Featured in Physics

A Polynomial-Time Classical Algorithm for Noisy Quantum Circuits

Thomas Schuster,[1,*] Chao Yin,[2,*] Xun Gao,[2,3] and Norman Y. Yao[4]

[1]Walter Burke Institute for Theoretical Physics and Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA
[2]Department of Physics and Center for Theory of Quantum Matter, University of Colorado, Boulder, Colorado 80309, USA
[3]JILA, University of Colorado, Boulder, Colorado 80309, USA
[4]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

(Received 29 October 2024; revised 3 April 2025; accepted 15 August 2025; published 3 November 2025)

We provide a polynomial-time classical algorithm for noisy quantum circuits. The algorithm computes the expectation value of any observable for any circuit, with a small average error over input states drawn from an ensemble (e.g., the computational basis). Our approach is based upon the intuition that noise exponentially damps nonlocal correlations relative to local correlations. This enables one to classically simulate a noisy quantum circuit by keeping track of only the dynamics of local quantum information. Our algorithm also enables sampling from the output distribution of a circuit in quasipolynomial time, so long as the distribution anticoncentrates. A number of implications are discussed, including a fundamental limit on the efficacy of noise mitigation strategies: For constant noise rates, any quantum circuit for which error mitigation succeeds in polynomial-time on most input states can also be classically simulated in polynomial-time on most input states. Our algorithms scale exponentially in the inverse noise rate, which is fundamental and makes them impractical for current quantum devices.

DOI: 10.1103/xct1-7kf2 Subject Areas: Quantum Physics, Quantum Information

I. INTRODUCTION

Quantum computers are believed to yield exponential computational advantages over their classical counterparts for certain tasks [1–4]. However, near-term quantum devices are inevitably impacted by noise. This raises a key question [5–51]: To what degree does noise *fundamentally* limit any quantum advantage over classical computation?

The answer to this question is well understood in two limits. On the one hand, for noise rates above a large threshold, seminal works have shown that noisy quantum circuits can be simulated classically [5–8]. On the other hand, for sufficiently low noise rates, one can utilize quantum error correction to perform fault-tolerant computation [9–12]. Nevertheless, the reality of modern quantum experiments lies precisely in between these two limits: Noise rates are small, but error correction is not typically employed. Progress in this intermediate regime has been restricted to several relatively specific contexts: bounds on the circuit depth but not the complexity [13–16], tasks that assume access to a noiseless oracle [17], hardness results for sampling with extremely high precision [35,52] or error detection [26,38,53], and classical algorithms for instantaneous quantum polynomial (IQP) circuits [26,27] and random quantum circuits [28–36]. To date, however, there are few rigorous results on the computational power of general quantum circuits with low noise rates but without error correction.

In this work, we provide a classical algorithm for computing expectation values in *any* noisy quantum circuit on *most* input states. The restriction to "most" input states is fundamental, since for a fixed input state and a sufficiently low noise rate, one can immediately perform quantum error correction. To this end, with a high probability over input states drawn from an ensemble, our algorithm succeeds in computing the expectation value of any observable (to within small error) for any noisy quantum circuit. We emphasize that our restriction to ensembles of input states is relatively weak—a wide range of ensembles are allowed, including the computational basis states. The algorithm runs in either polynomial or quasipolynomial time, depending on the observable and noise model of interest.

To compute the expectation value, our algorithm simulates the Heisenberg time evolution of the observable in the Pauli basis. Building upon seminal recent algorithms for random quantum circuits [28–36] and many-body dynamics [46–51,54–59], the key insight underlying our algorithm is

*These authors contributed equally to this work.

