}} = \langle \psi_{\text{ideal},n} | \rho_U | \psi_{\text{ideal},n} \rangle. \quad (\text{A15})$$

Using Eq. (A6) and the orthogonality of supports (ρ_{leak} has no support on $\mathcal{H}_n$), we obtain

$$F_{\text{state}} = (1 - L_n) \langle \psi_{\text{ideal},n} | \rho_n | \psi_{\text{ideal},n} \rangle. \quad (\text{A16})$$

Combining Eq. (A16) with the depolarizing form (A10) (and noting $\langle \psi_{\text{ideal},n} | I_n / D_n | \psi_{\text{ideal},n} \rangle = 1 / D_n$) shows

$$F_{\text{state}} = (1 - L_n) \left(F_n + (1 - F_n) \times \frac{1}{D_n} \right). \quad (\text{A17})$$

Thus, the deviation between the MLXEB expectation and the state fidelity is suppressed by the subspace dimension:

$$|\mathbb{E}[F_{\text{MLXEB},n}] - F_{\text{state}}| = (1 - L_n)(1 - F_n) \times \frac{1}{D_n} \leq \frac{1}{D_n}. \quad (\text{A18})$$

In words, once D_n is large, the two quantities essentially coincide.

APPENDIX B: PRACTICAL CHOICE OF THE INITIAL PARTICLE NUMBER n

In practice, the choice of n should adapt to hardware noise levels and available classical resources. There is an inherent trade-off between classical tractability and statistical/algorithmic robustness: increasing n enlarges the sector dimension $D_n = \binom{N}{n}$, which tends to accelerate in-sector

scrambling (so that Porter-Thomas (PT) distribution/2-design statistics are reached at shallower depths), but it also increases the classical cost of evaluating $p_n(x)$.

(1) High-noise regime (current NISQ devices): When gate and readout errors are on the order of 10^{-3} to 10^{-2} , meaningful MLXEB values require that circuits reach the PT regime before noise accumulation dominates. Larger n accelerates diffusion within $\mathcal{H}_n$, so PT-like statistics appear at smaller depth d . In this regime, it is therefore advisable to choose the largest n that remains classically tractable with the available resources, in order to benchmark with shallow circuits and suppress the impact of noise.

(2) Low-noise regime (concrete guidance): Let $F(d)$ denote the circuit-level fidelity at depth d (within the chosen sector), and let $\eta > 0$ be the per-layer effective decay rate so that $F(d) \approx e^{-\eta d}$ in the low-noise regime. As physical error rates η decrease, shallow circuits yield $F(d) \approx 1$, while the sampling standard deviation of the estimator scales as $c/\sqrt{N_s}$ for a modest constant c . To resolve a percycle decay rate under a finite-shot budget N_s , the departure from unity must exceed this standard deviation:

$$|1 - F(d)| \gtrsim \frac{c}{\sqrt{N_s}} \implies 1 - e^{-\eta d} \gtrsim \frac{c}{\sqrt{N_s}}. \quad (\text{B1})$$

For small ηd , using $1 - e^{-\eta d} \approx \eta d$ yields the practical criterion

$$d \gtrsim \frac{c}{\eta \sqrt{N_s}}. \quad (\text{B2})$$

We will refer to the right-hand side of Eq. (B2) as the statistical resolution depth and denote it by $d_{\text{stat}}(\eta, N_s)$ below.

Definitions used below.

(a) $d_{\text{mix}}(n)$ (PT/2-design mixing depth): The minimal depth at which the sector $\mathcal{H}_n$ exhibits Porter-Thomas/unitary 2-design statistics (typically decreasing with n);

(b) $d_{\text{stat}}(\eta, N_s)$ (statistical resolution depth): The minimal depth at which the expected decay is resolvable under N_s shots, i.e., the quantity defined by Eq. (B2);

(c) $d_{\text{sim}}(N, n, \mathcal{R}_{\text{class}})$: The largest depth for which the ideal sector distribution $p_n(x)$ is computable within the classical resource envelope $\mathcal{R}_{\text{class}}$ (time, memory, tolerance).

With these, a practically informative experiment requires

$$\boxed{\max\{d_{\text{mix}}(n), d_{\text{stat}}(\eta, N_s)\} \leq d \leq d_{\text{sim}}(N, n, \mathcal{R}_{\text{class}})}. \quad (\text{B3})$$

This interval defines the operational range (also referred to as a “window”) for (N, n) . When η is small, the lower bound $d_{\text{stat}}(\eta, N_s)$ pushes the target depth upward. To keep the ideal distribution computable at the chosen N and d , one then selects n so that D_n remains feasible under $\mathcal{R}_{\text{class}}$. In practice, once d_{stat} becomes the binding constraint, this often motivates a smaller n than in the high-noise regime, without implying that “smaller n is always better.”

