Effect of Nonunital Noise on Random-Circuit Sampling

Bill Fefferman, [1,*] Soumik Ghosh, [1,†] Michael Gullans, [2,‡] Kohdai Kuroiwa, [3,4,§] and Kunal Sharma [5,¶]

[1] *Department of Computer Science, University of Chicago, Chicago, Illinois 60637, USA*
[2] *Joint Center for Quantum Information and Computer Science and Joint Quantum Institute, University of Maryland and National Institute of Standards and Technology (NIST), College Park, Maryland 20742, USA*
[3] *Institute for Quantum Computing, University of Waterloo, Ontario N2L 3G1, Canada*
[4] *Perimeter Institute for Theoretical Physics, Ontario, N2L 2Y5, Canada*
[5] *IBM Quantum, IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA*

(Received 8 March 2024; revised 9 June 2024; accepted 17 June 2024; published 26 July 2024)

In this work, drawing inspiration from the type of noise present in real hardware, we study the output distribution of random quantum circuits under practical nonunital noise sources with constant noise rates. We show that even in the presence of unital sources such as the depolarizing channel, the distribution, under the combined noise channel, never resembles a maximally entropic distribution at any depth. To show this, we prove that the output distribution of such circuits never anticoncentrates—meaning that it is never too "flat"—regardless of the depth of the circuit. This is in stark contrast to the behavior of noiseless random quantum circuits or those with only unital noise, both of which anticoncentrate at sufficiently large depths. As a consequence, our results shows that the complexity of random-circuit sampling under realistic noise is still an open question, since anticoncentration is a critical property exploited by both state-of-the-art classical hardness and easiness results.

DOI: 10.1103/PRXQuantum.5.030317

I. INTRODUCTION

The defining feature of quantum systems today is noise [1]. A fundamental question in this era of noisy intermediate-scale quantum (NISQ) computers is whether noise renders any demonstration of quantum advantage with these systems useless or whether some advantage is still salvageable for specific tasks [2,3]. To study questions related to quantum advantage, a popular paradigm is the random quantum circuit model (see, e.g., Refs. [3–6]). This is because for large system sizes, sampling from the output distribution of these circuits is a task that is easy for quantum computers but provably hard for classical computers, under plausible complexity-theoretic assumptions [4,7]. Amongst other phenomena, these circuits can model quantum chaos [8], quantum pseudorandomness [9], and the ansatz for certain types of variational quantum algorithms used in optimization tasks [10,11]. An understanding of the behavior of these circuits under physically motivated noise models and limited system sizes is crucial to our understanding of quantum advantage.

One central feature of the output distribution, found in random quantum circuits of sufficiently high depth, is anticoncentration: it is a "flatness property" or, more formally, it means that the distribution is not concentrated on a sufficiently small number of outcomes. Anticoncentration is believed to be a key ingredient in both easiness and hardness proofs of random-circuit sampling (see, e.g., Refs. [6,11,12]). Importantly, anticoncentration is necessary for the final state of the system to have an output distribution that mimics the uniform distribution. If the system were to have that property, then the system would be simulable, because sampling from the uniform distribution is classically easy. This naturally prompts the following question:

Do random quantum circuits, under the influence of physically motivated noise models, anticoncentrate?

In this work, we answer this question in the negative: we show how random quantum circuits, under noise models inspired by real hardware, which is a mixture of both unital and nonunital noise of certain types, exhibit a lack of anticoncentration, regardless of the depth of the circuit. This shows that the output distribution does not

*Contact author: wjf@uchicago.edu
†Contact author: soumikghosh@uchicago.edu
‡Contact author: mgullans@umd.edu
§Contact author: kkuroiwa@uwaterloo.ca
¶Contact author: kunals@ibm.com

resemble the uniform distribution or close variants of the same. Interestingly, we find that this behavior is exhibited *even when* there are also unital noise sources, such as the depolarizing noise present in the system, the property of which is to push the system toward the maximally mixed state—which, when measured in the standard basis, gives the uniform distribution. In this sense, these nonunital sources "dominate" the other sources.

Hence, our work emphasizes the importance and necessity of investigations on whether there is a quantum signal in these distributions and sampling from them is indeed classically hard or whether the system tends toward a more sophisticated classically simulable final state. Much is unknown about random-circuit sampling under such realistic noise models and our work provides one of the first rigorous theoretical analyses of this regime.

II. OVERVIEW OF MAIN RESULTS AND CONSEQUENCES

In this section, we give a brief overview of consequences of this work. In Sec. II A, we provide a broad summary of our results. In Sec. II B, we review the previous approaches to put our results in context. In Sec. II C, we discuss implications of our results in the context developed in Sec. II B.

A. Summary of main results

In this work, we prove the lack of anticoncentration with respect to either of the two popular definitions: a strong definition of anticoncentration with respect to the convergence of scaled collision probabilities and a weak definition of anticoncentration with respect to high probability mass of typical probabilities. The definitions and connections between them are made explicit in Sec. IV.

Note that the terms "strong" and "weak" definition are relics from studying these definitions with respect to noiseless random circuits or random circuits with depolarizing noise, where the strong definition implies the weak definition. However, this is not the case for the nonunital noise channels that we consider, as we elaborate in Sec. IV C. Hence, our proofs of lack of anticoncentration, with respect to these two definitions, are independent. Nonetheless, we stick with the existing nomenclature to refer to these definitions succinctly in different parts of the paper.

Broadly, we prove the following three categories of results.

(1) First, in Sec. VI, we show how the scaled collision probabilities for our noisy ensembles, where noise is modeled as a mixture of amplitude-damping and depolarizing noise, diverge: this means that anticoncentration fails with respect to the strong definition. This has a clean proof, which involves "removing" the last layer of noise by using the adjoint of the noise channel and then using properties of the local Haar measure, such as translational invariance and explicit formulas for second moments [13], to prove a lower bound.

(2) Then, we show how typical output probabilities for strings that have high Hamming weights are small. This is done in Secs. VII, VIII, and IX. This shows that anticoncentration fails with respect to the weak definition. This has a more complicated proof, involving light-cone arguments and a statistical model.

(3) Finally, we discuss how extensions of our proofs hold for a wide variety of generic noise models: this is done in Secs. X and XI. In particular, we prove how lack of anticoncentration, with respect to the strong definition, is exhibited whenever the noise channel, acting on the identity operator, puts nonzero constant weight on the Pauli-Z operator.

B. Prior approaches and results

Here, we review previous results on random-circuit sampling. Depending on how strong the noise is, we can divide setups of random-circuit sampling into two different complexity-theoretic regimes.

1. High-noise regime

The first regime is that of *high noise*, when the noise rate is a constant that is independent of the system size, even when the number of qubits grows asymptotically. Near-term devices are susceptible to constant noise rates [14]. It is an equally reasonable model for scaled-up fault-tolerant systems, because to achieve fault tolerance, suppressing the noise below a certain constant threshold suffices—one does not need noise to go down with system size [15–17].

If we model the noise as only depolarizing noise in the high-noise regime, then after sufficient depth, random-circuit sampling becomes an easy task classically. It is known that a trivial classical algorithm that just samples from the uniform distribution achieves a total variation distance error, which exponentially decays with the circuit depth, from the target noisy random-circuit distribution. The upper bound is due to Aharonov and Ben-Or [18] and, more recently, the lower bound has been found by Deshpande *et al.* [6]. As is evident, for depth strictly greater than logarithmic, this trivial sampler achieves a total variation distance error that is smaller than any inverse polynomial.

At logarithmic depth, in a recent work, Aharonov *et al.* [19] have proposed another classical sampler from the output distribution of random circuits with depolarizing noise, which instead of achieving a total variation distance error that is inversely proportional to a fixed polynomial, provides a way to fine tune the total variation distance error to *any* polynomial function of our choice. In Appendix A, it is elaborated how, because of the special property of depolarizing noise, the guarantee that the noiseless ensemble of Ref. [19] satisfies anticoncentration is necessary for the authors' current analysis to work. This is why their sampler works well only for logarithmic depth and beyond, because anticoncentration needs at least logarithmic depth to kick in, and before that, anticoncentration fails [6,20].

Although depolarizing noise, along with anticoncentration, implies a classical sampler from random circuits, such noise is not the only type of noise source present in real hardware. There are nonunital effects in all known experimental hardware, e.g., those in Refs. [21–25]. These sources are fundamentally different from unital sources, such as the depolarizing channel, in the following sense: the depolarizing channel *increases* the entropy of the system by pushing it toward the maximally mixed state; however, nonunital noise channels can *decrease* the entropy of the system and actually push it toward a pure state. For the low-noise regime, there is some evidence that depolarizing noise remains a good approximation to all the noise sources present in the system [20], if the system only has unital noise. However, apart from some very special cases that we discuss in Sec. B, this approximation is not valid in the high-noise regime, especially when the system also has nonunital noise sources.

2. Low-noise regime

The second regime is that of *low noise*. Here, the strength of the noise is inversely proportional to the number of qubits. This can be thought to be a good approximation of the noise present in relatively small fixed-sized systems, such as those used in sampling hardness demonstrations (see, e.g., Refs. [22,26]), where the number of qubits is fixed, and the noise is a fixed constant that is inversely related to the number of qubits. However, without further investigation, it is unclear if asymptotic analysis of the low-noise regime is relevant to studying the properties of finite system sizes: we cannot ensure by current technology that the noise rate continues to go down with the number of qubits when the latter is increased.

The low-noise regime provides advantages to tasks such as benchmarking, where the fidelity of the output state is a figure of merit: below a certain noise threshold, the linear cross-entropy test (XEB) [22,27,28] corresponds to the fidelity of the output state of the noisy circuit and gives us a sample efficient way of estimating the fidelity of that state using only standard basis measurement. Originally, linear cross-entropy was proposed as a heuristic proxy for fidelity in Refs. [4,11,29]. Later, in Refs. [26,30,31], it has been observed empirically that the XEB score is a good proxy to circuit fidelity in a substantially low-noise regime, for both unital and amplitude-damping noise. This sharply contrasts with the high-noise regime, where XEB ceases to be a good proxy for fidelity and is actually classically spoofable for both these noise models, as found in Ref. [30].

C. Implications of our results

Our results show that the phenomenon of anticoncentration is a function of the noise present in the circuit: it *does not* hold for reasonable physically motivated nonunital noise models in the high-noise regime. Many of our techniques can be generalized to work for a wide variety of noise models and setups, including when we do not have the last layer of noise and only have noise in the middle layers, which we discuss in Sec. XI.

The failure to anticoncentrate implies that the final state of the system does not resemble a maximally mixed state and the uniform distribution is not a good proxy for the output distribution of the system, as elaborated on in Appendix C. So, sampling from the uniform distribution, or close variants of the same, no longer works as an effective strategy to classically spoof the output distribution. Additionally, the failure to anticoncentrate also implies that no known techniques can be harnessed to show that more sophisticated samplers, such as the one in Ref. [19], succeed in spoofing the output distribution.

Thus, our work leaves open whether sampling from this distribution, in the asymptotic limit, is classically hard under plausible complexity-theoretic hardness conjectures or whether the final state converges to a classically simulable fixed point that is much more sophisticated than just a maximally mixed state. It could also be interesting to investigate whether the lack of anticoncentration implies any advantage in the computation of the expectation value of certain cost functions in variational setups. We discuss many more open problems in Sec. XII.

D. Other recent work on nonunital noise

Finally, we mention a few recent complementary works on nonunital noise in NISQ devices.

1. Lack of noise-induced barren plateau under nonunital noise

For a circuit ensemble, if the output distribution anticoncentrates or is primarily “flat,” it renders the ensemble useless for any gradient-descent-based optimization tasks, because the optimization subroutine runs into the barren-plateau problem [32–34]. More concretely, the gradient of the cost function of the optimization task vanishes in the “flat” landscape and the optimization gets stuck. On the other hand, if a distribution lacks anticoncentration, it could potentially escape this phenomenon because the second moment of a distribution diverges with respect to the number of qubits and convergence of the second moment is necessary for barren plateaus for certain optimization setups, as has been rigorously shown in Ref. [34]. Recently, in Ref. [35], it has indeed been shown that under nonunital noise, the "barren-plateau" phenomenon can be escaped.

2. Classical estimation of expectation value of observable under nonunital noise

Even though anticoncentration is a key ingredient of existing easiness of sampling results from random quantum circuits, easiness of computing the expectation value of certain observables may not require anticoncentration. In Ref. [36], a polynomial time classical estimator has been proposed to compute such expectation values for random quantum circuits with only depolarizing noise, by exploiting a special property of the noise—the fact that only polynomially many Pauli paths have nontrival path weights. This is sketched in Appendix D and this technique does not require anticoncentration. Moreover, in Ref. [35], a classical estimator has recently been proposed that works for a noisy random circuit with nonunital noise and arbitrary depth.

The proposed algorithm has a polynomial run time for one-dimensional (1D) architecture and a quasipolynomial run time for higher-dimensional architectures, for computing expectation values up to inverse-polynomial additive precision. However, very crucially, the run time of this algorithm is inversely related to the desired additive precision. So, this algorithm only works for classically computing expectation values and does not shed light on the complexity of sampling from random circuits with nonunital noise. For sampling, we usually simulate certain quantities that can be written as expectation values—e.g., output probabilities and marginals—to inverse exponential additive precision. However, since the run time of this algorithm is inversely proportional to the additive precision of the simulation, this results in an exponentially worse running time if this algorithm is naively used to sample. Note that it still remains open if a cleverer algorithm does any better or if classically sampling from these circuits is indeed complexity-theoretically hard, as we will also mention in Sec. XII A.

III. PRELIMINARIES

In this section, we provide background materials for this work.

A. Notation and general definitions

In this section, we introduce the notation used in this paper. Throughout this paper, we use I_N to denote the $N \times N$ identity matrix. For example, I_2 represents the single-qubit identity operator.

The single-qubit Pauli operators are defined as follows:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1}$$

For a string $p \in \{0,1,2,3\}^n$, ω_p is defined as the Hamming weight of p, i.e., the number of nonzero symbols in p. For a string $p = p_1 p_2 \ldots p_n$ with $p_k \in \{0,1,2,3\}$ for each k, we define

$$\sigma_p := \bigotimes_{k=1}^{n} \sigma_{p_k}, \tag{2}$$

where we define $\sigma_0 = I_2$, $\sigma_1 = \sigma_x$, $\sigma_2 = \sigma_y$, and $\sigma_3 = \sigma_z$.

A quantum channel is a linear map of operators that is completely positive and trace preserving (CPTP). Any quantum channel $\mathcal{N}$ has Kraus operators $\{N_i\}_i$, with which we may write

$$\mathcal{N}(X) = \sum_i N_i X N_i^\dagger \tag{3}$$

for any input X. We use $\mathcal{I}$ to denote the single-qubit identity channel. A quantum channel $\mathcal{N}$ is a unital channel if $\mathcal{N}(I) = I$, where I denotes the identity operator.

Regarding the asymptotic notation, for $x \in \mathbb{R}$ and functions $f(x)$ and $g(x)$, we write

$$f(x) = \mathcal{O}(g(x)) \tag{4}$$

if $\limsup_{x\to\infty} |f(x)|/g(x) < \infty$. Also, we write

$$f(x) = \omega(g(x)) \tag{5}$$

if $\lim_{x\to\infty} f(x)/g(N) > \infty$.

B. Circuit architectures

Here, we define the circuit architecture that we use in the rest of the paper. We assume familiarity with basic terminologies in quantum computing, such as qubits, quantum circuits, and circuit depth. Unless otherwise stated, a quantum circuit, usually denoted by $\mathcal{C}$, is taken to be a CPTP map and the final measurements, after applying a quantum circuit, are always performed in the standard basis.

Definition 1 (Parallel quantum circuit). An n-qubit *parallel* quantum circuit is a quantum circuit in which every qubit is involved in a one- or two-qubit gate, at every depth instance.

Definition 2 (Geometrically local quantum circuit). An n-qubit *geometrically local* quantum circuit is a quantum circuit in which every quantum gate acts on nearest-neighbor qubits.

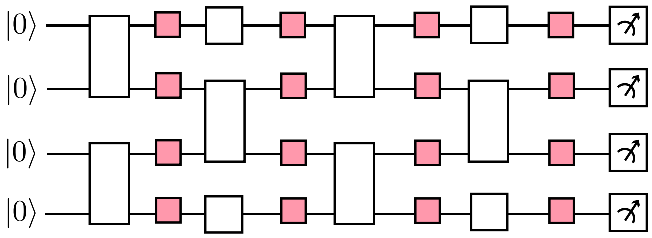


FIG. 1. The circuit model that we use for our analysis. Every pink rectangle is a single-qubit noise channel. Every white rectangle is either a single-qubit or a two-qubit Haar-random gate. In the end, the final state is measured in the standard basis.

Definition 3 (Noisy quantum circuit). An n-qubit *noisy* quantum circuit is one in which every quantum gate is followed by a single-qubit noise channel $\mathcal{N}$ on every qubit involved in the gate.

Definition 4 (Random quantum circuit). An n-qubit *random* quantum circuit is one in which every quantum gate, acting on k qubits, for a constant k, is drawn from the Haar measure on $\mathsf{U}(2^k)$: this is the set of all unitary matrices, of dimension $2^k \times 2^k$.

Alternatively, a random quantum circuit can be interpreted as a quantum circuit picked uniformly at random from an *ensemble* of quantum circuits. In this paper, unless otherwise stated, we consider parallel, geometrically local, noisy, and random circuits, as depicted in Fig. 1. We usually use $\mathcal{B}$ to denote an ensemble of noisy random circuits. In this work, we often compute the expectation value of a function f of noisy random circuits $\mathcal{C}$ from an ensemble $\mathcal{B}$, which is denoted by $\mathop{\mathbb{E}}_{\mathcal{B}}[f(\mathcal{C})]$. For example, we consider the output probability distribution or the expectation value of some observable as a function f.

C. Noise model

Quantum devices are affected by two sources of noise: unital and nonunital quantum noise channels. Dephasing, bit-flip, and depolarizing noise channels are examples of unital quantum channels. On the other hand, the amplitude-damping channel is an example of a nonunital quantum channel, which models the T_1 noise in superconducting quantum devices (see also Ref. [37] for these examples of noise). Unital and nonunital noise sources have the opposite behavior. While amplitude-damping noise "biases" the system toward a particular fixed state ($|0\rangle$ state), unital sources push the system toward the maximally mixed state.

Note that in this paper, we will often consider properties of noisy random quantum circuits *in expectation*: unless otherwise stated, the expectation will always be taken only over the choice of random gates and *not* over the noise channels.

In the next few sections, we will use a combination of the depolarizing channel and the amplitude-damping channel to model the noise after each single-qubit gate. Amplitude-damping noise is emblematic of the T_1 noise [25,38] and the depolarizing channel is emblematic of the type of unital noise just described [19,22]. In later sections, we discuss how our analysis and techniques are general enough to apply to a wide variety of noise channels.

1. Amplitude-damping noise

This type of noise pushes a qubit toward the state $|0\rangle\langle 0|$. It is represented by two Kraus operators, as given by Definition 5. The first Kraus operator "dampens" the $|1\rangle\langle 1|$ term and the second Kraus operator takes the state $|1\rangle\langle 1|$ to $|0\rangle\langle 0|$ state, with a prefactor. Both of the operators serve to make the contribution of $|0\rangle\langle 0|$ dominate in the final state that we obtain after the channel is applied.

Definition 5 (Amplitude-damping noise channel). Let $0 \le q \le 1$ be a real parameter. A single-qubit amplitude-damping noise $\mathcal{N}_q^{(\mathrm{amp})}$ with noise strength q is a quantum channel with the following Kraus operators:

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-q} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{q} \\ 0 & 0 \end{pmatrix}. \tag{6}$$

Therefore, for a single-qubit linear operator

$$X = \begin{pmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \end{pmatrix}, \tag{7}$$

the action of the amplitude-damping channel, $\mathcal{N}_q^{(\mathrm{amp})}(X)$, is given by

$$\mathcal{N}_q^{(\mathrm{amp})}(X) = \begin{pmatrix} x_{00} + q x_{11} & \sqrt{1-q}\,x_{01} \\ \sqrt{1-q}\,x_{10} & (1-q)x_{11} \end{pmatrix}. \tag{8}$$

2. Depolarizing noise

Definition 6 (Depolarizing noise channel). Let $0 \le p \le 1$ be a real parameter. A single-qubit depolarizing noise $\mathcal{N}_p^{(\mathrm{dep})}$ with noise strength p is a quantum channel with the following Kraus operators:

$$K_0 = \sqrt{1 - \frac{3p}{4}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad K_1 = \sqrt{\frac{p}{4}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$K_2 = \sqrt{\frac{p}{4}} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad K_3 = \sqrt{\frac{p}{4}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{9}$$

Therefore, for any single-qubit linear operator X,

$$X = \begin{pmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \end{pmatrix}, \tag{10}$$

the action of the depolarizing channel, $\mathcal{N}_p^{(\mathrm{dep})}(X)$, is given by

$$\mathcal{N}_p^{(\mathrm{dep})}(X) = \begin{pmatrix} (1-p)\,x_{00} + \frac{p}{2}x_{11} & (1-p)x_{01} \\ (1-p)x_{10} & (1-p)\,x_{11} + \frac{p}{2}x_{00} \end{pmatrix} \tag{11}$$

$$= (1-p)X + \frac{p}{2}\mathrm{Tr}(X)I_2, \tag{12}$$

with the single-qubit identity operator I_2.

The notion that the amplitude-damping channel is fundamentally different from the depolarizing channel is elaborated on in works such as Ref. [39], where it is shown how quantum circuits with an uncorrected amplitude-damping channel can be used to do exponential-time quantum computation, in the worst case, by using the noise as a resource to generate fresh ancilla qubits. However, this cannot be done for quantum circuits with the depolarizing noise channel.

IV. ANTICONCENTRATION AND THE LACK THEREOF

In this section, we introduce the notion of anticoncentration and discuss what it means to *not* anticoncentrate. First, we will use the definition of anticoncentration in Ref. [20], where it is defined with respect to the collision probability. Then, we will talk about another definition of anticoncentration, which is found in, e.g., Refs. [6,7,40], and discuss connections between the two definitions.

A. Strong definition: With respect to the scaled collision probability

Definition 7. The *output probability* p_x, of a string $x \in \{0,1\}^n$, for a quantum circuit $\mathcal{C}$ is given by

$$p_x = \mathrm{Tr}\big(|x\rangle\langle x|\, \mathcal{C}(|0^n\rangle\langle 0^n|)\big). \tag{13}$$

Definition 8. For an ensemble $\mathcal{B}$, the scaled collision probability is defined as

$$\mathcal{Z} = 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} p_x^2\right] - 1. \tag{14}$$

In other words,

$$\mathcal{Z} = 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \mathrm{Tr}(|x\rangle\langle x|\, \mathcal{C}(|0\rangle\langle 0|))^2\right] - 1. \tag{15}$$

Definition 9. An ensemble $\mathcal{B}$ of n-qubit quantum circuits is defined to be *anticoncentrated* if

$$\mathcal{Z} = \mathcal{O}(1). \tag{16}$$

Intuitively, Definition 9 says that the probability of seeing a collision, after sampling twice from the output distribution of a circuit in $\mathcal{B}$, is extremely small in expectation. In other words, "most" n-bit strings have sufficiently high probability weight in the output distribution. Definition 9 is a "stronger" definition of anticoncentration because, for certain types of circuits, it implies another well-studied "weaker" definition of anticoncentration, as we will see in Sec. IV B.

Definition 10. An ensemble $\mathcal{B}$ of n-qubit quantum circuits is said to exhibit *lack of anticoncentration* if

$$\mathcal{Z} = \omega(1). \tag{17}$$

The right-hand side means that the quantity is strictly more than a constant: it grows as some function of n. Intuitively, a distribution that satisfies Definition 10 is not "evenly" supported on all n-bit strings: some strings have much more probability weight than others. So, there is a much larger probability of seeing a collision when sampling multiple times from this distribution.