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

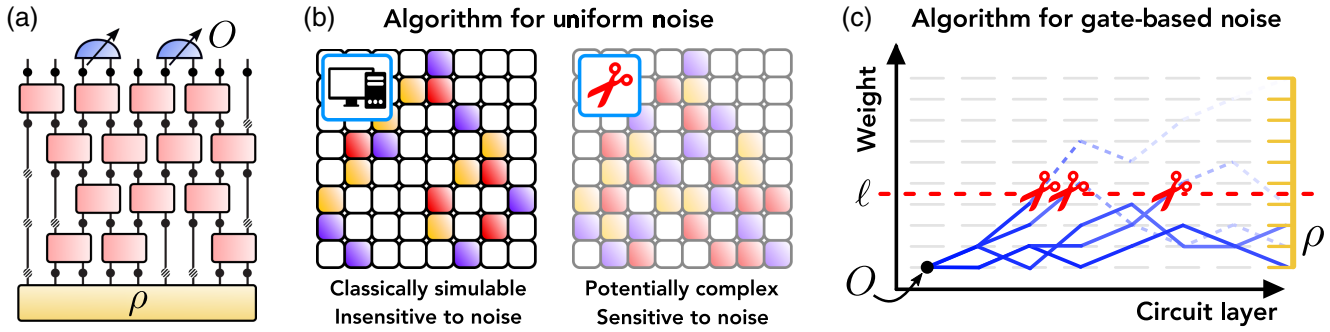


FIG. 1. (a) Schematic of a noisy quantum circuit. The input state ρ is acted on by a circuit of arbitrary two-qubit gates (pink) and local noise channels (dots), and concludes with measurement of an observable O. In the gate-based noise model, the noise acts on a qubit only when a gate is performed (black dots). In the uniform noise model, even when the qubit is idle, noise can occur (dashed dots). (b) For circuits with uniform noise, our classical algorithm decomposes the expectation value of O as a sum of Pauli paths. We depict each path by a space-time grid, where each square denotes whether the path is the identity (white) or X, Y, or Z (red, yellow, or purple, respectively) at that qubit and circuit layer. Our algorithm computes the sum of all low-weight paths (left), and truncates high-weight paths (right) since they are strongly damped by noise (fading). (c) For circuits with gate-based noise, our algorithm instead simulates the Heisenberg time evolution of O within the subspace of low-weight Pauli operators. That is, at each layer, we truncate all Pauli operators with weight above a threshold ℓ (dashed red line). The plot depicts the weight of various components of the time-evolved operator (blue lines), as they are acted on by noise (fading) and, potentially, truncated by our algorithm (red scissors).

the close connection between sets of Pauli operators that are hard to classically simulate and those that are strongly affected by noise. By carefully truncating such high-weight Pauli operators, we achieve a provably efficient and accurate simulation. We emphasize that, despite the prevalence of Pauli truncation methods for nearly six years to date [28], our work provides the first rigorous proof that such methods can succeed for nonrandom circuits. This is enabled by substantially new proof techniques compared to any previous work.

Our results have wide-ranging consequences for quantum experiments. First, we show that, for noise rates that do not improve as the number of qubits increases, any strategy for quantum error mitigation [60–62] can scale efficiently [63–66] only for quantum circuits that are easy to classically simulate. Like our main results, this statement holds for any quantum circuit on most input states. Second, we build upon our results to construct a simple and strict test for whether a given quantum circuit can exhibit quantum advantage. We find that any circuit exhibiting advantage on most input states must be macroscopically sensitive to noise. Finally, we provide extensions of our algorithm to sampling from circuits with output distributions that anticoncentrate, as well as to computing expectation values in random quantum circuits with nonunital noise [36,67], which improves in both run-time and generality compared to existing algorithms [36].

A. Classical algorithm for noisy quantum circuits

Consider a noisy quantum circuit on n qubits, of the form

$$\mathcal{C}\{\rho\} = \mathcal{D}_0\{U_1\mathcal{D}_1\{\ldots U_d\mathcal{D}_d\{\rho\}U_d^\dagger\ldots\}U_1^\dagger\}, \qquad (1)$$

where ρ is the input state, U_t are depth-1 unitaries of arbitrary two-qubit gates, and $\mathcal{D}_t$ are local depolarizing noise channels. Here, d is the circuit depth, and $t = 1, \ldots, d$ indexes the circuit layers in reverse order. After the circuit is applied, we are interested in either computing the expectation value of an observable O, $\mathrm{tr}(\mathcal{C}\{\rho\}O)$, or sampling from the outcomes when $\mathcal{C}\{\rho\}$ is measured in the computational basis.

We consider two archetypal noise models [Fig. 1(a)]: uniform noise and gate-based noise. In the *uniform noise* case, every qubit is affected by noise at every circuit layer. That is, we take $\mathcal{D}_t = \otimes_{j=1}^n \mathcal{D}_j$, where $\mathcal{D}_j\{\rho\} = e^{-\gamma}\rho + (1-e^{-\gamma})\mathrm{tr}_j(\rho)$ is a local depolarizing channel of strength γ. This noise model is relevant for many current experiments, in which idle errors affect qubits even when they are not involved in gates. However, the uniform noise model allows computation only up to a depth logarithmic in the number of qubits, after which the system is close to the maximally mixed state [13,68]. To this end, we also consider the so-called *gate-based noise* model, in which qubits are affected by noise only when they participate in nonidentity gates, $\mathcal{D}_t = \otimes_{j\in U_t}\mathcal{D}_j$. This model allows for "fresh qubits" unaffected by idle errors to be inserted at any circuit layer, which is the standard setting for proofs of fault-tolerant quantum computation [9–12,69]. In both noise models, we incorporate readout noise by taking the final noise channel $\mathcal{D}_0$ to act on all qubits. Finally, although we focus on depolarizing noise for the sake of simplicity, our results immediately generalize to any single-qubit noise channel in the so-called depolarizing class. This includes both biased and inhomogeneous noise channels (see Appendix F for further details).