(3) Adaptive strategy: Combining these observations, a natural strategy is to decrease n gradually as hardware improves. On noisier devices, one may begin with the largest classically feasible n to enable shallow-depth benchmarking, and subsequently transition to smaller n (with larger depth) as error rates decrease and shot-noise resolution becomes the dominant limitation. If ample classical resources are available,

TABLE I. Comparison of MLXEB and other benchmarking protocols.

	MLXEB [U(1) sector]	Cycle benchmarking	Standard/interleaved/ simultaneous RB	Elided-circuit LXEB
Core assumption	U(1)-symmetric random circuits form a 2-design within $\mathcal{H}_n$; leakage is explicit	Fixed Clifford(like) cycles with Pauli twirling; SPAM-robust Pauli decays	Clifford twirling over gate sets; Markovian, SPAM-robust decay model	Real-space cuts that suppress entanglement across boundaries to keep simulation tractable
What it estimates	Circuit-level fidelity $(1 - L_n)F_n$ within sector $\mathcal{H}_n$ (non-Clifford allowed)	Average (or lower-bounded) process fidelity of the chosen cycle	Average gate-set fidelity; specific-gate fidelity via interleaved RB	Effective circuit-level fidelity for a modified (elided) random-circuit family
Correlated noise	In-sector correlations averaged into F_n ; leakage and strong structure seen via L_n and PT deviations	Within-cycle correlations mapped to Pauli noise; inter-cycle structure mostly averaged	Long-range/temporal correlations largely averaged out; only coarse signatures accessible	Correlations across the cut not fully captured because boundary scrambling is weakened
Gate-set fit	Natural for native U(1)-symmetric gates (e.g., fSim-like, $CZ + R_z$)	Well matched to hardware organized into repeated parallel cycles of 1- and 2-qubit gates	Best suited to Clifford-friendly gate sets; non-Clifford requires compilation	Requires architectures admitting low-width real-space cuts with modest SWAP overhead
Scalability	Classical cost $\text{poly}(D_n)$ with $D_n = \binom{N}{n}$; feasible for $n=O(1)$ at large N	Number of Pauli decays independent of N ; depth linear in cycle count, scalable in qubit number	Number of experiments grows with gate types; deep large- N non-Clifford workloads are costly	Classical cost $\text{poly}(N) \exp[O(w)]$ with cut width w ; scalable in N when w is small

a larger n may be retained while still satisfying the depth and sampling requirements.

APPENDIX C: COMPARISON BETWEEN MLXEB AND OTHER BENCHMARKING PROTOCOLS

We position MLXEB against randomized benchmarking (RB), split here into (i) cycle benchmarking [48] and (ii) standard/interleaved/simultaneous RB [48], (iii) LXEB with elided circuits introduced in Google's work [2] on real-space subregions. We compare (1) what is estimated/assumed, (2) sensitivity to correlated/leakage-inducing errors, (3) gate-set affinity and experiment realism, (4) scalability/depth dependence, and (5) practical guidance. A concise summary of these comparisons is given in Table I.

1. What quantity is estimated (and under what assumptions)

a. MLXEB (this work)

Within a fixed particle-number sector $\mathcal{H}_n$, MLXEB estimates the circuit-level fidelity $\mathbb{E}[F_{\text{MLXEB},n}] = (1 - L_n)F_n$ (Appendix A), i.e., the product of the no-leak probability and the in-sector depolarizing weight. Under sector-wise $PT/2$ -design statistics, this matches the state fidelity up to $O(1/D_n)$. Thus, MLXEB reports a directly interpretable deep-circuit fidelity for the actual non-Clifford circuits executed [subject to a U(1) constraint].

b. Cycle benchmarking (scalable RB)

Reports a per-cycle average error parameter via a SPAM-robust exponential fit over repetitions of a fixed cycle layout. It yields an ensemble/cycle-average metric rather than the fidelity of an arbitrary deep non-Clifford workload.

c. Standard/interleaved/simultaneous RB

Standard RB returns a gate-set average fidelity under Clifford twirling; interleaved RB targets a specific gate's average fidelity; simultaneous RB probes crosstalk by running RB on multiple subsets. All are SPAM-robust averages, not direct deep-circuit fidelities.

d. Elided-circuit LXEB [2]

Reports an XEB score for a modified task in which entanglement is deliberately reduced across geometry-defined cuts of the 2D lattice; excellent for scaling studies. Consequently, it characterizes a circuit ensemble distinct from the fully entangling target.