Remark 1. The connection of Definition 9 to easiness proofs is elaborated on in Appendix A. In short, the guarantee that ensembles satisfy Definition 9 is necessary for the current techniques to analyze the accuracy of the classical sampler, proposed by Aharonov *et al.* [19], to work. This is why the sampler only works after logarithmic depth; logarithmic depth is the threshold at which these ensembles are known to start anticoncentrating [20].

B. Weak definition: With respect to individual probabilities

There is another definition of anticoncentration, which defines it in terms of how large individual probabilities are. This can be found in works such as Refs. [4,6,7,41].

Definition 11. An ensemble $\mathcal{B}$ of n-qubit quantum circuits is defined to be *anticoncentrated* if, for every $x \in \{0,1\}^n$, there exists a choice of $\alpha, \beta \in (0,1]$ such that

$$\Pr_{\mathcal{B}}\left[p_x \geq \frac{\alpha}{2^n}\right] \geq \beta. \tag{18}$$

We define the lack of anticoncentration in a similar way.

Definition 12. An ensemble $\mathcal{B}$ of n-qubit quantum circuits is defined to *lack anticoncentration* if there exists an $x \in \{0,1\}^n$, such that, for any $\alpha \in (0,1]$,

$$\lim_{n\to\infty} \Pr_{\mathcal{B}}\left[p_x < \frac{\alpha}{2^n}\right] = 1. \tag{19}$$

The form of anticoncentration in Sec. 11 is believed to be useful in proving that the output distribution of some random-circuit ensembles are hard to classically sample from [4,7]. Section 11 is called the "weaker definition" of anticoncentration because Definition 9 implies Sec. 11, as we will now establish.

C. Connection between two definitions

Before establishing the connection between Definition 9 and Sec. 11, let us state a version of the "hiding" property.

Definition 13. Let $\mathcal{B}$ be an ensemble of random quantum circuits. Then, $\mathcal{B}$ is said to satisfy hiding if

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[p_x^k\right] = \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_y^k\right] \tag{20}$$

for $x, y \in \{0,1\}^n$, $x \neq y$, and for any k.

Note that for ensembles that satisfy hiding,

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} p_x^2\right] = 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_y^2\right] \tag{21}$$

for any $y \in \{0,1\}^n$. Note that the hiding property follows form the left and right invariance of the Haar measure under unitary transformations. Furthermore, from Eq. (21), for any ensemble $\mathcal{B}$ satisfying both hiding and anticoncentration with respect to Definition 9,

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[p_x\right] = \frac{1}{2^n}, \quad \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_x^2\right] = \frac{\mathcal{O}(1)}{4^n} \tag{22}$$

for any $x \in \{0,1\}^n$. A number of ensembles satisfy hiding, including noiseless random quantum circuits and circuits with Pauli noise [6]. Let us state a proposition that follows from our discussion so far.

Proposition 1. Let $\mathcal{B}$ be an ensemble of random quantum circuits that satisfy hiding and anticoncentration with respect to Definition 9. Then, it also satisfies anticoncentration with respect to Sec. 11.

Proof. For any $x \in \{0,1\}^n$, by the Paley-Zygmund inequality,

$$\Pr[p_x \geq \alpha \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_x\right]] \geq \frac{(1-\alpha)^2}{4^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_x^2\right]}. \tag{23}$$

Then, the proof follows from Eq. (22). ■

In this sense, Definition 9 is stronger than Sec. 11.

Remark 2. When an ensemble $\mathcal{B}$ does not satisfy hiding, the relation between Definition 9 and Sec. 11 is not clear. Note that hiding is not satisfied in most of the setups that we study in this paper, because of how we model our noise. The details of our noise model are given in Sec. III C.

D. Fine graining the weak definition

One can consider a fine-grained version of Sec. 11. The fine graining is important in setups in which hiding is not satisfied and analyzing the typical probability weight for one particular bit string does not tell us about the typical behavior of other bit strings.

Definition 14. An ensemble $\mathcal{B}$ of n-qubit quantum circuits is defined to be *k-anticoncentrated* if there exists a set $S = \{x : x \in \{0,1\}^n\}$, with $|S| = k$, such that, for every $x \in S$, there exists a choice of $\alpha, \beta \in (0,1]$ satisfying

$$\Pr_{\mathcal{B}}\left[p_x \geq \frac{\alpha}{2^n}\right] \geq \beta. \tag{24}$$

Note that after a sufficiently large depth, noiseless random circuits, or random circuits with Pauli noise, are 2^n anticoncentrated [20].

Our work shows that for the setups we discuss in the paper, anticoncentration fails both with respect to Definition 9 and Sec. 11. Our analysis is fine grained, in the spirit of Definition 14.

V. PROOF TECHNIQUES AND STRATEGY

A. Techniques

There are three main classes of techniques that we utilize in our proofs. For many of our calculations, such as those involving putting bounds on the expected collision probability or those involving computing the first moment of output probabilities, we first "remove" the last layer of noise, compute relevant quantities for the modified circuit, and then generalize our calculations to the actual circuit.

This is usually done by considering the adjoint map of the last noise layer. For our calculations about typical probabilities, we use light-cone arguments and a statistical model.

1. Removing the last layer of noise

The technique of removing the last layer of noise by considering the adjoint of the noise channel makes calculations convenient because if our circuit terminates with a last layer of single-qubit Haar-random gates, instead of a last layer of noise, then we can then use many properties of the Haar measure directly, such as translational invariance or explicit expressions of higher moments. Additionally, this makes many of our results extremely general, especially the ones involving lack of anticoncentration, by proving divergence of scaled collision probabilities, because this technique has *no* dependence on the underlying architecture, unlike similar divergence results, for noiseless and unital noise models, in Ref. [20], which involve a statistical model and, hence, are only known to be applicable for certain specific architectures. Moreover, because of this proof technique, our proof of divergence holds for *any* circuit depth, as long as there is a last layer of noise.

To be more formal, suppose that we have an ensemble $\mathcal{B}$ of noisy random quantum circuits with noise channel $\mathcal{N}$ and pick a quantum circuit $\mathcal{C} \in \mathcal{B}$. In our analysis, we usually "remove" the last layer of noise and deal with the adjoint of the noise. More specifically, let $\mathcal{C}'$ be the quantum circuit obtained by removing the last layer of noise; i.e.,

$$\mathcal{C} = \mathcal{N}^{\otimes n} \circ \mathcal{C}'. \tag{25}$$

Let $\mathcal{B}'$ be the set of quantum channels obtained by removing the last layer of noise from the circuits in $\mathcal{B}$. The expected probability of obtaining the result $x \in \{0,1\}^n$ is

$$\mathop{\mathbb{E}}_{\mathcal{B}} [\mathrm{Tr}\,(|x\rangle\langle x|\ \mathcal{C}(|0\rangle\langle 0|))]\,. \tag{26}$$

By the definition of the adjoint map, we have

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mathcal{B}} [\mathrm{Tr}\,(|x\rangle\langle x|\ \mathcal{C}(|0\rangle\langle 0|))]\\ &\quad= \mathop{\mathbb{E}}_{\mathcal{B}'} \left[\mathrm{Tr}\left((\mathcal{N}^{\dagger})^{\otimes n}(|x\rangle\langle x|)\ \mathcal{C}'(|0\rangle\langle 0|)\right)\right].\end{aligned} \tag{27}$$

By the linearity of the trace and $\mathbb{E}$,

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mathcal{B}'} \left[\mathrm{Tr}\left((\mathcal{N}^{\dagger})^{\otimes n}(|x\rangle\langle x|)\ \mathcal{C}'(|0\rangle\langle 0|)\right)\right]\\ &\quad= \mathrm{Tr}\left((\mathcal{N}^{\dagger})^{\otimes n}(|x\rangle\langle x|) \mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathcal{C}'(|0\rangle\langle 0|)\right]\right).\end{aligned} \tag{28}$$

Hence, to analyze this expected probability, we may evaluate

$$(\mathcal{N}^{\dagger})^{\otimes n}(|x\rangle\langle x|) \quad \text{and} \quad \mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathcal{C}'(|0\rangle\langle 0|)\right] \tag{29}$$

separately. $(\mathcal{N}^{\dagger})^{\otimes n}(|x\rangle\langle x|)$ can usually be evaluated straightforwardly when the description of noise $\mathcal{N}$ is given. Note that $\mathcal{C}'$ terminates with a last layer of two-qubit Haar-random gates.

Remark 3. Just as in the computation of first-moment quantities, the trick of "removing" the last layer of noise by taking its adjoint is useful even in bounding certain second-moment quantities, such as the collision probability, as we detail in Sec. VI.

2. Light-cone arguments

The second class of techniques that we utilize to study low-depth circuits are variants of light-cone-type arguments. These techniques are popular in the study of low-depth random circuits in different settings (see, e.g., Refs. [6,42]).

The qualitative intuition behind light-cone arguments is that by looking at the size of the light cone for each qubit marginal at low depth, one could argue that the circuit does not "scramble" the distribution too well for sufficiently many strings to have high probability mass. Then, by studying particular noise channels, one could argue that specific noises do not assist in the "scrambling" either. Additionally, because of small light-cone sizes, instead of directly looking at the random variable p_x—the output probability of a string $x \in \{0,1\}^n$—it suffices to look at the random variable

$$-\frac{1}{n}\log p_x, \tag{30}$$

and prove its concentration around the mean by Markov's inequality; stronger second-moment bounds are not needed. Let us emphasize that we succeed in developing techniques that are more general than variants that came before, which may be of independent interest. Previously, the analysis and application of these methods only extended to the noiseless or the unital case, as discussed in papers such as Refs. [6,20].

To use light-cone arguments to show how certain output strings, with high Hamming weight, have very low probability mass at low depth, the key ingredient is to show that the lower bound of the total variation distance between the noisy distribution and the noiseless distribution is exponentially decaying in the depth of the circuit. In Theorem 7, we prove how this holds true whenever the noise channel $\mathcal{N}$ satisfies

$$\left\langle |0\rangle\langle 0|\,, \mathcal{N}^{d}(|0\rangle\langle 0|)\right\rangle = \kappa + \tau\lambda^{d}, \tag{31}$$

where d is the depth of the circuit, $\mathcal{N}^d$ represents the quantum channel obtained by concatenating the channel $\mathcal{N}$ d times, and κ, τ, and λ are constants satisfying $\frac{1}{2} \le \kappa \le 1$, $0 \le \tau, \lambda \le 1$. More specifically, this statement is true when the noise channel under consideration is a mixture of amplitude damping and depolarizing noise. This analysis extends that in Ref. [6], where a similar technique has been used for the noiseless case and the case with only Pauli noise.

3. Mapping to classical partition functions

The third type of technique that we utilize, to study sufficiently deep circuits, consists of mappings to classical partition functions. These techniques have been developed to study the output distribution of random quantum circuits in different settings (see, e.g., Refs. [20,43]).

Since light-cone sizes blow up for superlogarithmic depths, light-cone arguments are no longer tight and we need stronger second-moment inequalities to study the output distribution. To use these inequalities, we explicitly upper bound the second moment of the distribution. This is done using mappings to classical partition functions.

To show that the same strings have low probability mass at sufficiently high depths by applying second-moment inequalities, we need to upper bound the second moment of their output probabilities using a statistical model.

To do this, one standard way is to iteratively replace each two-qubit random gate in the circuit with two copies of a single-qubit random gate; show, using the statistical model, that the collision probability of this modified circuit upper bounds the collision probability of the actual circuit; and then directly upper bound the collision probability of the modified circuit. The steps are explained in detail in Sec. IX, where we show the correctness of each step under the condition that

$$\tilde{M}_{U_1,\mathsf{N}}(I_4) = (1-a)I_4 + 2aS, \tag{32}$$

$$\tilde{M}_{U_1,\mathsf{N}}(S) = bI_4 + (1-2b)S, \tag{33}$$

with $a > 0$, and $b > 0$, where I_4 is the two-qubit identity operator and S is the two-qubit SWAP gate,

$$M_{U_1}[\rho] = \mathop{\mathbb{E}}_{U_1\sim\mathcal{U}_{\mathrm{Haar}}}\left[U_1^{\otimes 2}\rho U_1^{\dagger\otimes 2}\right], \tag{34}$$

$\mathsf{N} = \mathcal{N}\otimes\mathcal{N}$, and the operator

$$\tilde{M}_{U_1,\mathsf{N}} = M_{U_1}\circ\mathsf{N}\circ M_{U_1}. \tag{35}$$

Our analysis is inspired by the techniques in Refs. [6,20], the analysis of which only covers the noiseless and the unital cases. The authors' analysis does not work for general noise models, as it is nontrivial to prove that when we iteratively replace each two-qubit gate with two copies of a single-qubit gate in a noisy circuit, the statistical-mechanics representation is still a valid one—i.e., that there are no negative path weights for I and S—and the original collision probability is upper bounded by that of the modified circuits. We prove that this is indeed the case in Lemma 6.

B. Proof strategy

In this section, we summarize our proofs. The proofs are detailed in the subsequent sections. Just like Sec. II A, this section gives a bird's eye view of the rest of the paper—but it is more formal than Sec. II A.

First, we will prove that the distribution exhibits a lack of anticoncentration, according to Definition 10, at any depth. In particular, we will show that for a noisy ensemble $\mathcal{B}$, the scaled collision probability

$$\mathcal{Z} \geq (1+s)^n - 1 \tag{36}$$

for a non-negative constant s depends on the strength of the noises present.

Then, we show how the distribution lacks anticoncentration according to Definition 12, at any depth. Moreover, our results show that the distribution is *never* 2^{n-1} anticoncentrated, according to Definition 14. For this, first we calculate the first moment of output probabilities to show that, in expectation, the probability weight on a string is exponentially suppressed with respect to the Hamming weight of the string: strings with lower Hamming weight have exponentially more weight than strings with higher Hamming weight. Intuitively, this behavior comes from the fact that the fixed point of an n-fold tensor product of a single-qubit amplitude-damping channel is $|0^n\rangle\langle 0^n|$ and the fact that 0^n is a string with a Hamming weight of 0. So, the distribution is biased toward strings that are closer in Hamming distance to 0^n. We then show that for any string with Hamming weight at least $n/2$, the probability weight on that string is negligible, for most circuits in the ensemble $\mathcal{B}$. This calculation is divided into two parts: the low-depth and the high-depth regime, with different techniques for each regime. We use a variant of a light-cone argument in the low-depth regime, whereas the high-depth regime is analyzed using mappings to a statistical model. Note that mapping random circuits to statistical models, to study various quantities of interest, is a useful analysis tool and has been studied in Refs. [6,20,44]. Additionally, we generalize our techniques to an arbitrary noise channel. First, we show that for an arbitrary noise channel with a nonunital component, the distribution exhibits a lack of anticoncentration according to Definition 10, for a wide range of parameters. Thereafter, we derive a condition for which the noisy distribution shows a lack of anticoncentration according to Definition 12 and is never 2^{n-1} anticoncentrated. This result holds for a general noise model, when the noise is modeled as any CPTP map.

Note that a layer of random gates can, intuitively, be thought to "scramble" the output distribution. On the other hand, a layer of amplitude-damping noise can be thought to "unscramble" the distribution and push it back to a pure state. So, one might suspect that there is a "see-saw" effect and whether the distribution exhibits a lack of anticoncentration is dependent on whether we terminate our circuit with a layer of noise or a layer of noiseless gates. However, we do not think this is the case: we argue that amplitude-damping noise in the middle layers is sufficient to cause lack of anticoncentration. To elaborate on this conceptual point that we want to make, at the end of our paper, we discuss a setup in which we do not have the last layer of noise and instead terminate with a last layer of noiseless gates.

We argue how such setups also appear to lack anticoncentration, according to Definitions 10 and 12. Our results in this regime are not as general as they were before but, nonetheless, they strongly suggest that the phenomenon of lack of anticoncentration holds true even when we have nonunital noise in only the middle layers.

VI. LACK OF ANTICONCENTRATION USING SCALED COLLISION PROBABILITY

In this section, we will show how our random-circuit ensemble exhibits a lack of anticoncentration according to Definition 10. The noise is modeled by a mixture of amplitude damping and depolarizing noise. That is, the combined noise channel could be either $\mathcal{N}_q^{(\text{amp})} \circ \mathcal{N}_p^{(\text{dep})}$ or $\mathcal{N}_p^{(\text{dep})} \circ \mathcal{N}_q^{(\text{amp})}$.

Let us observe the following identities:

$$|0\rangle\langle 0| = \frac{I_2 + \sigma_z}{2}, \tag{37}$$

$$|1\rangle\langle 1| = \frac{I_2 - \sigma_z}{2}, \tag{38}$$

where I_2 is the single-qubit identity operator and σ_z is the single-qubit Pauli-Z operator. Using these identities, for an ensemble $\mathcal{B}$, one could rewrite Eq. (14) as

$$\mathcal{Z} = \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{p\in\{0,3\}^n, p\neq 0^n} \text{Tr}(\sigma_p\, \mathcal{C}(|0\rangle\langle 0|))^2\right], \tag{39}$$

where σ_p is an n-qubit Pauli operator. Note that $\sigma_0 = I_2$ and $\sigma_3 = \sigma_z$. The derivation is shown in more detail in Appendix E. Additionally, note that

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\text{Tr}(\sigma_p\, \mathcal{C}(|0\rangle\langle 0|))^2\right] = \mathop{\mathbb{E}}_{\mathcal{B}}\left[\text{Tr}(\sigma_p \otimes \sigma_p\, \mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|))\right]. \tag{40}$$

We prove the following theorem, which shows a lack of anticoncentration for our noise model.

Theorem 1. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits with noise channel $\mathcal{N}$, where $\mathcal{N}$ is either $\mathcal{N}_q^{(\text{amp})} \circ \mathcal{N}_p^{(\text{dep})}$ or $\mathcal{N}_p^{(\text{dep})} \circ \mathcal{N}_q^{(\text{amp})}$. Then,

$$\mathcal{Z} \geq \left(1 + r^2\right)^n - 1, \tag{41}$$

where

$$r := \begin{cases} q, & \mathcal{N} = \mathcal{N}_q^{(\text{amp})} \circ \mathcal{N}_p^{(\text{dep})}, \\ q(1-p), & \mathcal{N} = \mathcal{N}_p^{(\text{dep})} \circ \mathcal{N}_q^{(\text{amp})}. \end{cases} \tag{42}$$

Proof. For pedagogical reasons, before we generalize to n qubits, let us first consider just the single-qubit case. In the single-qubit case, we have

$$\mathcal{Z} = \mathop{\mathbb{E}}_{\mathcal{B}}\left[\text{Tr}(\sigma_z\, \mathcal{C}(|0\rangle\langle 0|))^2\right] \tag{43}$$

$$= \mathop{\mathbb{E}}_{\mathcal{B}}\left[\text{Tr}(\sigma_z \otimes \sigma_z\, \mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|))\right] \tag{44}$$

by definition. Let $\mathcal{N}$ be either $\mathcal{N}_q^{(\text{amp})} \circ \mathcal{N}_p^{(\text{dep})}$ or $\mathcal{N}_p^{(\text{dep})} \circ \mathcal{N}_q^{(\text{amp})}$. Let

$$\rho = \mathcal{N}\left(U_1(\tilde{\rho})U_1^\dagger\right), \tag{45}$$

where $\tilde{\rho}$ is the state just before the last block. Additionally, let

$$\rho' = U_1(\tilde{\rho})U_1^\dagger. \tag{46}$$

By definition of the adjoint map, we have

$$\mathcal{Z} = \mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(\sigma_z \otimes \sigma_z \rho \otimes \rho)\right] \tag{47}$$

$$= \mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(\sigma_z \otimes \sigma_z\, \mathcal{N}(\rho') \otimes \mathcal{N}(\rho'))\right] \tag{48}$$

$$= \mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(\mathcal{N}^\dagger(\sigma_z) \otimes \mathcal{N}^\dagger(\sigma_z)\, \rho' \otimes \rho')\right]. \tag{49}$$

In Eq. (49), we have used the definition of the adjoint map of a channel $\mathcal{N}$. We can explicitly compute $\mathcal{N}^\dagger(\sigma_z)$ as

$$\mathcal{N}^\dagger(\sigma_z) = rI_2 + (1-q)(1-p)\sigma_z, \tag{50}$$

where r is defined in Eq. (42). Using this expansion, we have

$$\begin{aligned} &\mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(\mathcal{N}^\dagger(\sigma_z) \otimes \mathcal{N}^\dagger(\sigma_z)\, \rho' \otimes \rho')\right] \\ &= r^2 \mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(I_2 \otimes I_2 \rho' \otimes \rho')\right] + c_1 \mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(I_2 \otimes \sigma_z \rho' \otimes \rho')\right] \\ &\quad + c_1 \mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(\sigma_z \otimes I_2 \rho' \otimes \rho')\right] \\ &\quad + c_2^2 \mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(\sigma_z \otimes \sigma_z \rho' \otimes \rho')\right], \end{aligned} \tag{51}$$

where

$$c_1 := \begin{cases} q(1-q)(1-p), & \mathcal{N} = \mathcal{N}_q^{(\text{amp})} \circ \mathcal{N}_p^{(\text{dep})}, \\ q(1-p)(1-p)^2, & \mathcal{N} = \mathcal{N}_p^{(\text{dep})} \circ \mathcal{N}_q^{(\text{amp})}, \end{cases} \tag{52}$$

$$c_2 := \begin{cases} (1-q), & \mathcal{N} = \mathcal{N}_q^{(\text{amp})} \circ \mathcal{N}_p^{(\text{dep})}, \\ (1-q)(1-p), & \mathcal{N} = \mathcal{N}_p^{(\text{dep})} \circ \mathcal{N}_q^{(\text{amp})}. \end{cases} \tag{53}$$

Now, we evaluate the terms one by one. The first term can be exactly computed as

$$\mathop{\mathbb{E}}_{U_1}\left[\text{Tr}(I_2 \otimes I_2 \rho' \otimes \rho')\right] = 1, \tag{54}$$

by definition.

The second term is evaluated as

$$\mathop{\mathbb{E}}_{U_1}\left[\mathrm{Tr}(I_2 \otimes \sigma_z \rho' \otimes \rho')\right] \tag{55}$$

$$= \mathop{\mathbb{E}}_{U_1}\mathrm{Tr}\left(\left[U_1^{\dagger\otimes 2}(I_2 \otimes \sigma_z)U_1^{\otimes 2}\right]\tilde{\rho} \otimes \tilde{\rho}\right) \tag{56}$$

$$= \mathrm{Tr}\left(\mathop{\mathbb{E}}_{U_1}\left[U_1^{\dagger\otimes 2}(I_2 \otimes \sigma_z)U_1^{\otimes 2}\right]\tilde{\rho} \otimes \tilde{\rho}\right) \tag{57}$$

$$= 0. \tag{58}$$

The second line follows because of the linearity of trace and expectation to interchange the two operations suitably. In the third line, we have used the fact that

$$M_{U_1}[\sigma_i \otimes \sigma_j] = \mathop{\mathbb{E}}_{U_1}\left[U_1^{\dagger\otimes 2}\sigma_i \otimes \sigma_j U_1^{\otimes 2}\right] \tag{59}$$

$$= \mathop{\mathbb{E}}_{U_1}\left[U_1^{\otimes 2}\sigma_i \otimes \sigma_j U_1^{\dagger\otimes 2}\right] \tag{60}$$

$$= 0, \tag{61}$$

for $i \neq j$. The second line follows because the Hermitian conjugate of a unitary is also a unitary and because the expectation is only over the unitary U_1. The third line is obtained by a straightforward calculation with Eq. (F9) in Sec. F. Here, let us give qualitative reasoning for this relation. Observe that M_{U_1} projects an input onto the symmetric subspace and the antisymmetric subspace (See Sec. F). The projector Π_{sym} onto the symmetric subspace and the projector Π_{antisym} onto the antisymmetric subspace can be expressed as

$$\Pi_{\mathrm{sym}} = |\Phi^+\rangle\langle\Phi^+| + |\Phi^-\rangle\langle\Phi^-| + |\Psi^+\rangle\langle\Psi^+| \tag{62}$$

$$\Pi_{\mathrm{antisym}} = |\Phi^-\rangle\langle\Phi^-|, \tag{63}$$

where

$$|\Phi^+\rangle := \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \tag{64}$$

$$|\Phi^-\rangle := \frac{|00\rangle - |11\rangle}{\sqrt{2}}, \tag{65}$$

$$|\Psi^+\rangle := \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \tag{66}$$

$$|\Psi^-\rangle := \frac{|01\rangle - |10\rangle}{\sqrt{2}} \tag{67}$$

are the Bell states. In fact, for any Bell state $|\Xi\rangle$, we have $\langle\Xi|\,|\sigma_i \otimes \sigma_j|\,|\Xi\rangle = 0$ when $i \neq j$. An intuitive justification can be given with the following simultaneous measurement scenario. Suppose that two parties—Alice and Bob—share a Bell state $|\Xi\rangle$. Alice has an observable σ_i and Bob has an observable σ_j. They measure their observables with the state $|\Xi\rangle$. In this case, the expectation value of this measurement is zero. Indeed, Bob obtains 1 with probability $\frac{1}{2}$ and obtains -1 with probability $\frac{1}{2}$ regardless of Alice's measurement result, because they share a maximally entangled state and they locally have different Pauli operators. Hence, from this observation, we have

$$\langle\sigma_i \otimes \sigma_j, \Pi_{\mathrm{sym}}\rangle = \langle\sigma_i \otimes \sigma_j, \Pi_{\mathrm{antisym}}\rangle = 0. \tag{68}$$

Hence, $\sigma_i \otimes \sigma_j$ has no component on the symmetric and antisymmetric subspaces. Thus,

$$M_{U_1}[\sigma_i \otimes \sigma_j] = 0. \tag{69}$$

Note that this does not contradict that M_{U_1} is a CPTP map because $\mathrm{Tr}[\sigma_i \otimes \sigma_j] = 0$ when $i \neq j$.