The central framework of our approach is the decomposition of the Heisenberg time-evolved observable O in the Pauli basis. Crucially, under uniform noise, a Pauli

TABLE I. Known results on the computational complexity of various classes of noisy quantum circuits, taking the standard limits $d, \varepsilon^{-1} = \mathrm{poly}(n)$ and $\gamma^{-1} = \mathcal{O}(1)$. In the first three columns, expectation values are with respect to observables that are sums of polynomially many Pauli operators; each run-time increases to quasipolynomial otherwise. In the final two columns, entries denote hardness results arising from quantum error correction [9–13]. We use $\widetilde{\mathrm{QNC}}_1$ to denote the class of quantum circuits with depth $d = \mathcal{O}(\log n/\mathrm{poly}\log\log n)$ [13], which is just below logarithmic. We note that simpler classical simulations may be obtained for circuits with specific depths [27] or geometric locality, as we show, for example, in Theorem E2.

	IQP circuits	Random circuits	Any circuit, most inputs (uniform noise)	Any circuit, most inputs (gate noise)	Any circuit, any input (uniform noise)	Any circuit, any input (gate noise)
Expectation values	poly(n) Ref. [26]	poly(n) Refs. [28,31]	poly(n) Theorem 1	quasipoly(n) Theorem 2	$\widetilde{\mathrm{QNC}}_1$ Ref. [13]	BQP Refs. [9–12]
Sampling ($*$: with anticoncentration)	quasipoly(n)* Ref. [26]	poly(n)* Ref. [31]	quasipoly(n)* Theorem E1	quasipoly(n)* Theorem E1	Samp$\widetilde{\mathrm{QNC}}_1$ Ref. [13]	SampBQP Refs. [9–12]

operator P with weight $w[P]$ is damped by an amount $e^{-\gamma w[P]}$, where the weight w is defined as the number of nonidentity elements in the operator. High-weight Pauli operators are thus almost entirely damped by noise, while low-weight operators are less affected. Since there are only polynomially many, $\mathcal{O}((3n)^w/w!)$, low-weight operators, this suggests that one might be able to efficiently approximate a time-evolved observable by only keeping track of such operators. This intuition has motivated a range of seminal algorithms for classically simulating noisy quantum circuits by truncating Pauli operators of high weight [26,28–34,45–50,54–59]. However, whether it is possible to turn this intuition into a rigorous classical algorithm for general circuits has remained an essential open question.

There are two challenges. First, although each individual high-weight operator is strongly damped by noise, there are, in principle, exponentially many such operators. To address this, existing classical algorithms [28–35] utilize a strongly simplifying property of *random circuits*: In a random circuit, different Pauli paths *do not coherently interfere* on average [70]. This enables simple bounds on the error associated to truncated paths. However, this property does not extend to general circuits. Second, for circuits with gate-based noise, the damping of each Pauli operator is given by $w_{U_t}[P]$, the number of nonidentity elements of P that are acted upon by gates in U_t. (Specifically, $w_{U_t}[P]$ counts the number of qubits that are both in the support of P and acted on by nonidentity gates in U_t.) However, a high-weight operator might be involved only in a small number of gates in any given circuit layer. Thus, in combination with the first challenge, any rigorous approach for bounding the truncation error *must* keep track of the *history* of operators across many circuit layers.

In what follows, we will begin by stating our main results (Theorems 1 and 2; Table I). In particular, we provide one classical algorithm for circuits with uniform noise, which runs in polynomial time for a large class of observables, and a second classical algorithm for circuits with gate-based noise, which runs in quasipolynomial time. We will then describe our algorithms, with a focus on delineating the key techniques and insights that allow us to overcome the two aforementioned challenges. As aforementioned, both of our algorithms apply to any noisy quantum circuit on most input states. We term the allowed ensembles of input states *low-average* ensembles, which we define immediately following the theorems below.

Theorem 1 (Polynomial-time algorithm for quantum circuits with uniform noise). Consider any quantum circuit $\mathcal{C}$ with local depolarizing noise of strength γ on each qubit at each circuit layer, any normalized observable O, and a low-average ensemble of states $\mathcal{E} = \{\rho\}$. Assume the Pauli coefficients of O and ρ can be efficiently computed. For circuits with uniform noise, there is a classical algorithm that computes the expectation values, $\mathrm{tr}(\mathcal{C}\{\rho\}O)$, in time:

$$\mathrm{poly}(n)\cdot(1/\varepsilon)^{\mathcal{O}(\log(1/\gamma)/\gamma^2)}, \tag{2}$$

if O is a sum of polynomially many Paulis,

$$n^{\mathcal{O}(\log(1/\gamma)/\gamma)+\mathcal{O}(\log(1/\varepsilon)/\gamma d)}\cdot(1/\varepsilon)^{\mathcal{O}(\log(1/\gamma)/\gamma^2)}, \tag{3}$$

for any O, with root-mean-square error ε over the ensemble $\mathcal{E}$.