2. Sensitivity to correlated and leakage-inducing errors

a. MLXEB

Errors that remain within $\mathcal{H}_n$ twirl to F_n under sector randomization; symmetry-breaking processes appear explicitly as a measured leakage L_n . Deviations from $PT/2$ -design statistics expose untwirled structure. Reporting $(F_{\text{MLXEB},n}, L_n)$ together separates two important error classes that are often conflated.

b. Cycle benchmarking

Cohesive/correlated errors are partly averaged by the twirl; structure is washed out unless specialized extensions are used. Leakage requires leakage-aware RB variants.

c. Standard/interleaved/simultaneous RB

Similar averaging effects; simultaneous RB can reveal crosstalk, but still does not yield a whole circuit-level fidelity.

d. Elided LXEB

Boundary or cross-partition correlations can be undercounted because the modified topology reduces precisely those correlations.

3. Gate-set affinity and experiment realism

a. MLXEB

Runs natively on $U(1)$ -symmetric entangling gates (e.g., $f\text{Sim} + R_z$ or $\text{CZ} + \text{SX} + R_z$) with shallow or no decomposition. It therefore probes the same control stack, crosstalk pathways, and calibration drifts that limit symmetry-preserving workloads (e.g., number-conserving VQE ansätze), improving external validity.

b. Cycle benchmarking

Most natural for fixed Clifford(like) cycle layouts; representative non-Clifford stress tests often require deeper compiled cycles, changing the workload being probed.

c. Standard/interleaved/simultaneous RB

Built on Clifford randomization, these methods provide SPAM-robust averaged metrics. Interleaved RB can target a specific non-Clifford gate, but RB does not directly measure the fidelity of deep non-Clifford random circuits. Emulating such circuits within RB generally entails compilations/decompositions, which increase depth and can change the failure modes relative to the native workload.

d. Elided LXEB

Require task-specific graph cuts/partitions that may not match native application circuits.

4. Scalability and depth dependence

a. MLXEB

For fixed particle number $n = O(1)$, the sector dimension is $D_n = \binom{N}{n} = O(N^n)$, i.e., it grows only polynomially in the total qubit number N . The classical cost to compute the ideal distribution $p_n(x)$ scales as $\text{poly}(D_n) = \text{poly}(N^n)$, rather than $\exp(O(N))$ as in full-Hilbert-space simulation, while the implemented circuits may be deep and non-Clifford. With a shot budget N_s and depth d (measured in MLXEB cycles), the experimental cost to estimate $F_{\text{MLXEB},n}$ scales as $O(d N_s)$, and the statistical uncertainty scales as $O(1/\sqrt{N_s})$. In practice, we fit $F_{\text{MLXEB},n}$ as a function of d to an exponential or stretched-exponential form to extract a per-cycle decay rate.

b. Cycle benchmarking

For a fixed cycle layout, the number of free parameters (cycle fidelities) does not grow with N , and the circuit depth scales linearly with the number of cycles L . The experimental cost is $O(L N_s)$ for N_s shots per length- L sequence, and classical post-processing is $\text{poly}(N)$ (dominated by simple fits and bookkeeping). The reported quantity, however, remains an ensemble/ cycle average and does not directly yield the fidelity of an arbitrary deep non-Clifford workload with a specific structure.

c. Standard/interleaved/simultaneous RB

In standard and interleaved RB, one typically runs a separate RB experiment per gate (or per small gate set), so the number of required experiments scales at least linearly with the number of target gate types M . For each gate type, the sequence length L grows linearly with the number of RB steps, and when non-Clifford resources are present, compilation into Clifford gates and native entangling gates can further increase L by a factor that depends on the desired accuracy. Thus, the total experimental cost scales on the order of $O(M L N_s)$, with modest classical postprocessing per experiment. This makes it less practical to scale these RB variants to very large N and deep non-Clifford random circuits when the goal is to benchmark specific circuit-level workloads rather than averaged gate fidelities.

d. Elided LXEB

If an elided circuit admits a real-space cut of width w in the (typically two-dimensional) device geometry, the classical simulation cost of the ideal distribution scales as $\text{poly}(N) \exp[O(w)]$ rather than $\exp[O(N)]$. For a fixed $w = O(1)$ and depth d , the experimental cost is again $O(d N_s)$. By construction, however, the benchmarked task has reduced entanglement across the spatial cuts and therefore differs from the fully entangling target circuit.