By a similar reasoning as the second term, the third term is

$$\mathop{\mathbb{E}}_{U_1}\left[\mathrm{Tr}(\sigma_z \otimes I_2 \rho' \otimes \rho')\right] = 0. \tag{70}$$

The fourth term can be bounded as

$$\mathop{\mathbb{E}}_{U_1}\left[\mathrm{Tr}(\sigma_z \otimes \sigma_z \rho' \otimes \rho')\right] = \mathop{\mathbb{E}}_{U_1}\left[\mathrm{Tr}(\sigma_z \rho')^2\right] \geq 0. \tag{71}$$

This is because $\mathrm{Tr}(\sigma_z \rho')^2$ is always a positive-valued random variable. Putting these analyses together, Eq. (51) can be lower bounded by r^2; i.e.,

$$\mathcal{Z} \geq r^2. \tag{72}$$

With this observation for the single-qubit case, we consider the general n-qubit case:

$$\mathcal{Z} = \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{p \in \{0,3\}^n, p \neq 0^n} \mathrm{Tr}(\sigma_p \otimes \sigma_p\, \mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|))\right]. \tag{73}$$

Let us fix $p \in \{0,3\}^n \backslash \{0^n\}$ and consider

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\mathrm{Tr}(\sigma_p \otimes \sigma_p\, \mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|))\right]. \tag{74}$$

Letting $\mathcal{C}$' denote the circuit obtained by removing the last layer of noise, we have

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\mathrm{Tr}(\sigma_p \otimes \sigma_p\, \mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|))\right]$$

$$= \mathop{\mathbb{E}}_{\mathcal{B}}\left[\mathrm{Tr}(\sigma_p \otimes \sigma_p\, \mathcal{N}^{\otimes n} \circ \mathcal{C}'(|0\rangle\langle 0|) \otimes \mathcal{N}^{\otimes n} \circ \mathcal{C}'(|0\rangle\langle 0|))\right] \tag{75}$$

$$= \mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathrm{Tr}((\mathcal{N}^\dagger)^{\otimes n}(\sigma_p) \otimes (\mathcal{N}^\dagger)^{\otimes n}(\sigma_p)\, \mathcal{C}'(|0\rangle\langle 0|) \otimes \mathcal{C}'(|0\rangle\langle 0|))\right], \tag{76}$$

where $\mathcal{B}'$ is the set of quantum channels obtained by removing the last layer of noise from the circuits in

$\mathcal{B}$. Now, we consider the expansion of $(\mathcal{N}^\dagger)^{\otimes n}(\sigma_p)$. Let us write $p = p_1p_2\cdots p_n$, where $p_i \in \{0,3\}$ for $1 \le i \le n$. Then,

$$(\mathcal{N}^\dagger)^{\otimes n}(\sigma_p) = \bigotimes_{i=1}^{n} \mathcal{N}^\dagger(\sigma_{p_i}). \tag{77}$$

Recalling Eq. (50),

$$\mathcal{N}^\dagger(\sigma_{p_i}) = \begin{cases} I_2, & p_i = 0, \\ rI_2 + (1-q)(1-p)\sigma_z, & p_i = 3. \end{cases} \tag{78}$$

Note that $\mathcal{N}^\dagger(I_2) = I_2$, since the adjoint map of a quantum channel is unital. Substituting this relation into Eq. (77), we have

$$(\mathcal{N}^\dagger)^{\otimes n}(\sigma_p) = r^{w_p} I_{2^n} + \sum_{\substack{q\in\{0,3\}^n\setminus\{0^n\} \\ w_q \le w_p}} c_q\sigma_q, \tag{79}$$

where the c_q are non-negative coefficients and w_i is the number of nonidentity Pauli operators in σ_i. Using this expression,

$$\begin{aligned} &\mathop{\mathbb{E}}_{\mathcal{B}}\left[\mathrm{Tr}(\sigma_p \otimes \sigma_p\ \mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|))\right] \\ &\quad = r^{2w_p} \mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathrm{Tr}(I_{2^n} \otimes I_{2^n}\ \mathcal{C}'(|0\rangle\langle 0|) \otimes \mathcal{C}'(|0\rangle\langle 0|))\right] \\ &\qquad + \sum_{\substack{q,q'\in\{0,3\}^n\setminus\{0^n\} \\ w_q \le w_p}} c_q c_{q'} \mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathrm{Tr}(\sigma_q \otimes \sigma_{q'}\ \mathcal{C}'(|0\rangle\langle 0|) \right. \\ &\qquad \left. \otimes\ \mathcal{C}'(|0\rangle\langle 0|))\right]. \end{aligned} \tag{80}$$

The first term is

$$r^{2w_p} \mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathrm{Tr}(I_{2^n} \otimes I_{2^n}\ \mathcal{C}'(|0\rangle\langle 0|) \otimes \mathcal{C}'(|0\rangle\langle 0|))\right] = r^{2w_p}. \tag{81}$$

If $q \neq q'$, then, with a similar argument as in Eqs. (58) and (70), we have

$$\mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathrm{Tr}(\sigma_q \otimes \sigma_{q'}\ \mathcal{C}'(|0\rangle\langle 0|) \otimes \mathcal{C}'(|0\rangle\langle 0|))\right] = 0. \tag{82}$$

Here, we have used the fact that the average statistics do not change even if one appends single-qubit Haar-random unitary gates after a two-qubit Haar-random unitary gate. In addition,

$$\mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathrm{Tr}(\sigma_q \otimes \sigma_q\ \mathcal{C}'(|0\rangle\langle 0|) \otimes \mathcal{C}'(|0\rangle\langle 0|))\right] \tag{83}$$

$$= \mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathrm{Tr}(\sigma_q\mathcal{C}'(|0\rangle\langle 0|))^2\right] \tag{84}$$

$$\ge 0, \tag{85}$$

by a similar discussion in Eq. (71). By combining these observations,

$$\mathcal{Z} \ge \sum_{p\in\{0,3\}^n, p\neq 0^n} r^{2w_p}. \tag{86}$$

Hence, by computing the right-hand side with the binomial theorem,

$$\mathcal{Z} \ge (1+r^2)^n - 1, \tag{87}$$

which completes the proof. ∎

The lower bound in Eq. (41), established in Theorem 1, indicates that the scaled collision probability diverges in the limit of $n \to \infty$, i.e., the given circuit is not anticoncentrated, if the noise parameter $r \neq 0$. For example, recall that $r = q$ when $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}$. Hence, the circuit is not anticoncentrated as long as the circuit is affected by the amplitude-damping noise, no matter how strong the depolarizing noise is. On the other hand, when $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}$, the noise parameter is $r = q(1-p)$. In this case, the circuit is not anticoncentrated for any positive q unless $p = 1$. This slight difference originates from the order of the two kinds of noise. When the completely depolarizing noise comes after the amplitude-damping noise, the completely depolarizing noise nullifies the effects of the amplitude-damping noise.

Remark 4. If we only have the depolarizing channel, then $q = 0$. Then, from Eq. (42), $r = 0$. Hence, from Eq. (86),

$$\mathcal{Z} \ge 0. \tag{88}$$

This gives us a vacuous bound and our techniques fail to prove the lack of anticoncentration. This is consistent with the observation that random quantum circuits with depolarizing noise indeed anticoncentrate [6].

VII. USEFUL PROPERTIES OF OUTPUT DISTRIBUTION OF NOISY RANDOM CIRCUITS

In the next sections, we will prove that our setup exhibits a lack of anticoncentration according to Definition 12. Before doing that, it will be helpful to prove some useful results.

To that effect, in this section, we calculate the first moment of two different noise models. We divide our proof into many subparts. First, let us calculate the exact expression for the first moment of the output probabilities of our distribution. We will then argue about typical probabilities, first in the low-depth regime, using a light-cone argument, and then in the high-depth regime, using a second-moment inequality.

A. First moment of output probabilities

Theorem 2. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits with noise channel $\mathcal{N}$. Furthermore, for a particular $x \in \{0,1\}^n$, let p_x be the corresponding outcome probability. Then, the following hold:

(1) If $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}$,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] = \frac{(1-q)^{w_x}(1+q)^{n-w_x}}{2^n}. \tag{89}$$

(2) If $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}$,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] = \frac{(1-(1-p)q)^{w_x}(1+(1-p)q)^{n-w_x}}{2^n}. \tag{90}$$

Proof. We will prove the two cases separately. We first prove Eq. (89). Observe that

$$\left(\mathcal{N}_p^{(\mathrm{dep})}\right)^\dagger \circ \left(\mathcal{N}_q^{(\mathrm{amp})}\right)^\dagger (|0\rangle\langle 0|) = \left(1-\frac{p}{2}+\frac{pq}{2}\right)|0\rangle\langle 0| + \left(q+\frac{p}{2}-\frac{pq}{2}\right)|1\rangle\langle 1|, \tag{91}$$

$$\left(\mathcal{N}_p^{(\mathrm{dep})}\right)^\dagger \circ \left(\mathcal{N}_q^{(\mathrm{amp})}\right)^\dagger (|1\rangle\langle 1|) = \left(\frac{p}{2}-\frac{pq}{2}\right)|0\rangle\langle 0| + \left(1-q-\frac{p}{2}+\frac{pq}{2}\right)|1\rangle\langle 1|, \tag{92}$$

where for a quantum channel $\mathcal{N}$, $\mathcal{N}^\dagger$ represents its adjoint map. Now, construct a new ensemble $\mathcal{B}'$ by removing the last layer of the noise channel from the circuits in $\mathcal{B}$. Let $\mathcal{C}$ be a quantum circuit in $\mathcal{B}$ and let $\mathcal{C}'$ be the circuit obtained by removing the last layer of noise, namely,

$$\mathcal{C} = \mathcal{N}^{\otimes n} \circ \mathcal{C}'. \tag{93}$$

By the definition of the adjoint map,

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] &= \mathop{\mathbb{E}}_{\mathcal{B}}\mathrm{Tr}\left[|x\rangle\langle x|\,\mathcal{C}\left(|0^n\rangle\langle 0^n|\right)\right] \\ &= \mathop{\mathbb{E}}_{\mathcal{B}'}\mathrm{Tr}\left[\left(\mathcal{N}^\dagger\right)^{\otimes n}(|x\rangle\langle x|)\,\mathcal{C}'\left(|0^n\rangle\langle 0^n|\right)\right].\end{aligned} \tag{94}$$

By computing $\left(\mathcal{N}^\dagger\right)^{\otimes n}(|x\rangle\langle x|)$ using Eqs. (91) and (92), we have

$$\left(\mathcal{N}^\dagger\right)^{\otimes n}(|x\rangle\langle x|) = \sum_{y\in\{0,1\}^n}\left(\prod_{k=1}^n c_k^{(x,y)}\right)|y\rangle\langle y|, \tag{95}$$

where $c_k^{(x,y)}$ is defined by

$$c_k^{(x,y)} = \begin{cases}\left(1-\frac{p}{2}+\frac{pq}{2}\right), & (x_k,y_k)=(0,0),\\ \left(q+\frac{p}{2}-\frac{pq}{2}\right), & (x_k,y_k)=(0,1),\\ \left(\frac{p}{2}-\frac{pq}{2}\right), & (x_k,y_k)=(1,0),\\ \left(1-q-\frac{p}{2}+\frac{pq}{2}\right), & (x_k,y_k)=(1,1).\end{cases} \tag{96}$$

On the other hand, we have

$$\mathop{\mathbb{E}}_{\mathcal{B}'}\left[\mathcal{C}'\left(|0^n\rangle\langle 0^n|\right)\right] = \frac{I_{2^n}}{2^n}, \tag{97}$$

by considering the expectation over the last layer of random unitaries. Therefore, by combining these equations,

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] &= \sum_{y\in\{0,1\}^n}\left(\prod_{k=1}^n c_k^{(x,y)}\right)\mathrm{Tr}\left[|y\rangle\langle y|\frac{I_{2^n}}{2^n}\right] \\ &= \frac{1}{2^n}\sum_{y\in\{0,1\}^n}\left(\prod_{k=1}^n c_k^{(x,y)}\right).\end{aligned} \tag{98}$$

We can compute Eq. (98) as

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] = \frac{1}{2^n}\sum_{y\in\{0,1\}^n}\left(\prod_{k=1}^n c_k^{(x,y)}\right) \tag{99}$$

$$\begin{aligned}&= \frac{1}{2^n}\sum_{l=0}^{n-w_x}\binom{n-w_x}{l}\left(1-\frac{p}{2}+\frac{pq}{2}\right)^l \\ &\quad\times\left(q+\frac{p}{2}-\frac{pq}{2}\right)^{(n-w_x)-l}\sum_{l'=0}^{w_x}\binom{w_x}{l'}\left(\frac{p}{2}-\frac{pq}{2}\right)^{l'} \\ &\quad\times\left(1-q-\frac{p}{2}+\frac{pq}{2}\right)^{w_x-l'}\end{aligned} \tag{100}$$

$$\begin{aligned}&= \frac{1}{2^n}\left(1-\frac{p}{2}+\frac{pq}{2}+q+\frac{p}{2}-\frac{pq}{2}\right)^{n-w_x} \\ &\quad\times\left(\frac{p}{2}-\frac{pq}{2}+1-q-\frac{p}{2}+\frac{pq}{2}\right)^{w_x}\end{aligned} \tag{101}$$

$$= \frac{(1+q)^{n-w_x}(1-q)^{w_x}}{2^n}, \tag{102}$$

where to obtain Eq. (100), we have divided the cases based on the Hamming weight of string x. Then, Eq. (101) follows from the binomial theorem.

Next, we show Eq. (90). In this case, we have

$$\left(\mathcal{N}_q^{(\text{amp})}\right)^\dagger \circ \left(\mathcal{N}_p^{(\text{dep})}\right)^\dagger (|0\rangle\langle 0|) = \left(1-\frac{p}{2}\right)|0\rangle\langle 0| + \left(q(1-p)+\frac{p}{2}\right)|1\rangle\langle 1|, \tag{103}$$

$$\left(\mathcal{N}_q^{(\text{amp})}\right)^\dagger \circ \left(\mathcal{N}_p^{(\text{dep})}\right)^\dagger (|1\rangle\langle 1|) = \left(\frac{p}{2}\right)|0\rangle\langle 0| + \left(1-q(1-p)-\frac{p}{2}\right)|1\rangle\langle 1|. \tag{104}$$

With the same argument from case 1, we have

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] = \frac{(1+q(1-p))^{n-w_x}(1-q(1-p))^{w_x}}{2^n}. \tag{105}$$

■

From Theorem 2, it is evident that the expected output probabilities of strings with higher Hamming weights are exponentially suppressed with respect to the Hamming weight.

B. A discussion on marginal probabilities

Our calculations to prove lack of anticoncentration, with respect to Definition 12, in the low-depth regime require an argument based on light cones. As we will soon see, this necessitates that we compute marginal probabilities, where the order of the marginal is determined by the size of the light cone. The following corollary is immediate from Theorem 2.

Corollary 1. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits with noise channel $\mathcal{N}$. For a binary string $x \in \{0,1\}^n$, consider a substring y of x with length $|y|$ and Hamming weight w_y. Let p_y be the corresponding marginal probability. Then, the following hold:

(1) If $\mathcal{N} = \mathcal{N}_q^{(\text{amp})} \circ \mathcal{N}_p^{(\text{dep})}$,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_y] = \frac{(1-q)^{w_y}(1+q)^{|y|-w_y}}{2^{|y|}}. \tag{106}$$

(2) If $\mathcal{N} = \mathcal{N}_p^{(\text{dep})} \circ \mathcal{N}_q^{(\text{amp})}$,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_y] = \frac{(1-(1-p)q)^{w_y}(1+(1-p)q)^{|y|-w_y}}{2^{|y|}}. \tag{107}$$

Proof. We only show the proof for $|y| = n-1$. The other cases also follow similarly. Without loss of generality, we may assume $y = x_1x_2\cdots x_{n-1}$, where x_i denotes the ith bit of x; i.e., y is the substring obtained by discarding the last bit of x. Consider the case in which $\mathcal{N} = \mathcal{N}_q^{(\text{amp})} \circ \mathcal{N}_p^{(\text{dep})}$. The other case holds similarly.

Since the marginal probability $p_{x_1x_2\cdots x_{n-1}}$ is the sum of $p_{x_1x_2\cdots x_{n-1}0}$ and $p_{x_1x_2\cdots x_{n-1}1}$, we have

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[p_y\right] = \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{x_1x_2\cdots x_{n-1}}\right] \tag{108}$$

$$= \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{x_1x_2\cdots x_{n-1}0} + p_{x_1x_2\cdots x_{n-1}1}\right] \tag{109}$$

$$= \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{x_1x_2\cdots x_{n-1}0}\right] + \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{x_1x_2\cdots x_{n-1}1}\right]. \tag{110}$$

Now, from Theorem 2,

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{x_1x_2\cdots x_{n-1}0}\right] = \frac{(1+q)^{n-w_y}(1-q)^{w_y}}{2^n}, \tag{111}$$

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{x_1x_2\cdots x_{n-1}1}\right] = \frac{(1+q)^{n-(w_y+1)}(1-q)^{w_y+1}}{2^n}. \tag{112}$$

Therefore,

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[p_y\right] = \frac{(1+q)^{n-w_y}(1-q)^{w_y}}{2^n} + \frac{(1+q)^{n-(w_y+1)}(1-q)^{w_y+1}}{2^n} \tag{113}$$

$$= \frac{(1+q)^{(n-1)-w_y}(1-q)^{w_y}}{2^n}\left((1+q)+(1-q)\right) \tag{114}$$

$$= \frac{(1+q)^{(n-1)-w_y}(1-q)^{w_y}}{2^{n-1}} \tag{115}$$

$$= \frac{(1+q)^{|y|-w_y}(1-q)^{w_y}}{2^{|y|}}, \tag{116}$$

where the last equality follows because $|y| = |x_1x_2\cdots x_{n-1}| = n-1$. ■

VIII. LACK OF ANTICONCENTRATION WITH TYPICAL PROBABILITIES: LOW DEPTH

In this section, we prove that for a sublogarithmic depth noisy random-circuit ensemble, the probability weight on strings with Hamming weight at least $n/2$ is negligible in most circuits of the ensemble. This means that these circuits exhibit a lack of anticoncentration, as defined in Definition 12. Our analysis is fine grained, in the spirit of Definition 14.

We observe that by introducing random variables X_i, corresponding to the ith bit of n-bit string x, for any n-symbol permutation σ, we may write

$$p_x = \Pr[X_{\sigma(1)} = x_{\sigma(1)}]\Pr[X_{\sigma(2)} = x_{\sigma(2)}|X_{\sigma(1)} = x_{\sigma(1)}] \times \cdots \times \Pr[X_{\sigma(n)} = x_{\sigma(n)}|X_{\sigma(1)} = x_{\sigma(1)}, X_{\sigma(2)} = x_{\sigma(2)}, \ldots, X_{\sigma(n-1)} = x_{\sigma(n-1)}]. \tag{117}$$

Hence, we have

$$-\log p_x = -\frac{1}{n!}\sum_{\sigma\in S_n}\sum_{i=1}^{n}\log \Pr\Big(X_{\sigma(i)} = x_{\sigma(i)}|X_{\sigma(1)} = x_{\sigma(1)},\ldots,X_{\sigma(i-1)} = x_{\sigma(i-1)}\Big), \quad (118)$$

where S_n is the permutation group on n qubits. For a set of non-negative integers J_i, such that $i \notin J_i$, define

$$\langle Z_i\rangle_{J_i} = 2\Pr\big(X_i = x_i|\{X_j = x_j\}_{j\in J_i}\big) - 1. \quad (119)$$

If there is no conditional dependence, then we drop the subscript J_i. Finally, let

$$A_\sigma = -\sum_{i=1}^{n}\langle Z_{\sigma(i)}\rangle_{\sigma(1,2,\ldots,i-1)}. \quad (120)$$

Using a light-cone-type argument, the following lemma follows in a similar way to Ref. [6].

Lemma 1 ([6], Eq. (58)*).* For a quantum circuit $\mathcal{C}$, and for any $x \in \{0,1\}^n$, let p_x be the probability of obtaining x in the output. Then,

$$-\log p_x \ge n\log 2 + \frac{1}{n!}\sum_{\sigma\in S_n}A_\sigma + \frac{1}{4\times 4^d}\sum_{i=1}^{n}\langle Z_i\rangle^2. \quad (121)$$

Now, we prove our main theorem, showing that when the depth of the noisy random circuit with the mixture of the amplitude-damping channel and the depolarizing channel is $o(\log n)$, the circuit is never 2^{n-1} anticoncentrated.

Theorem 3. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits of depth d, with noise channel $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})}\circ\mathcal{N}_p^{(\mathrm{dep})}$ or $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})}\circ\mathcal{N}_q^{(\mathrm{amp})}$. Let $x \in \{0,1\}^n$ and w_x be the Hamming weight of x. Then, there exists a constant t for every x with $w_x \ge n/2$, and $\alpha \in (0,1]$, such that,

$$\lim_{n\to\infty}\Pr_{\mathcal{B}}\left[p_x < \frac{\alpha}{2^n}\right] = 1, \quad (122)$$

when $d < t\log n$.