Theorem 2 (Quasipolynomial time algorithm for quantum circuits with gate-based noise). For circuits with gate-based depolarizing noise of strength γ, there is a classical algorithm that runs in quasipolynomial time,

$$\mathcal{O}(d\cdot n^{(1/\gamma)\log(\sqrt{d+1}/\varepsilon)-1}), \tag{4}$$

for any normalized O, with root-mean-square error ε over the ensemble $\mathcal{E}$.

Before proceeding to our algorithms, a few brief comments are in order.

First, we say that an observable O is normalized if its normalized Frobenius norm is equal to one: $||O||_F^2 \equiv (1/2^n)\mathrm{tr}(O^\dagger O) = 1$. For example, this is true for any Pauli operator. If an observable is not normalized, the errors ε can be simply rescaled by the observable's norm.

ALGORITHM 1.

Classical algorithm for uniform noise

Input: State ρ, circuit $\mathcal{C}$, observable O; threshold ℓ
Output: Approximation of $\mathrm{tr}(\mathcal{C}\{\rho\}O)$
1: Enumerate Pauli paths $\vec{P}$ with $\sum_t w[P_t] \le \ell$
2: Compute amplitudes $c_{\vec{P}} = \prod_t a^{(t)}_{P_{t-1}P_t}$
3: Apply noise $c_{\vec{P}} \leftarrow e^{-\gamma \sum_t w[P_t]} \cdot c_{\vec{P}}$
4: **return** $\sum_{\vec{P}} c_{\vec{P}} \mathrm{tr}(\rho P_d)$

ALGORITHM 2.

Classical algorithm for gate-based noise

Input: State ρ, circuit $\mathcal{C}$, observable O; threshold ℓ
Output: Approximation of $\mathrm{tr}(\mathcal{C}\{\rho\}O)$
1: Initialize $c_P = e^{-\gamma w[P]}\mathrm{tr}(OP)/2^n$ for $w[P] \le \ell$
2: **for** $t = 1$ to d **do**
3: Update via $c_Q \leftarrow e^{-\gamma w_{U_t}[Q]} \sum_P a^{(t)}_{PQ} \cdot c_P$
4: **return** $\sum_P c_P \mathrm{tr}(\rho P)$

Second, while our results apply to any noisy quantum circuit, they still involve a small amount of randomness via the ensemble of input states. In particular, our metric of success is the root-mean-square error ε over a low-average ensemble of quantum states, which we define as any ensemble, $\mathcal{E} = \{\rho\}$, whose mixture is close to the maximally mixed state, $||(1/|\mathcal{E}|)\sum_\rho \rho||_\infty \le c/2^n$ for any $c = \mathcal{O}(1)$ (see Appendix B). As a simple example, any complete basis of pure states, such as the computational basis, forms a low-average ensemble (with $c = 1$). Intuitively, our attention to average-case performance excludes quantum error correction because syndrome qubits will be initialized randomly on average. Experiments with random input states are of significant interest in several settings, including in quantum learning [71,72] and quantum many-body dynamics [73–75].

Third, if preferred, our bounds on the root-mean-square error ε can be easily translated to probability bounds. That is, for any $\delta, \tilde{\varepsilon}$, our algorithms succeed in computing the expectation value $\mathrm{tr}(\mathcal{C}\{\rho\}O)$ to error $\tilde{\varepsilon}$ with high probability $1-\delta$ over input states ρ drawn from $\mathcal{E}$. This follows immediately from Theorems 1 and 2 by an application of Markov's inequality, setting $\varepsilon = \tilde{\varepsilon}\sqrt{\delta}$. The run-time of the algorithms remains polynomial and quasi-polynomial, respectively, for any inverse polynomial $\tilde{\varepsilon}$ and δ. Hence, we say that our algorithms succeed on "most" input states.

B. Classical algorithm for quantum circuits with uniform noise

Our algorithm for uniform noise (Algorithm 1) estimates the expectation value, $\mathrm{tr}(\mathcal{C}\{\rho\}O) = \mathrm{tr}(\rho\mathcal{C}^\dagger\{O\})$, by computing an approximation $\tilde{O}$ of the Heisenberg time-evolved observable $\mathcal{C}^\dagger\{O\}$. This approximation is given by the sum of all low-weight *Pauli paths* that contribute to $\mathcal{C}^\dagger\{O\}$. Here, a Pauli path $\vec{P} = (P_0, \ldots, P_d)$ is a sequence of Pauli operators corresponding to each layer of the circuit [Fig. 1(b)], and the exact time-evolved operator can immediately be expressed as a sum over Pauli paths,

$$\mathcal{C}^\dagger\{O\} = \sum_{\vec{P}} e^{-\gamma w[\vec{P}]} \cdot \frac{1}{2^n}\mathrm{tr}(OP_0) \cdot \left(\prod_{t=1}^{d} a^{(t)}_{P_{t-1}P_t}\right) \cdot P_d, \qquad (5)$$

with coefficients given by the transition amplitudes, $a^{(t)}_{PQ} = (1/2^n)\mathrm{tr}(QU_t^\dagger P U_t)$, at each layer t. Each path is damped by a factor $e^{-\gamma w[\vec{P}]}$ due to noise, where $w[\vec{P}] = \sum_t w[P_t]$ is the sum of the individual weights at each circuit layer.