5. When to use what (practical guidance)

For SPAM-robust routine performance monitoring of a device or device fleet, *cycle benchmarking* is well suited. For gate-specific certification, *interleaved RB* is the appropriate tool. For diagnosing and quantifying crosstalk, *simultaneous RB* is informative. For scaling demonstrations under a modified but classically tractable workload, *elided-circuit LXEB* is effective. When one needs a deep, large- N , non-Clifford, circuit-level fidelity under a physically relevant symmetry, and wishes to distinguish in-sector noise from symmetry-breaking leakage, *MLXEB* is appropriate and complementary.

6. Fairness and validation

We do not claim MLXEB is universally “better.” It targets a different point in the trade-off space: scalable, native, non-Clifford, circuit-level fidelity under a conserved-quantity constraint. In small-to-moderate instances where all metrics are feasible, we recommend cross-checks: (1) compare

cycle-RB per-cycle rates with MLXEB depth-attenuation; (2) verify stability of $F_{\text{MLXEB},n}$ under layer-order shuffles/phase

randomization and across n ; and (3) confirm that L_n is small when symmetry-breaking errors are weak.

-
- [1] A. M. Dalzell, S. McArdle, M. Berta, P. Bienias, C.-F. Chen, A. Gilyén, C. T. Hann, M. J. Kastoryano, E. T. Khabiboulline, A. Kubica, *et al.*, *Quantum Algorithms: A Survey of Applications and End-to-End Complexities* (Cambridge University Press, Cambridge, 2025).
- [2] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [3] A. Morvan, B. Villalonga, X. Mi, S. Mandra, A. Bengtsson, P. Klimov, Z. Chen, S. Hong, C. Erickson, I. Drozdov, *et al.*, Phase transitions in random circuit sampling, *Nature (London)* **634**, 328 (2024).
- [4] D. Gao, D. Fan, C. Zha, J. Bei, G. Cai, J. Cai, S. Cao, F. Chen, J. Chen, K. Chen, *et al.*, Establishing a new benchmark in quantum computational advantage with 105-qubit Zuchongzhi 3.0 processor, *Phys. Rev. Lett.* **134**, 090601 (2025).
- [5] M. DeCross, R. Haghsheenas, M. Liu, E. Rinaldi, J. Gray, Y. Alexeev, C. H. Baldwin, J. P. Bartolotta, M. Bohn, and E. A. Chertkov, The computational power of random quantum circuits in arbitrary geometries, *Phys. Rev. X* **15**, 021052 (2025).
- [6] R. Acharya, D. A. Abanin, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, N. Astrakhantsev, *et al.*, Google Quantum AI and Collaborators. Quantum error correction below the surface code threshold, *Nature (London)* **638**, 920 (2025).
- [7] A. Katabarwa, K. Gratsea, A. Caesura, and P. D. Johnson, Early fault-tolerant quantum computing, *PRX Quantum* **5**, 020101 (2024).
- [8] M. AbuGhanem, IBM quantum computers: Evolution, performance, and future directions, *J. Supercomput.* **81**, 687 (2025).
- [9] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [10] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nat. Rev. Phys.* **2**, 382 (2020).
- [11] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
- [12] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, *J. Math. Phys.* **43**, 4452 (2002).
- [13] Z. Zhou, A. Ji, and Y. Ding, Surface code error correction with crosstalk noise, *arXiv:2503.04642*.
- [14] J. F. Kam, S. Gicev, K. Modi, A. Southwell, and M. Usman, Detrimental non-Markovian errors for surface code memory, *Quantum Sci. Technol.* **10**, 035060 (2025).
- [15] T. Proctor, K. Young, A. D. Baczewski, and R. Blume-Kohout, Benchmarking quantum computers, *Nat. Rev. Phys.* **7**, 105 (2025).
- [16] 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).
- [17] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, *Nat. Phys.* **16**, 1184 (2020).
- [18] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, *Phys. Rev. A* **85**, 042311 (2012).
- [19] 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).
- [20] 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).
- [21] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [22] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, *ACM Trans. Quantum Comput.* **1**, 1 (2020).
- [23] R. Harper, W. Yu, and S. T. Flammia, Fast estimation of sparse quantum noise, *PRX Quantum* **2**, 010322 (2021).
- [24] S. T. Flammia and R. O'Donnell, Pauli error estimation via population recovery, *Quantum* **5**, 549 (2021).
- [25] S. T. Flammia, Averaged circuit eigenvalue sampling, in *17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)*, *Leibniz International Proceedings in Informatics (LIPIcs)* (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2022), Vol. 232, pp. 4:1–4:10.
- [26] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, *Nat. Phys.* **14**, 595 (2018).
- [27] S. Mullane, Sampling random quantum circuits: A pedestrian's guide, *arXiv:2007.07872*.
- [28] Y. Liu, M. Otten, R. Bassirianjahromi, L. Jiang, and B. Fefferman, Benchmarking near-term quantum computers via random circuit sampling, *arXiv:2105.05232*.
- [29] J. Chen, D. Ding, C. Huang, and L. Kong, Linear cross-entropy benchmarking with Clifford circuits, *Phys. Rev. A* **108**, 052613 (2023).
- [30] B. Cheng, F. Meng, Z.-J. Zhang, and M.-H. Yung, Generalized cross-entropy benchmarking for random circuits with ergodicity, *Chip* **4**, 100127 (2025).
- [31] D. Lacroix, E. A. R. Guzman, and P. Siwach, Symmetry breaking/symmetry preserving circuits and symmetry restoration on quantum computers: A quantum many-body perspective, *Eur. Phys. J. A* **59**, 3 (2023).
- [32] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, *et al.*, Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions, *Phys. Rev. A* **98**, 022322 (2018).
- [33] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, Exploiting symmetry in