Proof. From Appendix G, we have

$$\mathop{\mathbb{E}}_{\mathcal{B}}\big[\Pr\big(X_i = x_i|\{X_j = x_j\}_{j\in J_i}\big)\big] = \frac{1}{2}\left[\langle x_i|\,\mathcal{N}(I_2)\,|x_i\rangle\right]. \quad (123)$$

Here, $\langle x_i|\,\mathcal{N}(I_2)\,|x_i\rangle$ can be evaluated as follows:

(1) If $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})}\circ\mathcal{N}_p^{(\mathrm{dep})}$,

$$\langle x_i|\,\mathcal{N}(I_2)\,|x_i\rangle = \begin{cases}1+q & (x_i = 0),\\ 1-q & (x_i = 1).\end{cases} \quad (124)$$

(2) If $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})}\circ\mathcal{N}_q^{(\mathrm{amp})}$,

$$\langle x_i|\,\mathcal{N}(I_2)\,|x_i\rangle = \begin{cases}1+(1-p)q & (x_i = 0),\\ 1-(1-p)q & (x_i = 1).\end{cases} \quad (125)$$

Let us write

$$\langle x_i|\,\mathcal{N}(I_2)\,|x_i\rangle = \begin{cases}1+r, & (x_i = 0),\\ 1-r, & (x_i = 1),\end{cases} \quad (126)$$

with $0 \le r \le 1$, so that we can cover both cases. The expectation value of A_σ over $\mathcal{B}$ is computed as

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{B}}[A_\sigma] &= \sum_{i=1}^{n}\left(1 - \left[\langle x_i|\,\mathcal{N}(I_2)\,|x_i\rangle\right]\right)\\ &= (n-w_x)\,(1-(1+r)) + w_x\,(1-(1-r))\\ &= 2w_x r - nr. \end{aligned} \quad (127)$$

Additionally, from Appendix H,

$$\mathop{\mathbb{E}}_{\mathcal{B}_d}[\langle Z_i\rangle^2] \ge be^{-ad} \quad (128)$$

for some positive constants a and b, for any $i \in [n]$. Moreover, letting

$$X = -\log p_x, \quad (129)$$

from Ref. [6], it follows that

$$\mathop{\mathrm{Var}}_{\mathcal{B}}(X) \le 2n. \quad (130)$$

From Lemma 1 and Eq. (127),

$$\mathop{\mathbb{E}}_{\mathcal{B}}[X] \ge n\log 2 + (2w_x r - nr) + \frac{b}{4}ne^{-cd} \quad (131)$$

for $c = \log 4 + a$. By Chebyshev's inequality,

$$\Pr[|X - \mathop{\mathbb{E}}_{\mathcal{B}}[X]| \ge k] \le \frac{\mathop{\mathrm{Var}}_{\mathcal{B}}(X)}{k^2}. \quad (132)$$

Taking, say, $k = n^{0.01}\sqrt{\underset{\mathcal{B}}{\mathrm{Var}}(X)}$, we have

$$\Pr[|X - \underset{\mathcal{B}}{\mathbb{E}}[X]| \leq \mathcal{O}(n^{0.51})] \geq 1 - \frac{1}{n^{0.0001}}. \quad (133)$$

Hence,

$$\lim_{n\to\infty} \Pr[-X \leq -\underset{\mathcal{B}}{\mathbb{E}}[X] + \mathcal{O}(n^{0.51})] = 1. \quad (134)$$

Putting back the values, from Eqs. (129) and (131),

$$\lim_{n\to\infty} \Pr_{\mathcal{B}}\left[p_x \leq 2^{-n}\exp\left(-2w_x r + nr - \frac{b}{4}ne^{-cd} + \mathcal{O}(n^{0.51})\right)\right] = 1. \quad (135)$$

Now, if

$$2r\left(w_x - \frac{n}{2}\right) + \frac{b}{4}ne^{-cd} = \omega(n^{0.51}), \quad (136)$$

then

$$\exp\left(-2w_x r + nr - \frac{b}{4}ne^{-cd} + \mathcal{O}(n^{0.51})\right) = O(1), \quad (137)$$

and we have Eq. (122). For $w_x \geq n/2$, when we pick $d < 0.49/c \log n$, Eq. (136) is satisfied and thus the theorem follows. ∎

IX. LACK OF ANTICONCENTRATION WITH TYPICAL PROBABILITIES: HIGH DEPTH

We need a different technique to analyze sufficiently deep circuits and show that they satisfy Definition 10. This is because the light-cone-type arguments, in Lemma 1, break down when the light-cone sizes become too large for sufficiently deep circuits. Because of this, the technique that we use is to bound the second moment of the output probabilities and then use Chebyshev's inequality to show concentration around the mean for our desired probabilities. Just as in Sec. VIII, our results are fine grained, in the spirit of Definition 14. In the following theorem, we upper bound the second moment of an n-qubit depth-d noisy circuit. In the proof, we obtain the upper bound by repeatedly replacing a two-qubit Haar random unitary gate with two single-qubit Haar random unitary gates. We give the proof in Appendix I, using a statistical model.

Theorem 4. Consider an ensemble $\mathcal{B}$ of depth-d noisy random quantum circuits characterized by a noisy channel $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}$ or $\mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}$. Suppose that noise parameters (p, q) satisfies $(p, q) \neq (0, 0)$; i.e., we will not consider the noiseless case. Then, for $x \in \{0, 1\}^n$ with the Hamming weight $w_x \geq \frac{n}{2}$,

$$\underset{\mathcal{B}}{\mathbb{E}}[p_x^2] \leq \mu^n \eta^n \exp\left[n\frac{\nu}{\mu}\mathrm{e}^{-c(d-1)}\right], \quad (138)$$

where

$$\mu := \frac{1}{4} + \frac{r^2}{12c} \quad (\geq 0), \quad (139)$$

$$\nu := \frac{1}{12} - \frac{r^2}{12c} \quad (\geq 0), \quad (140)$$

$$\eta = 1 - r^2 \quad (\geq 0), \quad (141)$$

with

$$c := 1 - (1-p)^2(1-q)\left(1 - \frac{q}{3}\right) \quad (142)$$

$$r := \begin{cases} q, & \mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}, \\ q(1-p), & \mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}. \end{cases} \quad (143)$$

Remark 5. In the statement, since $(p, q) \neq (0, 0)$, c is always larger than 0 by definition. Thus, μ and ν are always well defined.

With the bound provided in Theorem 4, we can now show the lack of anticoncentration in the high-depth regime, which is formally stated below.

Theorem 5. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits of depth d. Let the noise channel be $\mathcal{N}$ and let $d = \Omega(\log n)$. Then, the following statements hold:

(1) When $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}$, then, for every $x \in \{0, 1\}^n$ with $w_x \geq n/2$ and $\alpha \in (0, 1]$,

$$\lim_{n\to\infty} \Pr_{\mathcal{B}}\left[p_x < \frac{\alpha}{2^n}\right] = 1, \quad (144)$$

as long as

$$\frac{q^2 + 2}{(q-3)(q-1)} > (1-p)^2. \quad (145)$$

(2) When $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}$, then, for every x with $w_x \geq n/2$ and $\alpha \in (0, 1]$,

$$\lim_{n\to\infty} \Pr_{\mathcal{B}}\left[p_x < \frac{\alpha}{2^n}\right] = 1, \quad (146)$$

as long as

$$q > \frac{3}{4} - \frac{1}{2(1-p)^2}. \quad (147)$$

Proof. From Theorem 2, we can write

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] = \frac{(1-r)^{w_x}(1+r)^{n-w_x}}{2^n}. \quad (148)$$

If $w_x \geq n/2$,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] = \frac{(1-r)^{w_x}(1+r)^{n-w_x}}{2^n} \leq \frac{(1-r^2)^{\frac{n}{2}}}{2^n}. \quad (149)$$

Hence, for any $\beta \in (0,1)$, there exists $\alpha \in (0,1)$ and sufficiently large n such that

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x] + \frac{\alpha}{2^n} < \frac{\beta}{2^n}, \quad (150)$$

unless $r = 0$. Now, recalling Chebyshev's inequality, we have

$$\Pr[|X - \mathbb{E}[X]| < k] \geq 1 - \frac{\mathrm{Var}(X)}{k^2} \geq 1 - \mathbb{E}[X^2] \quad (151)$$

for $k > 0$. With a choice of α and n satisfying Eq. (150), letting $k = \alpha/2^n$ in Eq. (151), we have

$$\Pr_{\mathcal{B}}\left[p_x < \frac{\alpha}{2^n} + \mathop{\mathbb{E}}_{\mathcal{B}}(p_x)\right] \geq 1 - \frac{4^n \mathop{\mathbb{E}}_{\mathcal{B}}[p_x^2]}{\alpha^2}. \quad (152)$$

By Eq. (152) and Theorem 4,

$$\Pr_{\mathcal{B}}\left[p_x < \frac{\beta}{2^n}\right] \geq 1 - \frac{(4\mu\eta)^n}{\alpha^2} \exp\left[n\frac{\nu}{\mu}\mathrm{e}^{-c(d-1)}\right]. \quad (153)$$

Since $c > 0$, when

$$d \geq \frac{\log n}{c}, \quad (154)$$

we have

$$\exp\left[n\frac{\nu}{\mu}\mathrm{e}^{-c(d-1)}\right] = \exp\left[\frac{\nu}{\mu}\mathcal{O}(1)\right] = \mathcal{O}(1). \quad (155)$$

Therefore, for such depth d, if

$$0 \leq 4\mu\eta < 1, \quad (156)$$

Eq. (144) is satisfied, because

$$\Pr_{\mathcal{B}}\left[p_x < \frac{\beta}{2^n}\right] \geq 1 - \frac{(4\mu\eta)^n}{\alpha^2}\mathcal{O}(1) \xrightarrow{n\to\infty} 1. \quad (157)$$

Thus, to satisfy Eq. (144), we have to make sure that $r \neq 0$ and $0 \leq 4\mu\eta < 1$.

We first consider the case $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}$. Since $r \neq 0$, we must have $q \neq 0$. By definition, we have

$$4\mu\eta = (1-q^2)\frac{q^2 + 3 - (1-p)^2(3-q)(1-q)}{3-(1-p)^2(3-q)(1-q)}. \quad (158)$$

Obviously, $0 \leq 4\mu\eta$. Thus, given parameters $0 \leq p, q \leq 1$, we only have to check if

$$(1-q^2)\frac{q^2 + 3 - (1-p)^2(3-q)(1-q)}{3-(1-p)^2(3-q)(1-q)} < 1. \quad (159)$$

This is equivalent to

$$0 < \left[1-(1-p)^2\right]q^2 + 4(1-p)^2 q + (2 - 3(1-p)^2). \quad (160)$$

If this condition is satisfied, we satisfy Eq. (146). Simplifying this, we obtain Eq. (145).

Next, we consider $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}$. Since $r \neq 0$, it must follow that $p \neq 1$ and $q \neq 0$. We have

$$4\mu\eta = (1-q^2(1-p)^2)\frac{q^2(1-p)^2 + 3 - (1-p)^2(3-q)(1-q)}{3-(1-p)^2(3-q)(1-q)}. \quad (161)$$

By definition, $0 \leq 4\mu\eta$. Thus, given parameters $0 \leq p, q \leq 1$, we only have to check if

$$(1-q^2(1-p)^2)\frac{q^2(1-p)^2 + 3 - (1-p)^2(3-q)(1-q)}{3-(1-p)^2(3-q)(1-q)} < 1. \quad (162)$$

Indeed, we can solve this inequality with respect to q as

$$q > \frac{3}{4} - \frac{1}{2(1-p)^2}. \quad (163)$$

If this condition is satisfied, we have Eq. (146). ■

Remark 6. Note that the restrictions on p and q, as given by Eqs. (145) and (147), are limitations of the proof technique and do not necessarily mean that the output distribution behaves any differently for values of p and q that do not satisfy these constraints.

Remark 7. Our results in Secs. VIII and IX mean that provided that the constraints in Secs. VIII and IX are satisfied, our setup is never 2^{n-1} anticoncentrated, according to Definition 14.

X. GENERALIZING TO ARBITRARY NOISE CHANNELS

In this section, we consider a general case, where the noise map $\mathcal{N}$ is characterized using parameters t_{ij} with $0 \leq i \leq 3$ and $1 \leq j \leq 3$ as

$$I_2 \to I_2 + t_{01}\sigma_x + t_{02}\sigma_y + t_{03}\sigma_z, \tag{164}$$

$$\sigma_x \to t_{11}\sigma_x + t_{12}\sigma_y + t_{13}\sigma_z, \tag{165}$$

$$\sigma_y \to t_{21}\sigma_x + t_{22}\sigma_y + t_{23}\sigma_z, \tag{166}$$

$$\sigma_z \to t_{31}\sigma_x + t_{32}\sigma_y + t_{33}\sigma_z. \tag{167}$$

Since a set $\{I_2, \sigma_x, \sigma_y, \sigma_z\}$ forms a basis for the space of single-qubit operators, an arbitrary single-qubit quantum channel can be expressed in this form. Note that, conversely, a map expressed in this form is not necessarily a quantum channel.

A. Lack of anticoncentration using collision probability

We show an extension of Theorem 1 and prove that the ensemble $\mathcal{B}$ fails to anticoncentrate.

Theorem 6. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits with the general noise channel $\mathcal{N}$. Then,

$$\mathcal{Z} \geq (1 + t_{03}^2)^n - 1. \tag{168}$$

Proof. For simplicity, let us first consider just the single-qubit case as in Theorem 1. Let

$$\rho = \mathcal{N}\left(U_1(\tilde{\rho})U_1^\dagger\right) \tag{169}$$

and let

$$\rho' = U_1(\tilde{\rho})U_1^\dagger, \tag{170}$$

where $\tilde{\rho}$ is the state just before the last block. For a single qubit, by the definition of the adjoint map,

$$\mathcal{Z} = \mathbb{E}_{U_1}\left[\mathrm{Tr}(\sigma_z \otimes \sigma_z \rho \otimes \rho)\right] \tag{171}$$

$$= \mathbb{E}_{U_1}\left[\mathrm{Tr}(\sigma_z \otimes \sigma_z\, \mathcal{N}(\rho') \otimes \mathcal{N}(\rho'))\right] \tag{172}$$

$$= \mathbb{E}_{U_1}\left[\mathrm{Tr}(\mathcal{N}^\dagger(\sigma_z) \otimes \mathcal{N}^\dagger(\sigma_z)\, \rho' \otimes \rho')\right] \tag{173}$$

$$= t_{03}^2 \mathbb{E}_{U_1}\left[\mathrm{Tr}(I_2 \otimes I_2 \rho' \otimes \rho')\right] + t_{13}^2 \mathbb{E}_{U_1}\left[\mathrm{Tr}(\sigma_x \otimes \sigma_x \rho' \otimes \rho')\right] \tag{174}$$

$$+ t_{23}^2 \mathbb{E}_{U_1}\left[\mathrm{Tr}(\sigma_y \otimes \sigma_y \rho' \otimes \rho')\right] + t_{33}^2 \mathbb{E}_{U_1} \times \left[\mathrm{Tr}(\sigma_z \otimes \sigma_z \rho' \otimes \rho')\right] \tag{175}$$

$$+ \sum_{i \neq j} t_{i3} t_{j3} \mathbb{E}_{U_1}\left[\mathrm{Tr}(\sigma_i \otimes \sigma_j \rho' \otimes \rho')\right]. \tag{176}$$

By using Eq. (59),

$$\mathbb{E}_{U_1}\left[\mathrm{Tr}(\sigma_i \otimes \sigma_j \rho' \otimes \rho')\right] = 0 \tag{177}$$

for all $i \neq j$. In addition, for $p = x, y, z$,

$$\mathbb{E}_{U_1}\left[\mathrm{Tr}(\sigma_p \otimes \sigma_p \rho' \otimes \rho')\right] = \mathbb{E}_{U_1}\left[\mathrm{Tr}(\sigma_p \rho')^2\right] \geq 0. \tag{178}$$

Therefore,

$$\mathcal{Z} \geq t_{03}^2 \mathbb{E}_{U_1}\left[\mathrm{Tr}(I_2 \otimes I_2 \rho' \otimes \rho')\right] = t_{03}^2. \tag{179}$$

With this observation, by a similar discussion as in Theorem 1, we analyze the general n-qubit case:

$$\mathcal{Z} = \mathbb{E}_{\mathcal{B}}\left[\sum_{p \in \{0,3\}^n, p \neq 0^n} \mathrm{Tr}(\sigma_p \otimes \sigma_p\, \mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|))\right] \tag{180}$$

$$\geq \sum_{p \in \{0,3\}^n, p \neq 0^n} t_{03}^{2w_p} \tag{181}$$

$$= (1 + t_{03}^2)^n - 1, \tag{182}$$

which completes the proof. ■

Theorem 6 implies that for any noise channel such that t_{03} is a nonzero constant, the ensemble $\mathcal{B}$ fails to anticoncentrate. By a similar argument to Theorem 6, one can show that for any noise channel such that t_{01} is a nonzero constant, the ensemble $\mathcal{B}$ fails to anticoncentrate when the collision probabilities are defined with respect to the Hadamard basis.

Note that for any unital channel, $t_{01} = t_{02} = t_{03} = 0$, so, by plugging into Eq. (182), $\mathcal{Z} \geq 0$ gives a vacuous bound.

B. Lack of anticoncentration using typical probabilities

We can use similar arguments to what we have done in Secs. VIII and IX to argue about the nature of the distribution. Just as there, here too the distribution has a lot of strings with very low probability weight. First, define the following quantities:

$$a = \frac{t_{01}^2 + t_{02}^2 + t_{03}^2}{3}, \tag{183}$$

$$b = \frac{1}{2} - \frac{t_{01}^2 + t_{02}^2 + t_{03}^2 + t_{11}^2 + t_{12}^2 + t_{13}^2 + t_{21}^2 + t_{22}^2 + t_{23}^2 + t_{31}^2 + t_{32}^2 + t_{33}^2}{6}, \tag{184}$$

$$c = a + 2b = 1 - \frac{t_{11}^2 + t_{12}^2 + t_{13}^2 + t_{21}^2 + t_{22}^2 + t_{23}^2 + t_{31}^2 + t_{32}^2 + t_{33}^2}{3}, \tag{185}$$

$$\mu = \frac{-t_{01}^2 - t_{02}^2 - t_{03}^2 + t_{11}^2 + t_{12}^2 + t_{13}^2 + t_{21}^2 + t_{22}^2 + t_{23}^2 + t_{31}^2 + t_{32}^2 + t_{33}^2 - 3}{4(t_{11}^2 + t_{12}^2 + t_{13}^2 + t_{21}^2 + t_{22}^2 + t_{23}^2 + t_{31}^2 + t_{32}^2 + t_{33}^2 - 3)}, \tag{186}$$

$$\nu = \frac{3(t_{01}^2 + t_{02}^2 + t_{03}^2) + t_{11}^2 + t_{12}^2 + t_{13}^2 + t_{21}^2 + t_{22}^2 + t_{23}^2 + t_{31}^2 + t_{32}^2 + t_{33}^2 - 3}{12(t_{11}^2 + t_{12}^2 + t_{13}^2 + t_{21}^2 + t_{22}^2 + t_{23}^2 + t_{31}^2 + t_{32}^2 + t_{33}^2 - 3)}, \tag{187}$$

$$\eta = \sqrt{\max\left\{(1+t_{03})^2, \frac{t_{13}^2}{2} + \frac{t_{23}^2}{2} + \frac{t_{33}^2}{2} + \frac{(1+t_{03})^2}{2}\right\}}. \tag{188}$$

1. Lack of anticoncentration at low depth

We have an analogue of Theorem 3, which shows a lack of anticoncentration for low-depth circuits.

Theorem 7. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits of depth d, with general noise channel $\mathcal{N}$. Let $x \in \{0,1\}^n$ and let w_x be the Hamming weight of x. Suppose further that

$$\left\langle |0\rangle\langle 0|, \mathcal{N}^d(|0\rangle\langle 0|)\right\rangle = \kappa + \tau\lambda^d, \tag{189}$$

with some $\kappa > \frac{1}{2}$, $\tau > 0$, and $\lambda \geq 0$. Then, if $t_{03} > 0$, there exists a constant t for every x with $w_x \geq n/2$, and $\alpha \in (0,1]$ such that

$$\lim_{n\to\infty} \Pr_{\mathcal{B}}\left[p_x < \frac{\alpha}{2^n}\right] = 1, \tag{190}$$

when $d < t\log n$.

In general, if we write

$$\mathcal{N}^n(|0\rangle\langle 0|) = \frac{1}{2}\left(I_2 + x_n\sigma_x + y_n\sigma_y + z_n\sigma_z\right), \tag{191}$$

x_n, y_n, and z_n are defined recursively as

$$x_{n+1} = t_{01} + t_{11}x_n + t_{21}y_n + t_{31}z_n, \tag{192}$$
$$y_{n+1} = t_{02} + t_{12}x_n + t_{22}y_n + t_{32}z_n, \tag{193}$$
$$z_{n+1} = t_{03} + t_{13}x_n + t_{23}y_n + t_{33}z_n, \tag{194}$$

for $n \geq 0$ with

$$x_0 = 0, \tag{195}$$
$$y_0 = 0, \tag{196}$$
$$z_0 = 1. \tag{197}$$

Thus, given noise $\mathcal{N}$ with parameters $\{t_{ij} : i = 0,1,2,3; j = 1,2,3\}$, we may check if the conditions $\kappa > \frac{1}{2}$, $\tau > 0$, and $\lambda \geq 0$ are satisfied by explicitly solving the recurrence relation introduced above.

2. Lack of anticoncentration at high depth

Theorem 8. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits of depth d, where each single-qubit noisy channel $\mathcal{N}$ is modeled as an arbitrary CPTP map with parameters $\{t_{ij}\}$. Let $d = \Omega(\log n)$. Then, for any $\alpha \in (0,1]$,

$$\lim_{n\to\infty} \Pr_{\mathcal{B}}\left[p_x < \frac{\alpha}{2^n}\right] = 1, \tag{198}$$

as long as $w_x \geq n/2$ and

$$\mu \geq 0, \tag{199}$$
$$\nu \geq 0, \tag{200}$$
$$0 < c \leq 1, \tag{201}$$
$$0 \leq 4\mu\eta < 1. \tag{202}$$

Proof. Recalling that

$$S = \frac{1}{2}\left(I_2 \otimes I_2 + \sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z\right), \tag{203}$$

we have

$$\tilde{M}_{U_1,\mathsf{N}}(I_4) = (1-a)I_4 + 2aS, \tag{204}$$

$$\tilde{M}_{U_1,\mathsf{N}}(S) = bI + (1-2b)S, \tag{205}$$

where the operators are as defined in Sec. IX. With a discussion similar to that in Sec. IX, we have

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x^2] \leq \mu^n \eta^n \exp\left[n\frac{\nu}{\mu}\mathrm{e}^{-c(d-1)}\right] \tag{206}$$

if $\mu, \nu \geq 0$ and $0 < c \leq 1$. Then, following the argument in Sec. IX, with the constraints being

$$\mu \geq 0, \tag{207}$$

$$\nu \geq 0, \tag{208}$$

$$0 < c \leq 1, \tag{209}$$

$$0 \leq 4\mu\eta < 1, \tag{210}$$

we obtain Eq. (198). ■

Remark 8. Just as in Remark 7, our calculations in Sec. X indicate that the setup in Sec. X is not 2^{n-1} anticoncentrated, provided that the constraints in Theorems 7 and 8 are satisfied.

XI. EFFECT OF THE LAST LAYER OF NOISE

Note that noises such as the amplitude-damping noise—our emblematic nonunital noise—try to push the output distribution toward a fixed state. However, a layer of random gates tries to "scramble" the distribution. In this sense, there are two opposite effects at play. This might lead one to conjecture that the behavior of the final distribution depends on which layer we end with—if ending with a layer of noiseless random gates causes anticoncentration and if ending with a layer of amplitude-damping noise causes lack of anticoncentration.

However, we provide strong evidence that this is not the case and that lack of anticoncentration occurs even if we terminate with a last layer of noiseless gates. Our results in this section are not as general as those of the other sections: they are only meant to justify our intuition. Furthermore, note that terminating with a last layer of gates is not a realistic assumption, as all known hardware has measurement noise immediately before the measurement operators, which can also be modeled as the amplitude-damping channel (see, e.g., Refs. [10,22]). Because of this, the circuit model that follows is just a toy model for analysis. What we will show is that if we "fix" a last layer of noiseless single-qubit gates, then for all choices of this layer, apart from a set of choices with measure zero, we obtain provable lack of anticoncentration, according to Definition 10.