In more detail, we can derive the above formula as follows. Prior to any time evolution, the observable O can be decomposed in the Pauli basis as $O = \sum_{P_0}(1/2^n)\mathrm{tr}(OP_0)P_0$. This sets the "initial condition" for the Pauli path. Next, we can time evolve O through the final circuit layer U_1. This produces the time-evolved operator, $U_1 O U_1^\dagger = \sum_{P_0} \mathrm{tr}(OP_0) U_1 P_0 U_1^\dagger$. As before, we can decompose this operator in the Pauli basis to find $U_1 O U_1^\dagger = \sum_{P_0}\sum_{P_1}(1/2^n)\mathrm{tr}(OP_0)\cdot(1/2^n)\mathrm{tr}(P_1 U_1^\dagger P_0 U_1)\cdot P_1$. Here $a^{(1)}_{P_0P_1} \equiv (1/2^n)\mathrm{tr}(P_1 U_1^\dagger P_0 U_1)$ is a *transition amplitude* from the Pauli operator P_0 to P_1 under the first layer of the circuit. Proceeding in this manner d times, we obtain Eq. (5), aside from the exponential factor involving γ.

To derive the exponential factor, we consider the behavior of the same circuit in the presence of local depolarizing noise. As discussed earlier, the local depolarizing noise channel has a very simple action in the Pauli basis, $\mathcal{D}\{P\} = e^{-\gamma w[P]}P$. Applying this channel at each layer of the circuit above yields an additional overall multiplicative factor of $e^{-\gamma w[P_0]}e^{-\gamma w[P_1]}\cdots e^{-\gamma w[P_d]}$, which we abbreviate as $e^{-\gamma w[\vec{P}]}$. This completes the derivation of Eq. (5).

Our classical algorithm's approximation for $\mathcal{C}^\dagger\{O\}$ is given by the sum over all Pauli paths with weight below a chosen threshold ℓ, i.e., all $\vec{P}$ with $w[\vec{P}] \le \ell$:

$$\tilde{O} = \sum_{\vec{P}: w[\vec{P}] \le \ell} e^{-\gamma w[\vec{P}]} \left(\prod_{t=1}^{d} a^{(t)}_{P_{t-1}P_t}\right) P_d. \qquad (6)$$

From this approximation, we estimate the expectation value of interest as $\mathrm{tr}(\rho\tilde{O})$.

We hasten to emphasize that our algorithm is in fact identical to those from random circuits [28,31]. To this end, our ability to rigorously extend the algorithm to general noisy quantum circuits owes to substantial improvements in a number of key proof techniques. In particular, to bound the error in the expectation values, we first prove the following simple lemma.

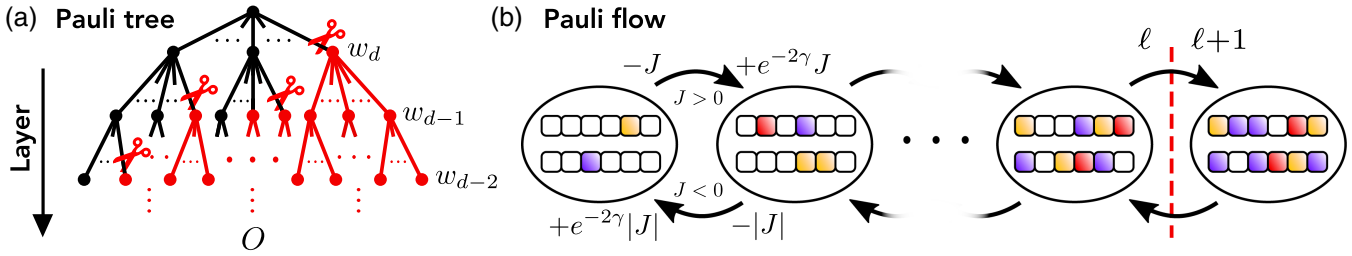


FIG. 2. (a) The Pauli tree framework used to prove Theorem 1, which groups Pauli paths according to their weight w_t at each circuit layer t. Our algorithm truncates paths with a summed weight above ℓ. We group these truncations into $\binom{\ell}{d}$ individual truncations (red scissors), which enables a tight bound on the algorithm's error. In the example shown, the branchings at each node correspond to weights $w = 1, \ldots, 5$ and we set $\ell = 6$, $d = 2$. (b) To prove Theorem 2, we analyze the flow of Pauli operators from one weight to another under circuit gates. Here, each bubble depicts the set of Pauli operators of a given weight, and arrows indicate flow from one weight to another. We show that this flow is lossy in the presence of gate-based noise, in the sense that a decrease $-J$ in support at weight w can increase the support at weight $w+1$ by at most $e^{-2\gamma}J$. This leads to our bound on the cumulative operator weight distribution in Lemma 2.