- variational quantum machine learning, *PRX Quantum* **4**, 010328 (2023).
- [34] J. M. Arrazola, O. Di Matteo, N. Quesada, S. Jahangiri, A. Delgado, and N. Killoran, Universal quantum circuits for quantum chemistry, *Quantum* **6**, 742 (2022).
- [35] B. M. Ayeni, A bottom-up approach to constructing symmetric variational quantum circuits, [arXiv:2308.08912](https://arxiv.org/abs/2308.08912).
- [36] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm, *npj Quantum Inf.* **6**, 10 (2020).
- [37] A. Summer, A. Nico-Katz, S. Dooley, and J. Goold, Anomalous transport in $U(1)$ -symmetric quantum circuits, *npj Quantum Inf.* (2026).
- [38] H. Liu, A. Hulse, and I. Marvian, Unitary designs from random symmetric quantum circuits, [arXiv:2408.14463](https://arxiv.org/abs/2408.14463).
- [39] Y. Mitsuhashi, R. Suzuki, T. Soejima, and N. Yoshioka, Unitary designs of symmetric local random circuits, *Phys. Rev. Lett.* **134**, 180404 (2025).
- [40] N. Khaneja, R. Brockett, and S. J. Glaser, Time optimal control in spin systems, *Phys. Rev. A* **63**, 032308 (2001).
- [41] B. Kraus and J. I. Cirac, Optimal creation of entanglement using a two-qubit gate, *Phys. Rev. A* **63**, 062309 (2001).
- [42] R. Suzuki, H. Katsura, Y. Mitsuhashi, T. Soejima, J. Eisert, and N. Yoshioka, More global randomness from less random local gates, [arXiv:2410.24127](https://arxiv.org/abs/2410.24127).
- [43] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, *Commun. Math. Phys.* **291**, 257 (2009).
- [44] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2010).
- [45] K. Mølmer, Y. Castin, and J. Dalibard, Monte Carlo wave-function method in quantum optics, *J. Opt. Soc. Am. B* **10**, 524 (1993).
- [46] H. Ueda, S. Yunoki, and T. Shimokawa, Quantum spin solver near saturation: QS3, *Comput. Phys. Commun.* **277**, 108369 (2022).
- [47] H. Ueda, D. Yamamoto, and T. Shimokawa, QS³ for Various Quantum Spin Systems, GitHub repository: QS-Cube/ED2, GitHub repository (2023), <https://github.com/QS-Cube/ED2>.
- [48] A. Hashim, L. B. Nguyen, N. Goss, B. Marinelli, R. K. Naik, T. Chistolini, J. Hines, J. Marceaux, Y. Kim, P. Gokhale, *et al.*, Practical introduction to benchmarking and characterization of quantum computers, *PRX Quantum* **6**, 030202 (2025).