In the following, we repeatedly use the characterization of a single-qubit gate U as

$$U(\theta,\phi) = \begin{pmatrix} \cos\theta e^{i\phi} & \sin\theta \\ -\sin\theta & \cos\theta e^{-i\phi} \end{pmatrix}. \tag{211}$$

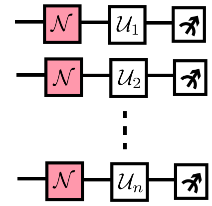


FIG. 2. The last layer of fixed single-qubit gates. In the circuit diagram, $\mathcal{N}$ is a noise channel and $U_1, U_2, \ldots, U_n$ are single-qubit gates.

Let $U_i(\theta_i, \phi_i)$ be the unitary applied to the ith qubit in the last layer.

We will consider a fixed last layer of single-qubit gates, as shown in Fig. 2, and show that for almost all choices of this layer, the output distribution exhibits a lack of anticoncentration.

Corollary 2. Let $\mathcal{B}$ be an ensemble of amplitude-damped random quantum circuits, with noise strength q. Additionally, before measurement, for every $i \in [n]$, let $U_i(\theta_i, \phi_i)$—a single-qubit noiseless gate—be applied to qubit i. Then,

$$\mathcal{Z} \geq (1 + q^2\cos^2 2\theta)^n - 1, \tag{212}$$

where

$$\theta := \mathop{\arg\min}_{\theta_j : j\in[n]} |\cos 2\theta_j|. \tag{213}$$

Proof. Note that by the action of $U_i\mathcal{N}(\cdot)U_i^\dagger$, the single-qubit identity operator I_2 will evolve as

$$I_2 \to I_2 - q\cos\phi\sin 2\theta_i\sigma_x + q\sin\phi\sin 2\theta_i\sigma_y + q\cos 2\theta_i\sigma_z. \tag{214}$$

Therefore, this last layer can be regarded as a noise map with $t_{03} = q\cos 2\theta_i$, from which the statement follows directly using Theorem 6. ■

From Corollary 2, it holds that for any value of θ, apart from those where $\cos 2\theta = 0$, the output distribution exhibits a lack of anticoncentration. The set of points for which this happens is a set of measure zero.

Note that we can also characterize the nature of the output distribution, for certain parameter regimes, and argue about lack of anticoncentration with respect to Definition 12, as shown in Appendix J 1. In fact, as we discuss later in Appendix J 2, which strings have low probability weight and which ones have higher probability weight is now determined not by the Hamming weight of the strings but by which gates have been applied in the last layer.

Qualitatively, we believe that if a layer of amplitude-damping noise is followed by a sufficiently shallow geometrically local random circuit, then the overall circuit still exhibits lack of anticoncentration. This is because amplitude damping "unscrambles" the output distribution and a shallow-depth geometrically local random circuit is not enough to counterbalance that and "scramble" it again, because shallow-depth random circuits themselves show lack of anticoncentration [6].

XII. OPEN PROBLEMS

Our paper motivates a number of open problems regarding the behavior of random circuits under nonunital noise.

A. Existence of efficient classical sampler

The most pertinent open question is whether the output distribution of random quantum circuits, with the nonunital noise models that we have studied, is classically hard to sample from. To answer this question, one potential approach is to figure out whether anticoncentration is a necessary feature in the classical sampling procedure devised in Refs. [19,45,46] or whether it just comes up as a proof artifact during analysis of the sampler and a different technique of analysis can potentially extend the authors' results to regimes for which there is no anticoncentration. If proof of classical hardness of sampling can be found, it might help in harnessing our results to design quantum advantage demonstrations with ensembles that have nonunital noise, which will complement existing quantum advantage demonstrations where the focus is on depolarizing noise [22]. To the best of our knowledge, no trivial sampling algorithm, e.g., one that samples from the fixed point of the noise channel, works for our nonunital noise models. While circuits with the depolarizing noise channel after every gate are at least inverse quasipolynomially close, in trace distance, to the maximally mixed state—the fixed point of the depolarizing channel—at sufficiently large depths, noise models such as amplitude damping are not known to show such behavior. Certain standard techniques to show this closeness, such as the data-processing inequality of quantum relative entropy [2,33,47], do not hold for the amplitude-damping channel. Note that if the nonunital noise present is only in the last layer and the rest of the circuit only has depolarizing noise, then techniques from Ref. [19] apply to classically sample from this circuit in polynomial time. One just stores an efficient truncated Fourier basis representation of the state, until the last layer of noise is encountered, and then just brute-force simulates the last layer of noise. But this trick does not work when every gate is followed by a noise channel that has a nonunital component.

Additionally, even though we obtain lack of anticoncentration, our results are different from those in Ref. [6] in the sense that the lack of anticoncentration, for our case, is not "catastrophic enough" to ensure easiness of computing output probabilities to additive precision 2^{-n}. It remains open whether that is a classically hard task.

Moreover, is there any dependence on classical simulation complexity and the rate of the noise? Is there a "percolation threshold"? That is, if the amplitude-damping noise strength is above a sufficiently high enough constant, do we obtain any phase transition in classical simulation complexity?

B. "Local" anticoncentration

Even though the global distribution does not show anticoncentration, could it still be "locally" anticoncentrated—i.e., could there be collections of bit strings such that the distribution looks flat "locally" when we consider the probability mass of only those bit strings? Depending on how much locally anticoncentrated the distribution is, one could either design new hardness conjectures or modify the existing classical samplers to work in this regime.

C. Nonunital noise at the low-noise regime

In Ref. [43], it has een observed that certain types of unital noise can be approximated by a global depolarizing noise in the low-noise regime: this is also known as the "white-noise approximation." However, it remains open whether this is also the case for nonunital noise; e.g., amplitude-damping noise. While techniques from Ref. [43] indicate that this might be true for circuits that only have a last layer of amplitude-damping noise, it remains open whether the authors' techniques could be generalized to circuits with the middle layer of amplitude-damping noise.

Rigorously proving the observation of Ref. [30]—which shows that linear cross entropy tracks fidelity in this regime—also remains open.

D. Strengthening our results

For our circuits, with respect to restricted parameter regimes, we have shown what certain strings of the output distribution look like, in Secs. IX and X and Appendix J 1. We believe that the fact that we have to restrict our parameter regimes is just an artifact of the proof technique. It remains open whether we could extend our results to a wider set of parameters.

Furthermore, our results in Secs. VI and VIII are agnostic to the choice of architecture, as long as the circuits are parallel and geometrically local. However, our calculations in Sec. IX make use of statistical models, which are known to work for 1D geometrically local circuits, but the techniques do not generalize to 2D architectures. An open question is whether architecture agnostic techniques help us in proving the bound in Sec. IX.

E. Finding a good benchmark for sampling

It is known that in the high-noise regime, the linear cross-entropy score can be classically spoofed for random circuits with the amplitude-damping noise [30]. This leaves open the question of finding a good benchmark that is not classically spoofable and that can be used to certify the results of a sampling experiment. It is also open whether the techniques of Ref. [30] extend beyond amplitude-damping noise to other noise models.

ACKNOWLEDGMENTS

S.G. thanks Changhun Oh, Yunchao Liu, Yinchen Liu, Jordan Docter, Abhinav Deshpande, and Alexander Dalzell for helpful comments on a draft of the manuscript. K.S. thanks Christa Zoufal for helpful comments on a draft of the manuscript, and Sergey Bravyi, Bryan A. O'Gorman, Oles Shtanko, and Ryan Sweke for insightful discussions. B.F. and S.G. acknowledge support from Air Force Office of Scientific Research (AFOSR) (Award No. FA9550-21-1-0008). K.K. was supported by a Mike and Ophelia Lazaridis Fellowship, the Funai Foundation, and a Perimeter Residency Doctoral Award. This material is based upon work partially supported by the National Science Foundation (NSF) under Grant No. CCF-2044923 [Faculty Early Career Development Program (CAREER)] and by the U.S. Department of Energy (DOE), Office of Science, National Quantum Information Science Research Centers as well as by DOE ""Quantum Information Science Enabled Discovery" (QuantISED) Grant No. DE-SC0020360. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

APPENDIX A: NOTE ON THE EASINESS OF SAMPLING RESULTS

In this appendix, we will sketch the easiness results of Ref. [19] and why anticoncentration is believed to be an important criterion for the current analysis techniques to go through. This is just a sketch, so we will not be too formal with the definitions and proofs.

Without loss of generality, let $\mathcal{C}$ be a unitary quantum circuit. By $p_x(\mathcal{C})$, for $x \in \{0,1\}^n$, let us denote the probability of obtaining string x in the output distribution of $\mathcal{C}|0^n\rangle$. Using techniques from Ref. [19], we can rewrite $p_x(\mathcal{C})$ in terms of Pauli paths as

$$p_x(\mathcal{C}) = \sum_s f(\mathcal{C}, x, s), \tag{A1}$$

where $f(\mathcal{C}, x, s)$ are as defined in Ref. [19]—they are path weights for each Pauli path—and s is the number of nonidentity Pauli terms for each Pauli path. Similarly, let

$$\tilde{p}_x(\mathcal{C}) = \sum_s \tilde{f}(\tilde{\mathcal{C}}, x, s) \tag{A2}$$

be the output probability, written as Pauli paths, for the noisy version of $\mathcal{C}$, denoted by $\tilde{\mathcal{C}}$. Now, to sample from the output distribution of $\tilde{\mathcal{C}}|0^n\rangle$, we consider a new distribution $\tilde{\tilde{q}}$, given by

$$\tilde{\tilde{q}}_x(\tilde{\mathcal{C}}) = \sum_{s,|s|\leq l} \tilde{f}(\tilde{\mathcal{C}}, x, s), \tag{A3}$$

where l is some threshold that we choose. Consider an ensemble $\mathcal{B}$ of such noisy circuits. Now, let us calculate the total variation distance between $\tilde{p}$ and $\tilde{\tilde{q}}$:

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[|\tilde{p} - \tilde{\tilde{q}}|_1^2\right] \leq 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \left(\tilde{p}_x(\mathcal{C}) - \tilde{\tilde{q}}_x(\tilde{\mathcal{C}})\right)^2\right] \tag{A4}$$

$$\leq 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \sum_{s,|s|>l} \tilde{f}(\tilde{\mathcal{C}}, x, s)^2\right], \tag{A5}$$

where the first line follows from the Cauchy-Schwarz inequality and the second line follows from definitions. Note that for any choice of the cutoff l,

$$2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \sum_{s,|s|>l} \tilde{f}(\tilde{\mathcal{C}}, x, s)^2\right] \tag{A6}$$

$$\leq 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \sum_{s} \tilde{f}(\tilde{\mathcal{C}}, x, s)^2\right] \tag{A7}$$

$$\leq 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \tilde{p}_x^2\right], \tag{A8}$$

where the last line follows from the orthogonality of Pauli paths in a random circuit.

1. Special case of the depolarizing channel

For the special case of the depolarizing channel,

$$\tilde{f}(\tilde{\mathcal{C}}, x, s) = (1-q)^{|s|} f(\mathcal{C}, x, s). \tag{A9}$$

So, for a choice of cutoff l, using steps similar to Eqs. (A7) and (A8), Eq. (A6) can be upper bounded with the quantity

$$(1-q)^{2l}2^n \sum_{x\in\{0,1\}^n} p_x^2. \tag{A10}$$

It is known that for sufficiently deep circuits [20], the scaled noiseless collision probability,

$$2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} p_x^2\right], \tag{A11}$$

is $\mathcal{O}(1)$. So, by appropriately choosing l, one can make the total variation distance an inverse polynomial or less.

As another direction, one could also have directly upper bounded the noisy collision probability

$$2^n \sum_{x\in\{0,1\}^n} \tilde{p}_x^2 \tag{A12}$$

for the depolarizing channel. Indeed, this is done in Ref. [6]. However, in Ref. [6], the bound is $\mathcal{O}(1)$ and not as tight as the analysis in Appendix A 1. This is necessary for inverse-polynomial closeness in total variation distance but is not sufficient. So, we still need the techniques from Appendix A 1 to prove our bound.

Remark 9. The expression in Eq. (A9) is only true for the special case of the depolarizing channel and *not* true in general. So, this analysis, where it suffices to look at the convergence of the noiseless collision probability because that can, essentially, be "factored out" of the actual expression and dealt with separately, *does not* work in general. In such a general case, directly bounding the noisy collision probability is the best that we can hope for, which may not always give us tight bounds.

2. Lack of anticoncentration implies failure of proof technique

When the noisy ensemble $\mathcal{B}$ is anticoncentrated, with respect to Definition 9, then this means that

$$2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \tilde{p}_x^2\right] = \mathcal{O}(1), \tag{A13}$$

which implies

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[|\tilde{p}-\tilde{\tilde{q}}|_1^2\right] \leq 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n}\sum_{s,|s|>l} \tilde{f}\,(\tilde{\mathcal{C}},x,s)^2\right] \leq 2^n \mathop{\mathbb{E}}_{\mathcal{B}} \times \left[\sum_{x\in\{0,1\}^n} \tilde{p}_x^2\right] = \mathcal{O}(1). \tag{A14}$$

This alone does not guarantee that the classical sampler, from Appendix A, samples from a distribution that is inverse-polynomially close, in total variation distance, to the actual distribution. However, the satisfaction of Definition 9 is a *necessary condition* for present proof techniques to go through, in the following sense: if an ensemble were to exhibit lack of anticoncentration, according to Definition 10, then the quantity

$$2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \tilde{p}_x^2\right] \tag{A15}$$

would diverge with n, and the chain of inequalities in Eq. (A14) would not hold, *no matter where we chose the cutoff* l. This does not mean that there could not be better analysis techniques that do not need anticoncentration: however, to the best of our knowledge, no such technique is known.

APPENDIX B: EFFECT OF TWIRLING

In this appendix, we explain how twirling can be used to estimate the first moment of certain expressions related to random quantum circuits.

1. Preliminaries

Let $\mathcal{N}$ be an arbitrary single-qubit noise channel. For a density matrix ρ and a single-qubit Haar-random gate U, consider the following identity [48]:

$$\begin{aligned}\Phi(\rho) &= \mathop{\mathbb{E}}_{U}\left[U^\dagger\, \mathcal{N}(U\rho U^\dagger)U\right]\\ &= (1-\lambda)\rho + \frac{\lambda}{2} I_2,\end{aligned} \tag{B1}$$

for an appropriate choice of a constant λ. Note that the expression on the left-hand side is the expression of a depolarizing channel with noise strength λ. This operation, of averaging out an arbitrary error channel to convert it into a depolarizing channel, is known in literature as "twirling," and finds use in randomized benchmarking, randomized compiling, error mitigation, etc. [48,49].

Twirling is a useful tool to gain insight into the first moments of quantities of interest for random quantum circuits. However, it is not useful to analyze second or higher-order moments. Thus, it is not a valid tool with which to analyze collision probabilities.

We will show this with four examples. A brief comment about notation: all gates shown in the figures below are Haar-random gates, either single-qubit ones or two-qubit ones depending on the context, and all noise channels shown are arbitrary single-qubit CPTP maps.

2. Computing the first moment with a last layer of noiseless gates

Consider a noisy random quantum circuit ensemble $\mathcal{B}$ and let $\mathcal{N}$ be the noise after every gate, as shown in Fig. 3.

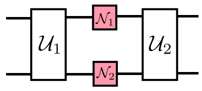


FIG. 3. A portion of the circuit drawn from the ensemble $\mathcal{B}$. U_1 and U_2 are two-qubit Haar-random gates and $\mathcal{N}_1$ and $\mathcal{N}_2$ are arbitrary noise channels.

Since the Haar measure is left and right invariant with respect to composition by a unitary, one could rewrite the entire circuit, without loss of generality, by doing the following before and after every noise channel.

Let $\rho^{(1)}$ be the marginal density matrix of the first qubit of Fig. 4 immediately before applying U_3 and let $\rho_f^{(1)}$ be the same immediately after applying $U_3^\dagger$. As is evident,

$$\rho_f^{(1)} = \mathop{\mathbb{E}}_{U_3}\left[U_3^\dagger\, \mathcal{N}_1(U_3\rho^{(1)}U_3^\dagger)\; U_3\right], \tag{B2}$$

which, from Eq. (B1), is a depolarizing channel. Every noise channel in the circuit can be equivalently modeled as a depolarizing channel in this way and calculating the first moment of the equivalent circuit suffices to calculate the first moment of the original circuit.

3. Computing the first moment with a last layer of noisy gates

If the circuit terminates with a last layer of noise, before the measurement layer, as illustrated in Fig. 5, then there is no way to twirl the last layer of noise, using the technique in Sec. B 2. Hence, these circuits cannot be twirled.

4. Computing the expected linear cross-entropy score

The linear cross entropy is given by the following quantity:

$$\begin{aligned}\text{XEB} &= \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} p_{\text{ideal}}(x)p_{\text{noisy}}(x)\right] \\ &= 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{\text{ideal}}(0^n)p_{\text{noisy}}(0^n)\right],\end{aligned} \tag{B3}$$

where $p_{\text{ideal}}(x)$ is the probability of seeing bit string x in the output distribution of a circuit drawn from $\mathcal{B}$ with all

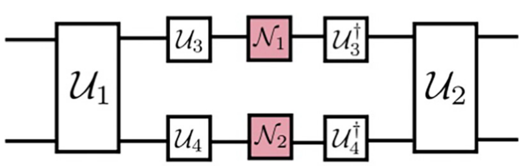


FIG. 4. An equivalent expression of the circuit shown in Fig. 3. In the circuit diagram, U_1 and U_2 are two-qubit Haar-random gates, U_3 and U_4 are single-qubit Haar-random gates, and $\mathcal{N}_1$ and $\mathcal{N}_2$ are arbitrary noise channels.

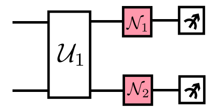


FIG. 5. If the circuit terminates with a last layer of noiseless gates, then there is no way to "sandwich" the last layer of noise between a Haar-random single-qubit unitary and its adjoint. So, twirling does not work.

the noise channels removed and $p_{\text{noisy}}(x)$ is the probability of seeing bit string x in the noisy output distribution. We have assumed that there is no last layer of noise. Note that Eq. (B3) can be written as

$$\text{XEB} = 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\text{Tr}\left(|0^n\rangle\langle 0^n| \otimes |0^n\rangle\langle 0^n|\rho \otimes \tilde{\rho}\right)\right], \tag{B4}$$

where ρ is the density matrix corresponding to the final state of the circuit with all the noise channels removed and $\tilde{\rho}$ is the density matrix corresponding to the circuit with noise. Twirling can be used to estimate this quantity, as is shown in Fig. 6.

5. Computing the second moment

For a noisy ensemble $\mathcal{B}$, let the task be to compute

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} p_x^2\right] = 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{0^n}^2\right], \tag{B5}$$

where we have assumed that there is no last layer of noise. Note that Eq. (B5) can be written as

$$2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[p_{0^n}^2\right] = 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\text{Tr}\left(|0^n\rangle\langle 0^n| \otimes |0^n\rangle\langle 0^n|\; \tilde{\rho} \otimes \tilde{\rho}\right)\right], \tag{B6}$$

where $\tilde{\rho}$ is the density matrix corresponding to the final state of the noisy circuit. Here, we cannot hope to "twirl"

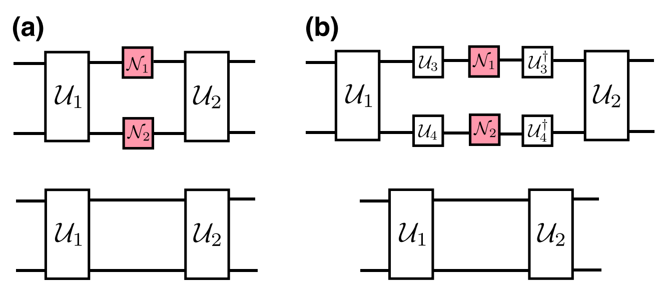


FIG. 6. (a) A portion of the circuit that prepares $\tilde{\rho}$ and a portion of the circuit that prepares ρ. (b) Note that $UU^* = \mathbb{I}$, where $\mathbb{I}$ is the single-qubit identity operator and U is any unitary operator. This means that there is no dependence of U_3 and U_4 on the "noiseless copy" of the circuit. Because of that, the quantity that we are evaluating can be simplified with Eq. (B1), just as we did for first-moment quantities previously.

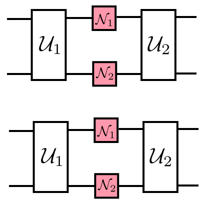


FIG. 7. Two copies of a portion of the circuit that prepares $\tilde{\rho}$.

the noise because both copies of the state are noisy, as illustrated in Fig. 7. So, checking whether the distribution anticoncentrates or not cannot be done by twirling. In Fig. 7, if we were to twirl $\mathcal{N}_1$ in a similar way as in Sec. B 4, we would need to evaluate

$$\mathbb{E}_U\big[U^\dagger \otimes U^\dagger\, \mathcal{N}_1^{\otimes 2}(U \otimes U\rho \otimes \rho U^\dagger \otimes U^\dagger)U \otimes U\big], \quad \text{(B7)}$$

where U is a single-qubit Haar-random unitary, which no longer simplifies to the depolarizing channel.

APPENDIX C: ANTICONCENTRATION AND CLOSENESS TO THE UNIFORM DISTRIBUTION

In this appendix, we show the relation between anticoncentration, according to Definition 9, and closeness to the uniform distribution. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits and let $\mathcal{C}$ be a random circuit from $\mathcal{B}$. Let

$$\rho = \mathcal{C}(|0^n\rangle\langle 0^n|). \quad \text{(C1)}$$

For $x \in \{0,1\}^n$, let

$$\mathbb{E}_{\mathcal{B}}[p_x] = \mathbb{E}_{\mathcal{B}}[\mathrm{Tr}(|x\rangle\langle x|\rho)]. \quad \text{(C2)}$$

Note that

$$\mathbb{E}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n}\left|p_x - \frac{1}{2^n}\right|^2\right] = \mathbb{E}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} p_x^2\right] - 2\frac{1}{2^n} + \frac{1}{2^n} \quad \text{(C3)}$$

$$= \mathbb{E}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} p_x^2\right] - \frac{1}{2^n}. \quad \text{(C4)}$$

Hence, if

$$\mathbb{E}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n}\left|p_x - \frac{1}{2^n}\right|^2\right] = \mathcal{O}(2^{-n}), \quad \text{(C5)}$$

then

$$\mathbb{E}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} p_x^2\right] = \mathcal{O}(2^{-n}), \quad \text{(C6)}$$

which means that Definition 9 is satisfied—and vice versa.

Now, let us analyze when Eq. (C5) is satisfied. Let Δ_n be the n-qubit completely dephasing channel with respect to the computational basis; i.e., for any n-qubit state σ,

$$\Delta_n(\sigma) := \sum_{x\in\{0,1\}^n} \langle x|\,\sigma\,|x\rangle\,|x\rangle\langle x|\,. \quad \text{(C7)}$$

Now, assume that the following equation holds:

$$\mathbb{E}_{\mathcal{B}}\left\|\rho - \frac{I_{2^n}}{2^n}\right\|_1^2 = \mathcal{O}(2^{-n}). \quad \text{(C8)}$$

Then, from monotonicity of trace distance, we obtain that

$$\mathbb{E}_{\mathcal{B}}\left\|\Delta_n(\rho) - \frac{I_{2^n}}{2^n}\right\|_1^2 = \mathcal{O}(2^{-n}). \quad \text{(C9)}$$

We have used the fact that $\Delta_n(I_{2^n}) = I_{2^n}$. Then, since the 2-norm is smaller than the trace norm, from Eq. (C9) we obtain

$$\mathbb{E}_{\mathcal{B}}\left\|\Delta_n(\rho) - \frac{I_{2^n}}{2^n}\right\|_2^2 = \mathcal{O}(2^{-n}), \quad \text{(C10)}$$

leading to Eq. (C5).