Lemma 1. Consider a low-average ensemble $\mathcal{E} = \{\rho\}$ and two observables O and $\tilde{O}$. The root-mean-square difference between the expectation value of O and $\tilde{O}$ is less than $||O - \tilde{O}||_F$.

Thus, if we are able to bound the Frobenius norm of the sum of truncated Pauli paths, then we can bound the root-mean-square error in our algorithm.

To bound the norm, we group the exponentially many Pauli paths according to their weight at each circuit layer. This organization is naturally viewed via a tree structure [Fig. 2(a)]. Each layer t of the tree corresponds to a layer of the quantum circuit, ordered from $t = d, \ldots, 1$. At every layer, each vertex branches into up to n additional vertices, which correspond to Pauli paths with weight $w_t = 1, \ldots, n$ at that layer. In this manner, each Pauli path $\vec{P}$ is uniquely associated with a particular downward sequence of vertices, $(w_d), (w_d, w_{d-1}), \ldots, (w_d, \ldots, w_0)$, through the tree.

The key benefits of the Pauli tree are that (i) the paths associated with each vertex are damped by at least $e^{-\gamma \sum_{t'=t}^{d} w_{t'}}$ due to noise and (ii) the tree organization naturally enables us to bound the norm of sums of Pauli paths, by using the orthogonality of Pauli operators at each circuit layer. We now outline how these properties enable a tight bound on the error of Algorithm 2. Let $\delta O(w, t)$ denote the sum of all Pauli paths from layer 0 to t with weight above w. We are interested in the Frobenius norm of $\delta O(\ell, d) = \mathcal{C}^\dagger\{O\} - \tilde{O}$, which bounds the root-mean-square error of Algorithm 1 via Lemma 1. To analyze the Frobenius norm, we leverage properties (i) and (ii) above to decompose it as a sum of contributions from each weight w_d at the final circuit layer, $||\delta O(\ell, d)||_F^2 = \sum_{w_d} ||\mathcal{P}_{w_d}\{\delta O(\ell, d)\}||_F^2$, where $\mathcal{P}_{w_d}$ projects onto Pauli operators of weight w_d. We can associate each term in the sum with a vertex (w_d) in the top layer of the Pauli tree. We can then bound the contribution of each vertex as $||\mathcal{P}_{w_d}\{\delta O(\ell, d)\}||_F^2 \leq e^{-2\gamma w_d}|\times |\delta O(\ell - w_d, d-1)||_F^2$, since any Pauli path with weight w_d at the final layer, and total weight above ℓ, must have total weight above $\ell - w_d$ on the preceding layers (see Appendix C for details). Combined with the first equality, we see that the norm of high-weight Pauli paths at layer d is bounded by a sum of analogous norms at layer $d-1$.

To complete our proof, we recursively apply this bound down each layer of the Pauli tree. That is, we bound the contribution of each top layer vertex (w_d) as a sum of contributions of vertices at the following layer, (w_d, w_{d-1}). Each such contribution arises from Pauli paths with weight above $\ell - w_d - w_{d-1}$ on layers up to $d-2$, and incurs an associated damping, $e^{-2\gamma(w_d + w_{d-1})}$. We then bound these contributions in terms of vertices at the next layer, (w_d, w_{d-1}, w_{d-2}), and so on. We halt each sequence of recursions once the summed weight, $\sum_{t'=t}^{d} w_{t'}$, surpasses $\ell - t$; from here, one can directly bound the remaining contribution by $e^{-2\gamma(\ell+1)}$, since every Pauli path from layer 0 to $t-1$ has weight at least t [76]. In effect, this recursive process groups the truncated Pauli paths according to their earliest subsequence of weights that sum to above $\ell - t$. Each grouping can be viewed as a single collective truncation by the algorithm [Fig. 2(a)]. The total error of the algorithm is determined by the number of such truncations, which a simple counting argument shows is $\binom{\ell}{d}$. This leads to a total error $\binom{\ell}{d} e^{-2\gamma(\ell+1)}$. We refer to Appendix C for full details.

From this error bound, we determine the run-time of Algorithm 1 by setting $\ell \approx \gamma^{-1} d + \gamma^{-1} \log(1/\varepsilon)$ to achieve a desired error ε. For circuits with uniform noise, the maximum nontrivial depth is $d = \mathcal{O}(\gamma^{-1} \log(1/\varepsilon))$, since after this depth, expectation values become ε-close to zero and the circuit can be trivially simulated [77]. The run-time of our classical algorithm is determined by the number of Pauli paths with weight below ℓ. This is upper bounded by $n^{\ell/d} \cdot 2^{\mathcal{O}(\ell)}$ for any O [31], and $m \cdot 2^{\mathcal{O}(\ell)}$ when O is a sum of m Pauli operators (see Appendix C). Plugging in the above values of ℓ and d yields Theorem 1.