APPENDIX D: EASINESS OF COMPUTING EXPECTATION VALUES

In this appendix, we will, very broadly, sketch the argument of Ref. [36] about computing the expectation value of certain observables, for random quantum circuits with depolarizing noise. The ideas are extremely similar to that of Appendix A but the argument does not require anticoncentration. To start with, note that for the depolarizing channel,

$$\tilde{f}(\tilde{\mathcal{C}}, x, s) = (1-q)^{|s|} f(\mathcal{C}, x, s), \quad \text{(D1)}$$

where f is a Pauli path of the noiseless circuit $\mathcal{C}$, $\tilde{f}$ is a Pauli path of the noisy circuit $\tilde{\mathcal{C}}$, $x \in \{0,1\}^n$, and $s \in \{0,1\}^{2n}$ is the number of nonidentity Pauli terms in the path. In Ref. [36], it is shown how the expectation value of any observable can be written as

$$\sum_s \tilde{f}(\tilde{\mathcal{C}}, x, s). \quad \text{(D2)}$$

Now, using Eq. (D1), any path such that $s = \omega(\log n)$ is at least inverse-superpolynomially suppressed and only polynomially many paths—those for which $s = \mathcal{O}(\log n)$—have at least $1/\mathrm{poly}(n)$ weight. Then, just by classically estimating those paths, which can be done in classical polynomial time, we obtain an estimate of the expectation value that is approximately $1/\mathrm{poly}(n)$ close to the original expectation value. Here, we emphasize that this technique does not need anticoncentration. Note that Eq. (D1) is a special property of the depolarizing channel: it is not evident whether the strategy described works for other noise channels.

APPENDIX E: DERIVATION OF EXPRESSION OF COLLISION PROBABILITY USING PAULI OPERATORS

In this appendix, we show the expression given in Eq. (39):

$$\mathcal{Z} = \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{p\in\{0,3\}^n, p\neq 0^n} \mathrm{Tr}(\sigma_p\, \mathcal{C}(|0\rangle\langle 0|))^2\right], \tag{E1}$$

where $\mathcal{Z}$ is the collision probability, defined as

$$\mathcal{Z} = 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \mathrm{Tr}(|x\rangle\langle x|\, \mathcal{C}(|0\rangle\langle 0|))^2\right] - 1. \tag{E2}$$

First, we rewrite $|x\rangle\langle x|$ using Pauli operators. Let us write $|x\rangle\langle x| = |x_1\rangle\langle x_1| \otimes |x_2\rangle\langle x_2| \otimes \cdots \otimes |x_n\rangle\langle x_n|$, where $x_i \in 0, 1$ for $i = 1, 2, \ldots, n$. Since

$$|0\rangle\langle 0| = \frac{I_2 + \sigma_z}{2}, \tag{E3}$$

$$|1\rangle\langle 1| = \frac{I_2 - \sigma_z}{2}, \tag{E4}$$

we can write

$$|x_i\rangle\,\langle x_i| = \frac{1}{2}\sum_{p_i\in\{0,3\}} (-1)^{b_{p_i}^{(x_i)}} \sigma_{p_i}, \tag{E5}$$

where $b_i^{(x_i)}$ is a binary digit defined by

$$b_i^{(x_i)} := \begin{cases} 0, & x_i = 0 \quad \text{or} \quad p_i = 0, \\ 1, & x_1 = 1 \quad \text{and} \quad p_i = 3. \end{cases} \tag{E6}$$

Hence,

$$|x\rangle\langle x| = \bigotimes_{i=1}^{n}\left(\frac{1}{2}\sum_{p_i\in\{0,3\}} (-1)^{b_{p_i}^{(x_i)}} \sigma_{p_i}\right) = \frac{1}{2^n}\sum_{p\in\{0,3\}^n} (-1)^{b_p^{(x)}} \sigma_p, \tag{E7}$$

where

$$b_p^{(x)} := \sum_{i=1}^{n} b_{p_i}^{(x_i)}. \tag{E8}$$

Next, using this relation, we have

$$\mathcal{Z} = 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \mathrm{Tr}(|x\rangle\langle x|\, \mathcal{C}(|0\rangle\langle 0|))^2\right] - 1 \tag{E9}$$

$$= 2^n \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n} \mathrm{Tr}((|x\rangle\langle x| \otimes |x\rangle\langle x|)(\mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|)))\right] - 1 \tag{E10}$$

$$= \frac{1}{2^n}\mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{x\in\{0,1\}^n}\sum_{p,p'\in\{0,3\}^n} (-1)^{b_p^{(x)}+b_{p'}^{(x)}} \mathrm{Tr}((\sigma_p \otimes \sigma_{p'})\, (\mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|)))\right] - 1 \tag{E11}$$

$$= \frac{1}{2^n}\mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{p,p'\in\{0,3\}^n}\left(\sum_{x\in\{0,1\}^n} (-1)^{b_p^{(x)}+b_{p'}^{(x)}}\right) \mathrm{Tr}((\sigma_p \otimes \sigma_{p'})(\mathcal{C}(|0\rangle\langle 0|) \otimes \mathcal{C}(|0\rangle\langle 0|)))\right] - 1. \tag{E12}$$

Now, we evaluate the term $\sum_{x\in\{0,1\}^n} (-1)^{b_p^{(x)}+b_{p'}^{(x)}}$. By the definition given in Eq. (E8), we have

$$\sum_{x\in\{0,1\}^n} (-1)^{b_p^{(x)}+b_{p'}^{(x)}} = \prod_{i=1}^{n}\left(\sum_{x_i\in\{0,1\}} (-1)^{b_{p_i}^{(x_i)}+b_{p_i'}^{(x_i)}}\right). \tag{E13}$$

Due to the definition given in Eq. (E6),

$$\sum_{x_i\in\{0,1\}} (-1)^{b^{(x_i)}_{p_i}+b^{(x_i)}_{p'_i}} = \begin{cases} 2, & p_i = p'_i, \\ 0, & p_i \neq p'_i, \end{cases} \tag{E14}$$

and thus we obtain

$$\sum_{x\in\{0,1\}^n} (-1)^{b^{(x)}_{p}+b^{(x)}_{p'}} = \begin{cases} 2^n, & p = p', \\ 0, & p \neq p'. \end{cases} \tag{E15}$$

Therefore,

$$\mathcal{Z} = \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{p\in\{0,3\}^n} \mathrm{Tr}((\sigma_p\otimes\sigma_p)\,(\mathcal{C}(|0\rangle\langle 0|)\otimes\mathcal{C}(|0\rangle\langle 0|)))\right] - 1 = \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{p\in\{0,3\}^n} \mathrm{Tr}(\sigma_p\,\mathcal{C}(|0\rangle\langle 0|))^2\right] - 1. \tag{E16}$$

Since

$$\mathrm{Tr}(I_2^{\otimes n}\,\mathcal{C}(|0\rangle\langle 0|)) = 1, \tag{E17}$$

we finally have

$$\mathcal{Z} = \mathop{\mathbb{E}}_{\mathcal{B}}\left[\sum_{p\in\{0,3\}^n, p\neq 0^n} \mathrm{Tr}(\sigma_p\,\mathcal{C}(|0\rangle\langle 0|))^2\right] \tag{E18}$$

as desired.

APPENDIX F: DISCUSSION ON THE EFFECT OF NOISE IN NOISY RANDOM CIRCUITS

In this appendix, we show that for any given noise $\mathcal{N}$, we may express the action of $\tilde{M}_{U_1,\mathsf{N}}$ on I_4 and S as

$$\tilde{M}_{U_1,\mathsf{N}}(I_4) = (1-a)I_4 + 2aS, \tag{F1}$$

$$\tilde{M}_{U_1,\mathsf{N}}(S) = bI + (1-2b)S, \tag{F2}$$

using some a and b. The action of M_{U_1} can be given as (see Example 7.25 of Ref. [50])

$$M_{U_1}(X) = \frac{\langle X,\Pi_{\mathrm{sym}}\rangle}{3}\Pi_{\mathrm{sym}} + \langle X,\Pi_{\mathrm{antisym}}\rangle\Pi_{\mathrm{antisym}}, \tag{F3}$$

where

$$\Pi_{\mathrm{sym}} := \frac{I_4+S}{2}, \tag{F4}$$

$$\Pi_{\mathrm{antisym}} := \frac{I_4-S}{2} \tag{F5}$$

are the projection operators onto the symmetric and antisymmetric subspaces, respectively. Suppose that X is given as

$$X = \sum_{i,j,k,l=0,1} x_{ijkl}\,|ij\rangle\langle kl|\,. \tag{F6}$$

A simple calculation leads to

$$\langle X,\Pi_{\mathrm{sym}}\rangle = x_{0000} + x_{1111} + \frac{x_{0101}+x_{1010}+x_{0110}+x_{1001}}{2}, \tag{F7}$$

$$\langle X,\Pi_{\mathrm{antisym}}\rangle = \frac{x_{0101}+x_{1010}-x_{0110}-x_{1001}}{2}. \tag{F8}$$

Thus,

$$M_{U_1}(X) = \left(\frac{x_{0000}}{6} + \frac{x_{1111}}{6} + \frac{x_{0101}+x_{1010}}{3} - \frac{x_{0110}+x_{1001}}{6}\right) I_4 + \left(\frac{x_{0000}}{6} + \frac{x_{1111}}{6} - \frac{x_{0101}+x_{1010}}{6} + \frac{x_{0110}+x_{1001}}{3}\right) S. \tag{F9}$$

When

$$X = \mathsf{N} \circ M_{U_1}[I_4] = \mathsf{N}[I_4], \tag{F10}$$

since $\mathrm{Tr}[X] = x_{0000} + x_{0101} + x_{1010} + x_{1111} = \mathrm{Tr}[I_4] = 4$, by setting

$$a = \frac{x_{0000}}{12} + \frac{x_{1111}}{12} - \frac{x_{0101} + x_{1010}}{12} + \frac{x_{0110} + x_{1001}}{6}, \tag{F11}$$

we have

$$\tilde{M}_{U_1,\mathsf{N}}[I_4] = (1-a)I_4 + 2aS. \tag{F12}$$

Similarly, when

$$X = \mathsf{N} \circ M_{U_1}[S] = \mathsf{N}[S], \tag{F13}$$

since $\mathrm{Tr}[X] = x_{0000} + x_{0101} + x_{1010} + x_{1111} = \mathrm{Tr}[S] = 2$, by setting

$$b = \frac{x_{0000}}{6} + \frac{x_{1111}}{6} + \frac{x_{0101} + x_{1010}}{3} - \frac{x_{0110} + x_{1001}}{6}, \tag{F14}$$

we have

$$\tilde{M}_{U_1,\mathsf{N}}[S] = bI + (1-2b)S. \tag{F15}$$

We see the cases of $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}$ and $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}$ as illustrative examples. When $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}$, we have

$$\mathsf{N}[I_4] = \begin{pmatrix} (1+q)^2 & 0 & 0 & 0 \\ 0 & 1-q^2 & 0 & 0 \\ 0 & 0 & 1-q^2 & 0 \\ 0 & 0 & 0 & (1-q)^2 \end{pmatrix}, \tag{F16}$$

$$\mathsf{N}[S] = \begin{pmatrix} (1-q)^2\left(\frac{p^2}{2} - p + 1\right) + 2q & 0 & 0 & 0 \\ 0 & \frac{1-q^2-(1-q)^2(1-p)^2}{2} & (1-q)(1-p)^2 & 0 \\ 0 & (1-q)(1-p)^2 & \frac{1-q^2-(1-q)^2(1-p)^2}{2} & 0 \\ 0 & 0 & 0 & (1-q)^2\left(\frac{p^2}{2} - p + 1\right) \end{pmatrix}. \tag{F17}$$

Therefore, we have

$$a = \frac{q^2}{3}, \qquad b = \frac{1}{2} - \frac{q^2}{6} - \frac{1}{6}(1-p)^2(1-q)(3-q). \tag{F18}$$

Similarly, when $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}$,

$$\mathsf{N}[I_4] = \begin{pmatrix} (1+(1-p)q)^2 & 0 & 0 & 0 \\ 0 & 1-(1-p)^2q^2 & 0 & 0 \\ 0 & 0 & 1-(1-p)^2q^2 & 0 \\ 0 & 0 & 0 & (1-(1-p)q)^2 \end{pmatrix}, \tag{F19}$$

$$\mathsf{N}[S] = \begin{pmatrix} \begin{matrix} q^2(1-p)^2 + 1 + \frac{p^2}{2} \\ -p(1+(1-p)q) \end{matrix} & 0 & 0 & 0 \\ 0 & q(1-q)(1-p)^2 - \frac{p^2}{2} + p & (1-q)(1-p)^2 & 0 \\ 0 & (1-q)(1-p)^2 & q(1-q)(1-p)^2 - \frac{p^2}{2} + p & 0 \\ 0 & 0 & 0 & \begin{matrix} (q(1-p)-1)^2 + \frac{p^2}{2} \\ -p(1+(1-p)q) \end{matrix} \end{pmatrix}. \tag{F20}$$

Therefore,

$$a = \frac{q^2(1-p)^2}{3}, \quad b = \frac{1}{2} - \frac{q^2(1-p)^2}{6} - \frac{1}{6}(1-p)^2(1-q)(3-q). \tag{F21}$$

APPENDIX G: DISCUSSION ON CONDITIONAL PROBABILITIES

In this appendix, we will evaluate the first moment of conditional probability.

Lemma 2. Let $\mathcal{B}$ be an ensemble of noisy random quantum circuits with noise channel $\mathcal{N}$. Let $i \in \{1, 2, \ldots, n\}$ and let J_i be a subset of $\{1, 2, \ldots, n\}$ that does not contain i. Then,

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\Pr\left(X_i = x_i | \{X_j = x_j\}_{j \in J_i}\right)\right] = \frac{1}{2}\left[\langle x_i | \mathcal{N}(I_2) | x_i \rangle\right]. \tag{G1}$$

Proof. We only give proof for

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\Pr\left(X_i = x_i | \{X_j = x_j\}_{j \in \{1,2,\ldots,i-1\}}\right)\right] = \frac{1}{2}\left[\langle x_i | \mathcal{N}(I_2) | x_i \rangle\right]. \tag{G2}$$

The other cases follow similarly.

Let $\mathcal{C}$ be the given quantum circuit. We may assume that $\mathcal{C}$ can be written in the following form:

$$\mathcal{C} = \mathcal{N}^{\otimes n} \circ \left(\bigotimes_{i=1}^{n} \mathcal{U}_i\right) \circ \tilde{\mathcal{C}}, \tag{G3}$$

where $\mathcal{N}$ is the noise channel, $\mathcal{U}_i$ is a single-qubit Haar-random unitary channel, and $\tilde{\mathcal{C}}$ is the circuit without the last layer of noise.

Then, the expectation over $\mathcal{B}$ is decomposed as

$$\mathop{\mathbb{E}}_{\mathcal{B}} = \mathop{\mathbb{E}}_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_n \\ \sim \text{Haar}}} \mathop{\mathbb{E}}_{\mathcal{B}'}, \tag{G4}$$

where Haar represents the set of single-qubit Haar unitaries and $\mathcal{B}'$ is the set of noisy random quantum circuits without the last layer of noise. Then,

$$\Pr\left(X_i = x_i | \{X_j = x_j\}_{j \in \{1,2,\ldots,i-1\}}\right) \tag{G5}$$

$$= \frac{\Pr\left(X_1 = x_1, \ldots, X_i = x_i\right)}{\Pr\left(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}\right))} \tag{G6}$$

$$= \frac{\langle x_1 \cdots x_{i-1} x_i | \mathcal{N}^{\otimes i} \circ \left(\bigotimes_{j=1}^{i} \mathcal{U}_j\right)\left(\rho_{\tilde{\mathcal{C}}}^{(i)}\right) | x_1 \cdots x_{i-1} x_i \rangle}{\sum_{y \in \{0,1\}} \langle x_1 \cdots x_{i-1} y | \mathcal{N}^{\otimes i} \circ \left(\bigotimes_{j=1}^{i} \mathcal{U}_j\right)\left(\rho_{\tilde{\mathcal{C}}}^{(i)}\right) | x_1 \cdots x_{i-1} y \rangle}, \tag{G7}$$

where

$$\rho_{\tilde{\mathcal{C}}}^{(i)} = \mathrm{Tr}_{i+1\cdots n}\left[\left(\mathcal{I}^{\otimes i} \otimes \mathcal{N}^{\otimes(n-i)}\right) \circ \left(\mathcal{I}^{\otimes i} \otimes \bigotimes_{j=i+1}^{n} \mathcal{U}_j\right) \circ \tilde{\mathcal{C}}\left(|0^n\rangle\langle 0^n|\right)\right]. \tag{G8}$$

Now, let us write

$$\left(\mathcal{N}^{\otimes(i-1)} \otimes \mathcal{I}\right) \circ \left(\bigotimes_{j=1}^{i-1} \mathcal{U}_j \otimes \mathcal{I}\right)\left(\rho_{\tilde{\mathcal{C}}}^{(i)}\right) = \sum_{y_1 \cdots y_{i-1}, z_1 \cdots z_{i-1} \in \{0,1\}^{i-1}} \eta_{(z_1 \cdots z_{i-1})}^{(y_1 \cdots y_{i-1})} |y_1 \cdots y_{i-1}\rangle \langle z_1 \cdots z_{i-1}| \otimes \sigma_{(z_1 \cdots z_{i-1})}^{(y_1 \cdots y_{i-1})}, \tag{G9}$$

where the $\eta^{(y_1\cdots y_{i-1})}_{(z_1\cdots z_{i-1})}$ are appropriately defined coefficients so that $\sigma^{(y_1\cdots y_{i-1})}_{(z_1\cdots z_{i-1})}$ will be density operators when $y_1 \cdots y_{i-1} = z_1 \cdots z_{i-1}$. Then, we may write

$$\Pr\big(X_i = x_i | \{X_j = x_j\}_{j\in\{1,2,\ldots,i-1\}}\big) \tag{G10}$$

$$= \frac{\eta^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})} \langle x_i| (\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)|x_i\rangle}{\eta^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\left(\langle 0| (\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)|0\rangle + \langle 1| (\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)|1\rangle\right)} \tag{G11}$$

$$= \frac{\eta^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})} \langle x_i| (\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)|x_i\rangle}{\eta^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\mathrm{Tr}\left[(\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)\right]} \tag{G12}$$

$$= \frac{\eta^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})} \langle x_i| (\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)|x_i\rangle}{\eta^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}} \tag{G13}$$

$$= \langle x_i| (\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)|x_i\rangle . \tag{G14}$$

Here, we have used the fact that $\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}$ is a single-qubit density operator. Therefore,

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\Pr\big(X_i = x_i | \{X_j = x_j\}_{j\in\{1,2,\ldots,i-1\}}\big)\right] \tag{G15}$$

$$= \mathop{\mathbb{E}}_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_n\\ \sim\text{Haar}}}\mathop{\mathbb{E}}_{\mathcal{B}'}\left[\Pr\big(X_i = x_i | \{X_j = x_j\}_{j\in\{1,2,\ldots,i-1\}}\big)\right] \tag{G16}$$

$$= \mathop{\mathbb{E}}_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_n\\ \sim\text{Haar}}}\mathop{\mathbb{E}}_{\mathcal{B}'}\left[\langle x_i| (\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)|x_i\rangle\right] \tag{G17}$$

$$= \mathop{\mathbb{E}}_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_{i-1}\\ \sim\text{Haar}}}\mathop{\mathbb{E}}_{\substack{\mathcal{U}_{i+1},\ldots,\mathcal{U}_n\\ \sim\text{Haar}}}\mathop{\mathbb{E}}_{\mathcal{B}'}\left[\langle x_i| \mathop{\mathbb{E}}_{\mathcal{U}_i\sim\text{Haar}}\left[(\mathcal{N}\circ\mathcal{U}_i)\left(\sigma^{(x_1\cdots x_{i-1})}_{(x_1\cdots x_{i-1})}\right)\right]|x_i\rangle\right] \tag{G18}$$

$$= \mathop{\mathbb{E}}_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_{i-1}\\ \sim\text{Haar}}}\mathop{\mathbb{E}}_{\substack{\mathcal{U}_{i+1},\ldots,\mathcal{U}_n\\ \sim\text{Haar}}}\mathop{\mathbb{E}}_{\mathcal{B}'}\left[\langle x_i| \mathcal{N}\left(\frac{I_2}{2}\right)|x_i\rangle\right] \tag{G19}$$

$$= \frac{1}{2}\left[\langle x_i| \mathcal{N}(I_2)|x_i\rangle\right], \tag{G20}$$

which we aimed to show. ■

APPENDIX H: COMPUTATION OF THE SECOND MOMENT OF A SPECIAL OBSERVABLE FOR NOISY RANDOM QUANTUM CIRCUITS

In this appendix, we give a lower bound on the second moment of the single-qubit Pauli-z operator and describe applications of the lower bound to our setup.

Lemma 3. Consider the single-qubit channel $\mathcal{N}$ applied d times to $|0\rangle\langle 0|$. Denote the resultant channel by $\mathcal{N}^d := \underbrace{\mathcal{N}\circ\mathcal{N}\circ\cdots\circ\mathcal{N}}_{d\text{ times}}$. Suppose that

$$\left\langle |0\rangle\langle 0|, \mathcal{N}^d(|0\rangle\langle 0|)\right\rangle = \kappa + \tau\lambda^d, \tag{H1}$$

with some $\kappa \geq \frac{1}{2}$, $\tau, \lambda \geq 0$. Then,

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\langle Z_i\rangle^2\right] \geq \frac{4\left(\kappa - \frac{1}{2} + \tau\lambda^d\right)^2}{30^d}, \tag{H2}$$

where $\mathcal{B}$ is an ensemble of noisy random quantum circuits of depth d, in which the noise is modeled by $\mathcal{N}$, and $\langle Z_i \rangle := 2p_i - 1$, with p_i being the marginal probability of obtaining outcome i for a single qubit.

Proof. The proof of this lemma proceeds similarly to Theorem 1 of Ref. [6]. Indeed, in Theorem 1 of Ref. [6], a special case of this lemma is shown with $\kappa = \frac{1}{2}, \tau = \frac{1}{2}$, and $\lambda = (1-p)$, where p is the strength of the depolarizing channel. ■

We may apply the above lemma to the noise models that we consider:

(1) When $\mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}$, a simple calculation leads to

$$\kappa = \frac{q + \frac{p}{2}(1-q)}{1-(1-p)(1-q)} \tag{H3}$$

$$\tau = 1 - \frac{q + \frac{p}{2}(1-q)}{1-(1-p)(1-q)} \tag{H4}$$

$$\lambda = (1-p)(1-q). \tag{H5}$$

It is obvious that $\lambda \geq 0$. To show $\kappa \geq \frac{1}{2}$ and $\tau \geq 0$, we show that

$$\frac{1}{2} \leq \frac{q + \frac{p}{2}(1-q)}{1-(1-p)(1-q)} \leq 1. \tag{H6}$$

The left inequality can be shown as

$$\frac{q + \frac{p}{2}(1-q)}{1-(1-p)(1-q)} = \frac{q + \frac{p}{2}(1-q)}{q + p(1-q)} \tag{H7}$$

$$\geq \frac{\frac{q}{2} + \frac{p}{2}(1-q)}{q + p(1-q)} \tag{H8}$$

$$= \frac{1}{2}. \tag{H9}$$

Similarly, the right inequality can be shown as

$$\frac{q + \frac{p}{2}(1-q)}{1-(1-p)(1-q)} = \frac{q + \frac{p}{2}(1-q)}{q + p(1-q)} \tag{H10}$$

$$\leq \frac{q + p(1-q)}{q + p(1-q)} \tag{H11}$$

$$= 1. \tag{H12}$$

(2) When $\mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}$, we have

$$\kappa = \frac{\frac{p}{2} + (1-p)q}{1-(1-p)(1-q)} \tag{H13}$$

$$\tau = 1 - \frac{\frac{p}{2} + (1-p)q}{1-(1-p)(1-q)} \tag{H14}$$

$$\lambda = (1-p)(1-q). \tag{H15}$$

With a similar argument, we have

$$\frac{1}{2} \leq \frac{\frac{p}{2} + (1-p)q}{1-(1-p)(1-q)} \leq 1, \tag{H16}$$

so $\kappa \geq \frac{1}{2}$ and $\tau, \lambda \geq 0$.