C. Classical algorithm for quantum circuits with gate-based noise

Although Algorithm 1 is quite efficient for quantum circuits with uniform noise, where the depth d is at most logarithmic, this efficiency does not extend to circuits with gate-based noise, where the depth d may be large. In such circuits, the run-time of Algorithm 1 diverges exponentially in d owing to the $\binom{\ell}{d}$ truncations.

To address this, our algorithm for gate-based noise (Algorithm 2) performs only a single truncation at each circuit layer, of all Pauli operators with weight greater than ℓ. In particular, at each layer t we store an approximation of the time-evolved operator within the subspace of low-weight Pauli operators:

$$\tilde{O}^{(t)} = \sum_{P:\, w[P]\le \ell} \tilde{c}_P^{(t)} P. \tag{7}$$

The coefficients are updated from layer to layer via standard Heisenberg time evolution:

$$\tilde{c}_Q^{(t)} = e^{-\gamma w_{U_t}[Q]} \sum_{P:\, w[P]\le \ell} a_{PQ}^{(t)} \tilde{c}_P^{(t-1)}. \tag{8}$$

At the final circuit layer, we compute the expectation value $\mathrm{tr}(\rho \tilde{O}^{(d)})$.

By performing only $d+1$ truncations in total, Algorithm 2 avoids the exponential blowup of Algorithm 1 at large depths. In particular, if we denote the truncated operators at layer t as $\delta O^{(t)} \equiv \mathcal{P}_{>\ell}\{\mathcal{D}_t\{U_t^\dagger \tilde{O}^{(t-1)} U_t\}\}$, where $\mathcal{P}_{>\ell} = \sum_{w=\ell+1}^{n} \mathcal{P}_w$, then we immediately have $||\delta O||_F^2 \le (d+1)\sum_{t=0}^{d} ||\delta O^{(t)}||_F^2$ by the Cauchy-Schwarz inequality. The trade-off is a moderate increase in run-time compared to Algorithm 1 at the same value of ℓ. In particular, Algorithm 2 performs d updates on $\mathcal{O}(n^\ell)$ low-weight Pauli coefficients, and thus has run-time $\mathcal{O}(dn^\ell)$; see Appendix D for full details.

The central difficulty in proving Theorem 2 is to bound the truncation norms $||\delta O^{(t)}||_F$ within the gate-based noise model. This would be straightforward within the uniform noise model, since a single application of uniform depolarizing noise damps the norm of any operator with weight above ℓ by at least $e^{-\gamma(\ell+1)}$. However, as aforementioned, it is not clear whether such a bound applies to the gate-based noise model, where the effect of noise can accumulate gradually over many circuit layers.

To show that the bound indeed extends to circuits with gate-based noise, our core idea is to track the *flow* of operator norm from one weight to another under each circuit gate [Fig. 2(b)]. Intuitively, each two-qubit gate can transfer operator norm from a weight w to adjacent weights $w \pm 1$. Thus, for a component of the operator to reach a high weight, it must participate in a high number of circuit gates, and thereby accrue a large damping due to noise. Despite its simplicity, formalizing this intuition can be remarkably subtle, due to processes wherein an operator component grows to high weight, then shrinks back to low weight, and then grows to high weight again. Our focus on the flow of the operator norm allows us to simultaneously capture damping during the growth in operator weight, while avoiding overcounting due to shrinking and regrowth processes.

In more detail, to quantify the norm of a time-evolved operator $O^{(g)}$ at weight w, we leverage an object of recent interest in many-body physics: the *operator weight distribution* [78–81]:

$$P^{(g)}(w) = \frac{1}{2^n}\mathrm{tr}(O^{(g)} \cdot \mathcal{P}_w\{O^{(g)}\}). \tag{9}$$

Here, we track the evolution through each circuit gate g instead of each circuit layer t. In keeping with our intuition, we show that the weight distribution changes in a "local" manner when any unitary two-qubit gate g is applied,

$$P^{(g)}(w) = P^{(g-1)}(w) + J^{(g)}(w) - J^{(g)}(w+1), \tag{10}$$

where $J^{(g)}(w)$ measures the flow in operator norm from weight $w-1$ to w under g. The crucial insight of our proof is to show that this flow becomes *lossy* for quantum circuits with gate-based noise, in the sense that

$$P^{(g)}(w) \le P^{(g-1)}(w) + \{1, e^{-2\gamma}\} \cdot J^{(g)}(w) - \{1, e^{-2\gamma}\} \cdot J^{(g)}(w+1), \tag{11}$$

where the bound holds for any choices of the constants 1 or $e^{-2\gamma}$ in each of the terms. The strongest bound is obtained by taking $e^{-2\gamma}$ when the term is positive, and 1 when negative. In the former case, the bound shows that increases in operator norm due to flows inward to weight w are damped by $e^{-2\gamma}$ due to noise; in the latter case, decreases in the norm due to outward flows are not damped [Fig. 2(b)].