This lemma implies that if a given noise channel $\mathcal{N}$ satisfies

$$\kappa - \frac{1}{2} + \tau\lambda^d > 0, \tag{H17}$$

then there exist positive numbers $a, b > 0$ such that

$$\mathop{\mathbb{E}}_{\mathcal{B}}\left[\langle Z_i \rangle^2\right] \geq b e^{-ad}. \tag{H18}$$

APPENDIX I: SECOND MOMENT OF OUTPUT PROBABILITIES

In this appendix, we provide proof of Theorem 4, which gives an upper bound on the second moment of a noisy random-circuit ensemble. In Appendix I 1, we show technical lemmas analyzing the effect of randomized noise, which are used in the proof. In Appendix I 2, we show our technique to evaluate the second moment by considering a slightly simplified circuit. In Appendix I 3, we finally give a full proof of Theorem 4.

1. Technical lemmas characterizing the effect of noise

In this section, we show intermediate lemmas used in the proof of Theorem 4. First, we characterize the effect of noise randomized by 2-copy single-qubit random unitaries on the input state.

Lemma 4. Let $\mathcal{N}$ be a single-qubit noise channel and let $\mathsf{N} = \mathcal{N} \otimes \mathcal{N}$ be two copies of a single-qubit channel $\mathcal{N}$. Define a two-qubit operator

$$\tilde{M}_{U_1,\mathsf{N}} = M_{U_1} \circ \mathsf{N} \circ M_{U_1} \tag{I1}$$

with

$$M_{U_1}[\rho] = \mathop{\mathbb{E}}_{U_1 \sim \mathcal{U}_{\mathrm{Haar}}}\left[U_1^{\otimes 2} \rho U_1^{\dagger \otimes 2}\right]. \tag{I2}$$

Suppose that

$$\tilde{M}_{U_1,\mathsf{N}}(I_4) = (1-a)I_4 + 2aS, \tag{I3}$$

$$\tilde{M}_{U_1,\mathsf{N}}(S) = bI_4 + (1-2b)S \tag{I4}$$

with $a > 0$ and $b > 0$, where I_4 is the two-qubit identity operator and S is the two-qubit SWAP gate. Then,

$$M_{U_1} \circ \underbrace{(\mathsf{N} \circ M_{U_1}) \circ (\mathsf{N} \circ M_{U_1}) \circ \cdots \circ (\mathsf{N} \circ M_{U_1})}_{m \text{ times}} \left[|0\rangle\langle 0|^{\otimes 2}\right]$$
$$= \frac{1}{10}(2I+S) + \left[\frac{-2a+b}{10(a+2b)} + (1-a-2b)^m \left(-\frac{1}{30} - \frac{-2a+b}{10(a+2b)}\right)\right](I_4 - 2S). \tag{I5}$$

Remark 10. As discussed in Appendix F, without loss of generality, for any noise $\mathcal{N}$, the effect of noise versus I_4 and S in the noisy random circuit can be expressed as in Eqs. (I3) and (I4).

Proof. First, note that by properties of the Haar measure,

$$M_{U_1}[\rho] = M_{U_1} \circ M_{U_1}[\rho]. \tag{I6}$$

Hence,

$$M_{U_1} \circ \underbrace{(\mathsf{N} \circ M_{U_1}) \circ (\mathsf{N} \circ M_{U_1}) \circ \cdots \circ (\mathsf{N} \circ M_{U_1}))}_{m \text{ times}} \times \left[|0\rangle\langle 0|^{\otimes 2}\right] \tag{I7}$$

$$= \underbrace{(M_{U_1} \circ \mathsf{N} \circ M_{U_1}) \circ \cdots \circ (M_{U_1} \circ \mathsf{N} \circ M_{U_1})}_{m \text{ times}} \times \left[|0\rangle\langle 0|^{\otimes 2}\right] \tag{I8}$$

$$= \underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \cdots \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}} \circ M_{U_1} \left[|0\rangle\langle 0|^{\otimes 2}\right] \tag{I9}$$

$$= \underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \cdots \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}} \left[\frac{1}{6}(I_4+S)\right]. \tag{I10}$$

In the fourth line, we have used the fact that

$$M_{U_1}\left[|0\rangle\langle 0|^{\otimes 2}\right] = \frac{1}{6}(I_4+S). \tag{I11}$$

Note that

$$\frac{1}{6}(I_4+S) = \frac{1}{10}(2I+S) - \frac{1}{30}(I_4-2S). \tag{I12}$$

It can be verified that

$$\tilde{M}_{U_1,\mathsf{N}}[2I+S] = (2I+S) + (-2a+b)(I_4-2S), \tag{I13}$$

$$\tilde{M}_{U_1,\mathsf{N}}[I_4-2S] = (1-a-2b)(I_4-2S). \tag{I14}$$

Thus, using Eqs. (I13) and (I14) repeatedly, we have the following relation:

$$\underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \tilde{M}_{U_1,\mathsf{N}} \circ \cdots \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}} \left[\frac{1}{6}(I_4+S)\right] \tag{I15}$$

$$= \frac{1}{10}\underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \tilde{M}_{U_1,\mathsf{N}} \circ \cdots \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}}[(2I+S)] \tag{I16}$$
$$- \frac{1}{30}\underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \tilde{M}_{U_1,\mathsf{N}} \circ \cdots \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}}[(I_4-2S)]$$

$$= \frac{1}{10}(2I+S) + x_m(I_4-2S), \tag{I17}$$

where $\{x_n\}_n$ is defined by the recurrence relation

$$x_0 = -\frac{1}{30}, \tag{I18}$$

$$x_{n+1} = (1-a-2b)x_n + \frac{-2a+b}{10} \quad (n \geq 0). \tag{I19}$$

Solving this relation, we have

$$x_m = \frac{-2a+b}{10(a+2b)} + (1-a-2b)^m\left(-\frac{1}{30} - \frac{-2a+b}{10(a+2b)}\right), \tag{I20}$$

which leads to Eq. (I5). ■

Next, we derive formulas for the effect of noise on the identity operator and the SWAP operator.

Lemma 5. Let $\mathcal{N}$ be a single-qubit noise channel and let $\mathsf{N} = \mathcal{N} \otimes \mathcal{N}$ be two copies of $\mathcal{N}$ acting on two qubits. Define a two-qubit operator

$$\tilde{M}_{U_1,\mathsf{N}} = M_{U_1} \circ \mathsf{N} \circ M_{U_1} \tag{I21}$$

with

$$M_{U_1}[\rho] = \mathop{\mathbb{E}}_{U_1 \sim \mathcal{U}_{\text{Haar}}}\left[U_1^{\otimes 2} \rho U_1^{\dagger \otimes 2}\right]. \tag{I22}$$

Suppose that

$$\tilde{M}_{U_1,\mathsf{N}}(I_4) = (1-a)I_4 + 2aS, \tag{I23}$$

$$\tilde{M}_{U_1,\mathsf{N}}(S) = bI_4 + (1-2b)S \tag{I24}$$

with $a > 0$ and $b > 0$, where I_4 is the two-qubit identity operator and S is the two-qubit SWAP gate. Then,

$$\underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \tilde{M}_{U_1,\mathsf{N}} \circ \cdots \circ \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}}(I_4) = \left(1 - \frac{1-(1-a-2b)^m}{1+\frac{2b}{a}}\right) I_4 + \left(\frac{1-(1-a-2b)^m}{\frac{1}{2}+\frac{b}{a}}\right) S, \tag{I25}$$

$$\underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \tilde{M}_{U_1,\mathsf{N}} \circ \cdots \circ \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}}(S) = \left(\frac{1}{2} - \frac{\frac{1}{2}+\frac{b}{a}(1-a-2b)^m}{1+\frac{2b}{a}}\right) I_4 + \left(\frac{\frac{1}{2}+\frac{b}{a}(1-a-2b)^m}{\frac{1}{2}+\frac{b}{a}}\right) S. \tag{I26}$$

Proof. Let us write

$$\underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \tilde{M}_{U_1,\mathsf{N}} \circ \cdots \circ \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}}(I_4) = x_m I_4 + y_m S. \tag{I27}$$

Using Eqs. (I23) and (I24),

$$x_{m+1} = (1-a)x_m + b y_m, \tag{I28}$$
$$y_{m+1} = 2a x_m + (1-2b) y_m. \tag{I29}$$

This is equivalent to

$$x_{m+1} + \frac{1}{2} y_{m+1} = x_m + \frac{1}{2} y_m, \tag{I30}$$
$$x_{m+1} - \frac{b}{a} y_{m+1} = (1-a-2b)\left(x_m - \frac{b}{a} y_m\right). \tag{I31}$$

Therefore, we have

$$x_m + \frac{1}{2} y_m = x_0 + \frac{1}{2} y_0 = 1, \tag{I32}$$
$$x_m - \frac{b}{a} y_m = (1-a-2b)^m \left(x_0 - \frac{b}{a} y_0\right) = (1-a-2b)^m. \tag{I33}$$

Hence,

$$x_m = 1 - \frac{1-(1-a-2b)^m}{1+\frac{2b}{a}}, \tag{I34}$$
$$y_m = \frac{1-(1-a-2b)^m}{\frac{1}{2}+\frac{b}{a}}. \tag{I35}$$

With a very similar argument, letting us write

$$\underbrace{\tilde{M}_{U_1,\mathsf{N}} \circ \tilde{M}_{U_1,\mathsf{N}} \circ \cdots \circ \tilde{M}_{U_1,\mathsf{N}}}_{m \text{ times}}(S) = z_m I_4 + w_m S, \tag{I36}$$

we have

$$z_m = \frac{1}{2} - \frac{\frac{1}{2}+\frac{b}{a}(1-a-2b)^m}{1+\frac{2b}{a}}, \tag{I37}$$
$$w_m = \frac{\frac{1}{2}+\frac{b}{a}(1-a-2b)^m}{\frac{1}{2}+\frac{b}{a}}. \tag{I38}$$

■

2. Replacement of two-qubit random gate by single-qubit gates

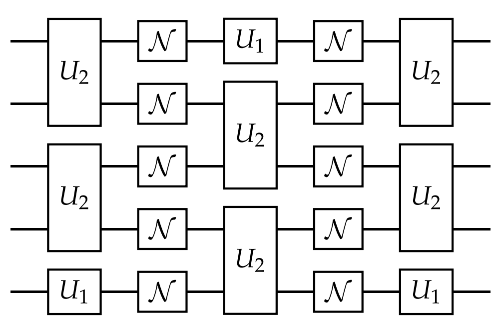
In the proof of Theorem 4, instead of directly evaluating the second moment of the original circuit, we consider a slightly simplified circuit, in which two-qubit random unitary gates are replaced with single-qubit random gates and the last layer of noise is removed. The circuit obtained by this replacement is composed solely of single-qubit Haar-random unitary gates and noise channel $\mathcal{N}$, without the last layer of noise. We prove that the second moment of the simplified circuit gives an upper bound of the original second moment that we desired to evaluate.

Lemma 6. Consider an ensemble $\mathcal{B}$ of noisy random quantum circuits with noise channel $\mathcal{N}$ satisfying Eqs. (I3) and (I4). Let $\mathcal{B}'$ denote the ensemble of circuits obtained by removing the last layer of noise from the noisy random circuits in $\mathcal{B}$. Consider another ensemble $\tilde{\mathcal{B}}'$ of circuits that can be obtained by replacing each two-qubit Haar-random gate U_2 in the circuits of $\mathcal{B}'$ as

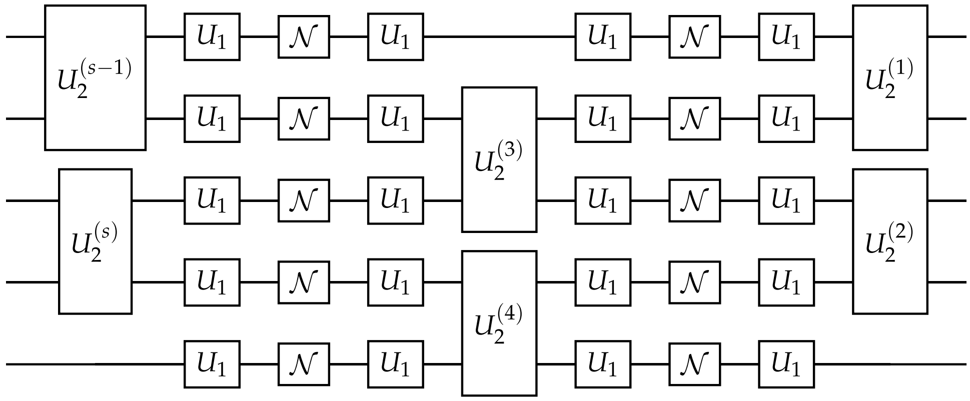
$$U_2 \to \left(U_1 \otimes U_1'\right), \tag{I39}$$

where U_1 and U_1' are independent single-qubit Haar-random gates. In this setup, if $0 \le 1-a-2b \le 1$, then $\mathbb{E}_{\mathcal{B}'}[p_x^2] \le \mathbb{E}_{\tilde{\mathcal{B}}'}[p_x^2]$ for any $x \in \{0,1\}^n$.

Proof. Any circuit $\mathcal{C}$ in $\mathcal{B}'$ can be written in the following form:



Let s be the number of two-qubit Haar-random gates in given circuit $\mathcal{C} \in \mathcal{B}'$. Let us introduce a numbering of the two-qubit Haar-random gates in $\mathcal{C}$: $U_2^{(k)}$ refers to the kth two-qubit Haar-random gate, where the counting starts from the top gate at the last layer. Rewiring the circuit $\mathcal{C}$, we have



(I40)

Construct a sequence of circuits $\{\mathcal{C}_j : j = 0, 1, 2, \ldots, s\}$ in the following recursive way.

(a) $\mathcal{C}_0 = \mathcal{C} \in \mathcal{B}'$.
(b) For $j = 1, 2, \ldots, s$, $\mathcal{C}_j$ is a circuit obtained by replacing the leading two-qubit Haar-random gate in $\mathcal{C}_{j-1}$, i.e., $U_2^{(j)}$, by two parallel independent single-qubit Haar-random gates.

By construction, $\mathcal{C}_s$ consists only of single-qubit Haar-random gates and noise channels and $\mathcal{C}_s \in \tilde{\mathcal{B}}'$. Let $x \in \{0,1\}^n$ be any n-bit string. Here, we aim to show that

$$\mathop{\mathbb{E}}_{\mathcal{C}_0 \sim \mathcal{B}'}[p_x^2] \leq \mathop{\mathbb{E}}_{\mathcal{C}_s \sim \tilde{\mathcal{B}}'}[p_x^2] \tag{I41}$$

and, to this end, it suffices to show that

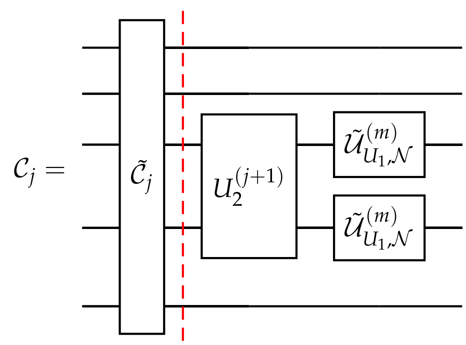
$$\mathop{\mathbb{E}}_{\mathcal{C}_j}[p_x^2] \leq \mathop{\mathbb{E}}_{\mathcal{C}_{j+1}}[p_x^2] \tag{I42}$$

for all $j = 0, 1, 2, \ldots, s-1$. For this purpose, fix j and let us look at $\mathcal{C}_j$ and $\mathcal{C}_{j+1}$. We will show that

$$\mathop{\mathbb{E}}_{\mathcal{C}_j}[p_x^2] \leq \mathop{\mathbb{E}}_{\mathcal{C}_{j+1}}[p_x^2] \tag{I43}$$

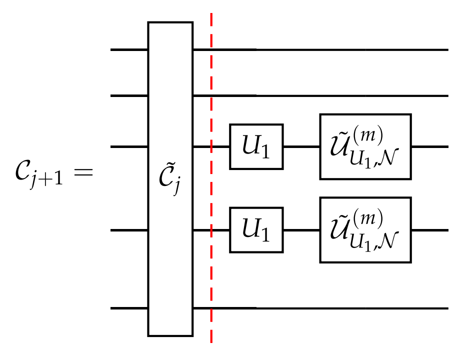
for this j. Suppose that $U_2^{(j)}$ acts on the kth and lth qubits. Recall the characterization shown in Eq. (I40).

Then, by absorbing the action of the single-qubit Haar-random gates and noise channels acting on qubits other than the kth or lth one to the preceding gates, we may write



(I44)

and



(I45)

where

$$\tilde{\mathcal{U}}_{U_1,\mathcal{N}}^{(m)} := \underbrace{(\mathcal{U}_1 \circ \mathcal{N} \circ \mathcal{U}_1) \circ \cdots \circ (\mathcal{U}_1 \circ \mathcal{N} \circ \mathcal{U}_1)}_{m \text{ times}}, \tag{I46}$$

with some non-negative integer m. Here, we consider a statistical model with $\{I_4, S\}^n$ bit-string representation, which has been introduced in Ref. [20]. Suppose that we have $\gamma \in \{I_4, S\}^n$ just before the red dashed line. Suppose that γ will change as

$$\gamma \xrightarrow{\mathcal{C}_j\text{:After red line}} \sum_{\gamma' \in \{I_4,S\}^n} c_j^{(\gamma')}\gamma' \tag{I47}$$

$$\gamma \xrightarrow{\mathcal{C}_{j+1}\text{:After red line}} \sum_{\gamma' \in \{I_4,S\}^n} c_{j+1}^{(\gamma')}\gamma'. \tag{I48}$$

Define

$$a_j^{(\gamma)} := \sum_{\gamma' \in \{I_4,S\}^n} c_j^{(\gamma')} \tag{I49}$$

$$a_{j+1}^{(\gamma)} := \sum_{\gamma' \in \{I_4,S\}^n} c_{j+1}^{(\gamma')}. \tag{I50}$$

It suffices to show that

$$a_j^{(\gamma)} \leq a_{j+1}^{(\gamma)} \tag{I51}$$

for all γ. Observe that after the red line, the gates only act on the kth and the lth qubits. So, let us focus on these two qubits:

(1) If $\gamma_{k,l} = II, SS$, since

$$II \xrightarrow{U_2^{(j+1)}} II, \tag{I52}$$

$$II \xrightarrow{U_1 \otimes U_1} II \tag{I53}$$

and

$$SS \xrightarrow{U_2^{(j+1)}} SS, \tag{I54}$$

$$SS \xrightarrow{U_1 \otimes U_1} SS, \tag{I55}$$

we have

$$a_j^{(\gamma)} = a_{j+1}^{(\gamma)} \tag{I56}$$

in this case.

(2) If $\gamma_{k,l} = IS, SI$, without loss of generality, we may assume that $\gamma_{k,l} = IS$. The proof for $\gamma_{k,l} = SI$ follows similarly. We have

$$IS \xrightarrow{U_2^{(j+1)}} \frac{2}{5}\left(II + SS\right), \tag{I57}$$

$$IS \xrightarrow{U_1 \otimes U_1} IS. \tag{I58}$$

By Lemma 5,

$$IS \xrightarrow{\mathcal{C}_j\text{:after red line}} \frac{2}{5}\left((x_m I_4 + y_m S)(x_m I_4 + y_m S) + (z_m I_4 + w_m S)(z_m I_4 + w_m S)\right), \tag{I59}$$

$$IS \xrightarrow{\mathcal{C}_{j+1}\text{:after red line}} (x_m I_4 + y_m S)(z_m I_4 + w_m S), \tag{I60}$$

where

$$x_m = 1 - \frac{1-(1-a-2b)^m}{1+\frac{2b}{a}}, \tag{I61}$$

$$y_m = \frac{1-(1-a-2b)^m}{\frac{1}{2}+\frac{b}{a}}, \tag{I62}$$

$$z_m = \frac{1}{2} - \frac{\frac{1}{2}+\frac{b}{a}(1-a-2b)^m}{1+\frac{2b}{a}}, \tag{I63}$$

$$w_m = \frac{\frac{1}{2}+\frac{b}{a}(1-a-2b)^m}{\frac{1}{2}+\frac{b}{a}}. \tag{I64}$$

Therefore,

$$a_j^{(\gamma)} = \frac{2}{5}\left(u_m^2 + v_m^2\right), \tag{I65}$$

$$a_{j+1}^{(\gamma)} = u_m v_m, \tag{I66}$$

where

$$u_m = x_m + y_m = 1 + \frac{1-(1-a-2b)^m}{1+\frac{2b}{a}}, \tag{I67}$$

$$v_m = z_m + w_m = \frac{1}{2} + \frac{\frac{1}{2}+\frac{b}{a}(1-a-2b)^m}{1+\frac{2b}{a}}. \tag{I68}$$

Now,

$$\begin{aligned} & a_j^{(\gamma)} \leq a_{j+1}^{(\gamma)} \\ \Leftrightarrow\ & \frac{2}{5}\left(u_m^2+v_m^2\right) \leq u_m v_m \\ \Leftrightarrow\ & 0 \leq (2v_m - u_m)(2u_m - v_m). \end{aligned} \tag{I69}$$

Here, if $0 \leq 1-a-2b$,

$$2v_m - u_m = \frac{\left(\frac{b}{a}+2\right)(1-a-2b)^m}{1+\frac{2b}{a}} \geq 0, \tag{I70}$$

and if $0 \le 1-a-2b \le 1$,

$$2u_m - v_m = \left(2 + \frac{2-2(1-a-2b)^m}{1+\frac{2b}{a}}\right) - \left(\frac{1}{2} + \frac{\frac{1}{2}+\frac{b}{a}(1-a-2b)^m}{1+\frac{2b}{a}}\right) \ge 2-1 = 1 \ge 0. \tag{I71}$$

Hence, if $0 \le 1-a-2b \le 1$, $a_j^{(\gamma)} \le a_{j+1}^{(\gamma)}$, which completes the proof.

■

3. Main proof

Now, we finally give our proof of Theorem 4. In the noise models that we consider, we have

$$a = \begin{cases} \dfrac{q^2}{3}, & \mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}, \\ \dfrac{q^2(1-p)^2}{3}, & \mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}, \end{cases} \tag{I72}$$

$$b = \begin{cases} \dfrac{1}{2} - \dfrac{q^2}{6} - \dfrac{1}{6}(1-p)^2(1-q)(3-q), & \mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}, \\ \dfrac{1}{2} - \dfrac{q^2(1-p)^2}{6} - \dfrac{1}{6}(1-p)^2(1-q)(3-q), & \mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}. \end{cases} \tag{I73}$$

We can easily verify that

$$1-a-2b = (1-p)^2(1-q)\left(1-\frac{q}{3}\right) \tag{I74}$$

in both cases. Hence, $0 \le 1-a-2b \le 1$. Now, we rewrite the second-moment probability by using the weighted trajectories of the strings in $\{I_4, S\}^n$ (see also the proof of Lemma 6 in Ref. [6]). Suppose that the circuits in $\mathcal{B}'$ contain s Haar-random gates. A trajectory $\gamma := (\gamma^1\gamma^2\cdots\gamma^s\gamma^{s+1}) \in \{I_4, S\}^{n\times(s+1)}$ is a sequence of n-bit string $\gamma^i \in \{I_4, S\}^n$. For $i = 1, 2, \ldots, s$, $\gamma^i \in \{I_4, S\}^n$ represents the n-bit string right before the sth Haar-random gate in the trajectory γ. In particular, γ^1 is the initial string of the trajectory γ. $\gamma^{s+1} \in \{I_4, S\}^n$ represents the final bit string of the trajectory γ at the end of the circuit. For each trajectory $\gamma \in \{I_4, S\}^{n\times(s+1)}$, we define the weight $\mathrm{wt}_{\mathcal{B}'}(\gamma)$ as

$$\mathrm{wt}_{\mathcal{B}'}(\gamma) := c_1c_2c_3\cdots c_s, \tag{I75}$$

where for $i = 1, 2, \ldots, s$,

$$c_i := \begin{cases} \text{coefficient of transformation } \gamma^i \to \gamma^{i+1}, & \gamma^i \to \gamma^{i+1} \quad \text{is possible}, \\ 0, & \gamma^i \to \gamma^{i+1} \quad \text{is impossible}. \end{cases} \tag{I76}$$

Thus, the second-moment probability with respect to $\mathcal{B}'$ can be expressed as

$$\mathbb{E}_{\mathcal{B}'}[p_x^2] = \sum_{\gamma_1,\gamma_2,\ldots,\gamma_n\in\{I_4,S\}} \left[\left(\sum_{\substack{\gamma\in\{I_4,S\}^{n\times(s+1)}\\ \gamma^{s+1}=\gamma_1\gamma_2\cdots\gamma_n}} \mathrm{wt}_{\mathrm{B}'}(\gamma)\right) \langle x|\langle x|(\gamma_1\gamma_2\cdots\gamma_n)|x\rangle|x\rangle\right], \tag{I77}$$

where in the right-hand side, we divide the cases based on the final bit string $\gamma^{s+1} = \gamma_1\gamma_2\cdots\gamma_n$. We may further rewrite as

$$\mathbb{E}_{\mathcal{B}'}[p_x^2] = \sum_{\gamma_1,\gamma_2,\ldots,\gamma_n\in\{I_4,S\}} \left[\left(\sum_{\substack{\gamma\in\{I_4,S\}^{n\times(s+1)}\\ \gamma^{s+1}=\gamma_1\gamma_2\cdots\gamma_n}} \mathrm{wt}_{\mathrm{B}'}(\gamma)\right)\left(\prod_{j=1}^n \langle x_j|\langle x_j|\gamma_j|x_j\rangle|x_j\rangle\right)\right]. \tag{I78}$$

For $\gamma_j \in \{I_4, S\}$, $\langle 0| \langle 0| \gamma_i |0\rangle |0\rangle = \langle 1| \langle 1| \gamma_i |1\rangle |1\rangle = 1$. Therefore,

$$\mathbb{E}_{\mathcal{B}'}[p_x^2] = \sum_{\gamma_1,\gamma_2,\ldots,\gamma_n \in \{I_4,S\}} \left[\left(\sum_{\substack{\gamma \in \{I_4,S\}^{n\times(s+1)} \\ \gamma^{s+1}=\gamma_1\gamma_2\cdots\gamma_n}} \mathrm{wt}_{\mathrm{B}'}(\gamma) \right) \right]. \tag{I79}$$

On the other hand, the second-moment probability with respect to the original ensemble $\mathcal{B}$ can be obtained by considering the last layer of noise. Thus, in this case, instead of Eq. (I78), we have

$$\mathbb{E}_{\mathcal{B}}[p_x^2] = \sum_{\gamma_1,\gamma_2,\ldots,\gamma_n \in \{I_4,S\}} \left[\left(\sum_{\substack{\gamma \in \{I_4,S\}^{n\times(s+1)} \\ \gamma^{s+1}=\gamma_1\gamma_2\cdots\gamma_n}} \mathrm{wt}_{\mathrm{B}'}(\gamma) \right) \left(\prod_{j=1}^{n} \langle x_j | \langle x_j | \mathsf{N}(\gamma_j) | x_j \rangle | x_j \rangle \right) \right]. \tag{I80}$$

Here, observe that

$$\langle x_j | \langle x_j | \mathsf{N}(\gamma_j) | x_j \rangle | x_j \rangle \leq \max\{\langle x_j | \langle x_j | \mathsf{N}(I_4) | x_j \rangle | x_j \rangle, \langle x_j | \langle x_j | \mathsf{N}(S) | x_j \rangle | x_j \rangle\} =: e_j \tag{I81}$$

for each j. Hence,

$$\mathbb{E}_{\mathcal{B}}[p_x^2] \leq \left[\sum_{\gamma_1,\gamma_2,\ldots,\gamma_n \in \{I_4,S\}} \left(\sum_{\substack{\gamma \in \{I_4,S\}^{n\times(s+1)} \\ \gamma^{s+1}=\gamma_1\gamma_2\cdots\gamma_n}} \mathrm{wt}_{\mathrm{B}'}(\gamma) \right) \right] (e_1 e_2 \cdots e_n). \tag{I82}$$

Using Eq. (I79),

$$\mathbb{E}_{\mathcal{B}}[p_x^2] \leq \mathbb{E}_{\mathcal{B}'}[p_x^2]\,(e_1 e_2 \cdots e_n). \tag{I83}$$

By Lemma 6, since $\mathbb{E}_{\mathcal{B}'}[p_x^2] \leq \mathbb{E}_{\tilde{\mathcal{B}}'}[p_x^2]$,

$$\mathbb{E}_{\mathcal{B}}[p_x^2] \leq \mathbb{E}_{\tilde{\mathcal{B}}'}[p_x^2]\,(e_1 e_2 \cdots e_n). \tag{I84}$$

Now, by the construction of $\tilde{\mathcal{B}}'$ and Lemma 4,

$$\mathbb{E}_{\tilde{\mathcal{B}}'}[p_x^2] = \left(\frac{3}{10} - \left[\frac{-2a+b}{10(a+2b)} + (1-a-2b)^{d-1} \left(-\frac{1}{30} - \frac{-2a+b}{10(a+2b)} \right) \right] \right)^n. \tag{I85}$$

Here, we have used the fact that $\langle 0| \langle 0| (2I+S) |0\rangle |0\rangle = \langle 1| \langle 1| (2I+S) |1\rangle |1\rangle = 3$ and $\langle 0| \langle 0| (I_4 - 2S) |0\rangle |0\rangle = \langle 1| \langle 1| (I_4 - 2S) |1\rangle |1\rangle = -1$. By definition of a and b, letting

$$c := 1 - (1-p)^2(1-q)\left(1 - \frac{q}{3}\right), \tag{I86}$$

$$r := \begin{cases} q, & \mathcal{N} = \mathcal{N}_q^{(\mathrm{amp})} \circ \mathcal{N}_p^{(\mathrm{dep})}, \\ q(1-p), & \mathcal{N} = \mathcal{N}_p^{(\mathrm{dep})} \circ \mathcal{N}_q^{(\mathrm{amp})}, \end{cases} \tag{I87}$$

we have

$$\mathbb{E}_{\tilde{\mathcal{B}}'}[p_x^2] = \left(\left(\frac{1}{4} + \frac{r^2}{12c} \right) + (1-c)^{d-1} \left(\frac{1}{12} - \frac{r^2}{12c} \right) \right)^n. \tag{I88}$$

By taking

$$\mu = \frac{1}{4} + \frac{r^2}{12c}, \tag{I89}$$

$$\nu = \frac{1}{12} - \frac{r^2}{12c}, \tag{I90}$$

we obtain

$$\mathbb{E}_{\tilde{\mathcal{B}}'}[p_x^2] = \left(\mu + (1-c)^{d-1}\nu\right)^n. \tag{I91}$$

Here, we will check $\mu, \nu \geq 0$. It follows trivially that $\mu \geq 0$ from the definition. Next, we evaluate ν. Since $r \leq q$ by definition,

$$\nu = \frac{1}{12} - \frac{r^2}{12c} \tag{I92}$$

$$\geq \frac{1}{12} - \frac{q^2}{12c} \tag{I93}$$

$$= \frac{c - q^2}{12c}. \tag{I94}$$

By substituting Eqs. (I86) and (I87),

$$\frac{c-q^2}{12c} = \frac{1-(1-p)^2(1-q)\left(1-\frac{q}{3}\right)-q^2}{12c}. \tag{I95}$$

Since $(1-p)^2 \leq 1$,

$$\frac{1-(1-p)^2(1-q)\left(1-\frac{q}{3}\right)-q^2}{12c}$$

$$\geq \frac{1-(1-q)\left(1-\frac{q}{3}\right)-q^2}{12c} \tag{I96}$$

$$= \frac{\frac{4q}{3}(1-q)}{12c} \tag{I97}$$

$$= \frac{q(1-q)}{9c}. \tag{I98}$$

Since $q \geq 0$ and $1-p \geq 0$, we have

$$\nu \geq 0. \tag{I99}$$

Since $1 + x \leq \mathrm{e}^x$ for all $x \in \mathbb{R}$,

$$\mathbb{E}_{\tilde{\mathcal{B}}'}[p_x^2] \leq \mu^n \exp\left[n\frac{\nu}{\mu}\mathrm{e}^{-c(d-1)}\right]. \tag{I100}$$

Using Eqs. (I84) and (I100),

$$\mathbb{E}_{\mathcal{B}}[p_x^2] \leq \mu^n \exp\left[n\frac{\nu}{\mu}\mathrm{e}^{-c(d-1)}\right](e_1 e_2 \cdots e_n). \tag{I101}$$

Now, we evaluate e_j:

$$e_j := \max\left\{\langle x_j | \langle x_j | \mathsf{N}(I_4) | x_j\rangle | x_j\rangle, \langle x_j | \langle x_j | \mathsf{N}(S) | x_j\rangle | x_j\rangle\right\}$$

$$= \begin{cases} (1+r)^2, & x_j = 0, \\ (1-r)^2, & x_j = 1. \end{cases} \tag{I102}$$

That is, if $w_x \geq \frac{n}{2}$,

$$e_1 e_2 \cdots e_n = (1+r)^{2(n-w_x)}(1-r)^{2w_x}$$

$$\leq (1+r)^n(1-r)^n = (1-r^2)^n = \eta^n. \tag{I103}$$

Combining Eqs. (I101) and (I103),

$$\mathbb{E}_{\mathcal{B}}[p_x^2] \leq \mu^n\eta^n \exp\left[n\frac{\nu}{\mu}\mathrm{e}^{-c(d-1)}\right], \tag{I104}$$

which completes the proof.

APPENDIX J: THE EFFECT OF LAST LAYER OF NOISELESS GATES

In this appendix, we analyze the effect of a layer of single-qubit gates that immediately follows the last layer of noise. This setup is equivalent to arbitrarily locally rotating the measurement basis. Moreover, we assume that these gates are noiseless. A single-qubit gate U is parametrized as

$$U(\theta, \phi) = \begin{pmatrix} \cos(\theta)e^{i\phi} & \sin(\theta) \\ -\sin(\theta) & \cos(\theta)e^{-i\phi} \end{pmatrix}. \tag{J1}$$

Let $U_i(\theta_i, \phi_i)$ be the unitary applied to the ith qubit in the last layer.

1. Lack of anticoncentration via weak definition

In Corollary 2, we have seen that a noisy random circuit with the last layer of noiseless gates is not anticoncentrated in terms of the collision probability (Definition 10) under a certain condition. Here, we investigate whether a noisy random circuit with the last layer of *noiseless* gates is anticoncentrated in the sense of Definition 12. We succeed in giving a condition on the last layer of noiseless gates under which the given circuit shows the lack of anticoncentration.

Theorem 9. Let $\mathcal{B}$ be an ensemble of amplitude-damped random quantum circuits, with noise strength q. Additionally, before measurement, for every $i \in [n]$, let $U_i(\theta_i, \phi_i)$—a single-qubit noiseless gate—be applied to qubit i. If

$$4 - \sqrt{15} < |\cos 2\theta_i| \tag{J2}$$

for all $i \in [n]$, then there exists a string $x \in \{0,1\}^n$ such that for any $q > 0$ and $d = \Omega(\log n)$,

$$\lim_{n\to\infty} \Pr_{\mathcal{B}}\left[p_x < \frac{\alpha}{2^n}\right] = 1 \tag{J3}$$

for any $\alpha \in (0, 1]$.

Proof. Following the same argument as in the proof of Theorem 4 in Appendix I, for sufficiently large depth and sufficiently large n, we have

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x^2] \leq \left(\left(\frac{1}{4-q}\right)^n (e_1 e_2 \cdots e_n)\right) \times \mathcal{O}(1), \tag{J4}$$

where

$$\begin{aligned} e_j &= \max\left\{\langle x_j | \langle x_j | (\mathcal{U}_j \otimes \mathcal{U}_j) \mathsf{N}(I_4) (\mathcal{U}_j \otimes \mathcal{U}_j)^\dagger |x_j\rangle |x_j\rangle, \right. \\ &\quad \left. \langle x_j | \langle x_j | (\mathcal{U}_j \otimes \mathcal{U}_j) \mathsf{N}(S) (\mathcal{U}_j \otimes \mathcal{U}_j)^\dagger |x_j\rangle |x_j\rangle \right\} \\ &= \begin{cases} (1 + q\cos 2\theta_j)^2, & x_j = 0, \\ (1 - q\cos 2\theta_j)^2, & x_j = 1 \end{cases} \end{aligned} \tag{J5}$$

for $j = 1, 2, \ldots, n$. Choose x so that

$$e_j = (1 - q|\cos 2\theta_j|)^2. \tag{J6}$$

In this case,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_x^2] \leq \left(\frac{(1 - q|\cos 2\theta|)^2}{4-q}\right)^n \times \mathcal{O}(1), \tag{J7}$$

where

$$\theta := \mathop{\arg\min}_{\theta_j : j \in [n]} |\cos 2\theta_j|. \tag{J8}$$

To see the concentration, it suffices to check if

$$4\frac{(1 - q|\cos 2\theta|)^2}{4-q} < 1, \tag{J9}$$

which is equivalent to

$$q\left(4|\cos 2\theta|^2 q - (8|\cos 2\theta| - 1)\right) < 0. \tag{J10}$$

When $q = 0$, this inequality cannot be satisfied, so $q > 0$. Under this condition, we have

$$4|\cos 2\theta|^2 q - (8|\cos 2\theta| - 1) < 0. \tag{J11}$$

When

$$4 - \sqrt{15} < |\cos 2\theta|, \tag{J12}$$

all $q > 0$ satisfy this condition, which completes the proof. ∎

2. Bias in output distribution

In this section, we see how the last layer of single-qubit noiseless gates introduces bias in the output distribution.

Theorem 10. Let $\mathcal{B}$ be an ensemble of amplitude-damped random quantum circuits, with noise strength q. Additionally, before measurement, for every $i \in [n]$, let $U_i(\theta_i, \phi_i)$—a single-qubit, noiseless gate—be applied to qubit i. Then,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_{i,b}] = \frac{1}{2} + \frac{(-1)^b q\cos 2\theta_i}{2}, \tag{J13}$$

where $p_{i,b}$ is marginal probabilities of obtaining $b \in \{0, 1\}$, in the ith qubit.

Proof. As per the earlier convention, let $\mathcal{N}$ be the single-qubit noise channel and let K_0 and K_1 be the corresponding Kraus operators. Let $\mathcal{C} \in \mathcal{B}$ be a noisy random circuit. Let $\tilde{\mathcal{C}}$ be the quantum circuit *without the last layer*; i.e.,

$$\mathcal{C} = \mathcal{N}^{\otimes n} \circ \tilde{\mathcal{C}}. \tag{J14}$$

For $i \in [n]$, define ρ_i as follows:

$$\rho_i = \mathop{\mathbb{E}}_{\mathcal{B}}[\mathrm{Tr}_{1,\ldots,i-1,i+1,\ldots n}(\tilde{\mathcal{C}}(|0^n\rangle\langle 0^n|))], \tag{J15}$$

where I_2 is the single-qubit identity matrix. In other words, ρ_i is the expected reduced density matrix on just the ith qubit. By definition, we have

$$\rho_i = \frac{I_2}{2}. \tag{J16}$$

In the last noiseless layer, since the parameters are implicit, we will drop the subscripts and refer to the ith single-qubit gate as just U_i. Then, by the definition of the adjoint map,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_{i,0}] = \mathop{\mathbb{E}}_{\mathcal{B}}[\mathrm{Tr}(|0\rangle\langle 0| U_i \mathcal{N}(\rho_i) U_i^\dagger)] \tag{J17}$$

$$= \mathop{\mathbb{E}}_{\mathcal{B}}[\mathrm{Tr}(U_i^\dagger |0\rangle\langle 0| U_i \mathcal{N}(\rho_i))] \tag{J18}$$

$$= \mathop{\mathbb{E}}_{\mathcal{B}}[\mathrm{Tr}(U_i^\dagger |0\rangle\langle 0| U_i \mathcal{N}(\rho_i))] \tag{J19}$$

$$\begin{aligned} &= \mathop{\mathbb{E}}_{\mathcal{B}}[\mathrm{Tr}(K_0^\dagger U_i^\dagger |0\rangle\langle 0| U_i K_0 \rho_i)] \\ &\quad + \mathop{\mathbb{E}}_{\mathcal{B}}[\mathrm{Tr}(K_1^\dagger U_i^\dagger |0\rangle\langle 0| U_i K_1 \rho_i)], \end{aligned} \tag{J20}$$

where we have repeatedly used the cyclic property of trace. In Eq. (J20), we have used the Kraus operators K_0 and

K_1 of the amplitude-damping noise channel. Since $\rho_i = \frac{I_2}{2}$, Eq. (J20) is equal to

$$\frac{1}{2}\left(\mathrm{Tr}(K_0^\dagger U_i^\dagger |0\rangle\langle 0| U_i K_0) + \mathrm{Tr}(K_1^\dagger U_i^\dagger |0\rangle\langle 0| U_i K_1)\right). \tag{J21}$$

By explicitly computing each term, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{B}}[p_{i,0}] &= \frac{1}{2}\left((\cos^2\theta_i + (1-q)\sin^2\theta_i) + q\cos^2\theta_i\right)\\ &= \frac{1+q(\cos^2\theta_i - \sin^2\theta_i)}{2}\\ &= \frac{1}{2} + \frac{q\cos 2\theta_i}{2}.\end{aligned} \tag{J22}$$

■

Let us define the "bias" β_i of qubit i as

$$\beta_i = q\cos 2\theta_i. \tag{J23}$$

Note that $\beta_i \in [-q, q]$. Hence,

$$\mathop{\mathbb{E}}_{\mathcal{B}}[p_{i,b}] = \frac{1}{2} + \frac{\beta_i}{2}, \quad \mathop{\mathbb{E}}_{\mathcal{B}}[p_{i,b}] = \frac{1}{2} - \frac{\beta_i}{2}. \tag{J24}$$

One way to interpret Theorem 10 is to observe that there is an "effective" amplitude-damping channel acting on each qubit but each such channel has a "tunable" noise strength, which can now also be negative.

A negative noise strength means that the outcome 0 is suppressed and the outcome 1 is assigned higher weight. However, *which* strings are suppressed, and by *how much*, depends on the "bias" of each qubit. By changing the single-qubit gates, we can control the bias.

[1] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).
[2] D. S. França and R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, Nat. Phys. **17**, 1221 (2021).
[3] A. Bouland, B. Fefferman, Z. Landau, and Y. Liu, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Denver, CO, USA, 2022).
[4] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, On the complexity and verification of quantum random circuit sampling, Nat. Phys. **15**, 159 (2018).
[5] R. Movassagh, Efficient unitary paths and quantum computational supremacy: A proof of average-case hardness of random circuit sampling, ArXiv:1810.04681.
[6] A. Deshpande, P. Niroula, O. Shtanko, A. V. Gorshkov, B. Fefferman, and M. J. Gullans, Tight bounds on the convergence of noisy random circuits to the uniform distribution, PRX Quantum **3**, 040329 (2022).
[7] S. Aaronson and A. Arkhipov, in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing, STOC '11* (Association for Computing Machinery, San Jose, California, USA, 2011), pp. 333–342.
[8] A. Nahum, S. Vijay, and J. Haah, Operator spreading in random unitary circuits, Phys. Rev. X **8**, 021014 (2018).
[9] A. Bouland, B. Fefferman, and U. Vazirani, Computational pseudorandomness, the wormhole growth paradox, and constraints on the AdS/CFT duality, ArXiv:1910.14646.
[10] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).
[11] 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).
[12] 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).
[13] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys. **264**, 773 (2006).
[14] S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, The future of quantum computing with superconducting qubits, J. Appl. Phys. **132**, 160902 (2022).
[15] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput. **38**, 1207 (2008).
[16] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation, Science **279**, 342 (1998).
[17] A. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. (NY) **303**, 2 (2003).
[18] D. Aharonov and M. Ben-Or, in *Proceedings of 37th Conference on Foundations of Computer Science* (Burlington, VT, USA, 1996), pp. 46–55.
[19] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023* (Association for Computing Machinery, Orlando, FL, USA, 2023), pp. 945–957.
[20] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits anticoncentrate in log depth, PRX Quantum **3**, 010333 (2022).
[21] S. Debnath, N. M. Linke, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, Demonstration of a small programmable quantum computer with atomic qubits, Nature **536**, 63 (2016).
[22] F. Arute, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).
[23] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. S. Allman, C. H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, C. Ryan-Anderson, and B. Neyenhuis, Demonstration of the trapped-ion quantum CCD computer architecture, Nature **592**, 209 (2021).
[24] Q. Zhu, *et al.*, Quantum computational advantage via 60-qubit 24-cycle random circuit sampling, Sci. Bull. **67**, 240 (2022).

[25] M. Carroll, S. Rosenblatt, P. Jurcevic, I. Lauer, and A. Kandala, Dynamics of superconducting qubit relaxation times, npj Quantum Inf. **8**, 132 (2022).

[26] A. Morvan, *et al.*, Phase transition in random circuit sampling, arXiv:2304.11119 [quant-ph].

[27] S. Aaronson and L. Chen, in *Proceedings of the 32nd Computational Complexity Conference, CCC '17* (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Riga, Latvia, 2017).

[28] S. Aaronson and S. Gunn, On the classical hardness of spoofing linear cross-entropy benchmarking, Theory Comput. **16**, 1 (2020).

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

[30] 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, PRX Quantum **5**, 010334 (2024).

[31] B. Ware, A. Deshpande, D. Hangleiter, P. Niroula, B. Fefferman, A. V. Gorshkov, and M. J. Gullans, A sharp phase transition in linear cross-entropy benchmarking, ArXiv:2305.04954.

[32] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[33] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[34] J. Napp, Quantifying the barren plateau phenomenon for a model of unstructured variational ansätze, ArXiv:2203.06174.

[35] A. A. Mele, A. Angrisani, S. Ghosh, S. Khatri, J. Eisert, D. S. França, and Y. Quek, Noise-induced shallow circuits and absence of barren plateaus, ArXiv:2403.13927.

[36] Y. Shao, F. Wei, S. Cheng, and Z. Liu, Simulating quantum mean values in noisy variational quantum algorithms: A polynomial-scale approach, ArXiv:2306.05804.

[37] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2010).

[38] A. Kubica, A. Haim, Y. Vaknin, H. Levine, F. Brandão, and A. Retzker, Erasure qubits: Overcoming the T_1 limit in superconducting circuits, Phys. Rev. X **13**, 041022 (2023).

[39] M. Ben-Or, D. Gottesman, and A. Hassidim, Quantum refrigerator, ArXiv:1301.1995.

[40] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Average-case complexity versus approximate simulation of commuting quantum computations, Phys. Rev. Lett. **117**, 080501 (2016).

[41] W. J. Fefferman, Ph.D. thesis, Computer Science, Engineering and Applied Science, California Institute of Technology, 2014.

[42] J. C. Napp, R. L. La Placa, A. M. Dalzell, F. G. S. L. Brandão, and A. W. Harrow, Efficient classical simulation of random shallow 2D quantum circuits, Phys. Rev. X **12**, 021021 (2022).

[43] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, Random quantum circuits transform local noise into global white noise, Commun. Math. Phys. **405**, 78 (2024).

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

[45] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Achieving quantum supremacy with sparse and noisy commuting quantum computations, Quantum **1**, 8 (2017).

[46] X. Gao and L. Duan, Efficient classical simulation of noisy quantum computation, ArXiv:1810.03176.

[47] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of noisy reversible computation, ArXiv:quant-ph/9611028.

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

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

[50] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).