To bound the truncation error, we apply Eq. (11) for every circuit gate g and sum over weights $w > \ell$. The currents between high weights cancel, and we find that the total norm above weight ℓ is less than the *net current* from ℓ to $\ell+1$, $\sum_g J^{(g)}(\ell)$, damped by $e^{-2\gamma}$. By bounding the norm in terms of the net current, we naturally avoid subtleties due to shrinking and regrowth processes. Furthermore, the net current from weight ℓ to $\ell+1$ is necessarily bounded by the maximum norm at weight ℓ. This is, by a similar inequality, less than the net current from $\ell-1$ to ℓ damped $e^{-2\gamma}$. Iterating ℓ times, we find that the total norm above weight ℓ is less than $e^{-2\gamma(\ell+1)}$ for any circuit with gate-based noise (see Appendix D for full details).

Lemma 2 (Upper bound on weight distributions in noisy quantum circuits). Consider the time evolution $O^{(g)}$ of any normalized observable O in any quantum

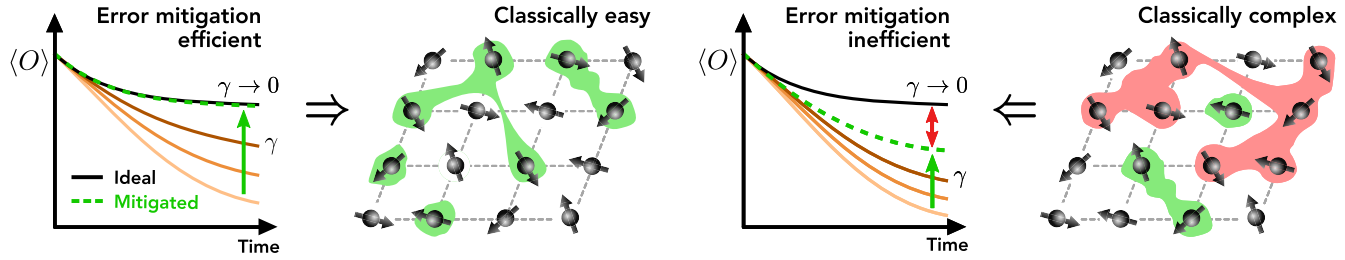


FIG. 3. Schematic of our results' implications for quantum error mitigation. For concreteness, we depict a specific error mitigation strategy, zero noise extrapolation, in which one measures an expectation value $\langle O\rangle$ for several different noise rates (orange lines), and performs an extrapolation (green arrow and dashed line) to estimate the ideal expectation value (black line). (a) If error mitigation succeeds in recovering the ideal expectation value, then the expectation value must be dominated by low-weight Pauli operators (green). Thus, our classical algorithm can also compute the ideal expectation value. (b) On the other hand, if the ideal expectation value is hard to compute classically, then it must contain contributions from high-weight Pauli operators (red). Error mitigation cannot capture these contributions, since they are exponentially suppressed by noise. Thus, the extrapolated expectation value necessarily differs from the ideal value (red arrows).

circuit with gate-based noise. The cumulative weight distribution is less than

$$\sum_{w'=w+1}^{\infty} P^{(g)}(w') \le e^{-2\gamma(w+1)} \tag{12}$$

for all γ, w, t. This also holds if we replace $O^{(g)}$ with our classical approximation $\tilde{O}^{(g)}$.

Applying the lemma Algorithm 2, we find that the total truncation error is less than $(d+1)e^{-2\gamma(\ell+1)}$. Setting $\ell \approx \gamma^{-1}\log(\sqrt{d}/\varepsilon)$ to ensure a small root-mean-square error gives Theorem 2. We remark that our proof of Lemma 2, as outlined above, introduces substantially new techniques compared to any previous Pauli truncation algorithm that we are aware of.

D. Applications and Implications

Let us now consider the implications of our algorithms. We summarize a few of the most prominent ones below, and direct the interested reader to the Appendixes E–L for details and additional results.

Lower bound on quantum error mitigation. Error mitigation seeks to estimate the output of an ideal quantum circuit via experiments on noisy circuits [60–62,82–87]. While the sampling overhead for certain mitigation strategies (such as probabilistic error cancellation) is explicitly exponential in the number of qubits [60–62], $\sim e^{\gamma nd}$, for other strategies (such as zero noise extrapolation), an exponential scaling is known only for certain worst-case circuits [63–65]. This has raised an intriguing open question: Can error mitigation scale efficiently for any quantum circuits of practical interest [88–90]?

Unfortunately, our results imply a somewhat negative answer (Fig. 3). For a large class of mitigation strategies, any circuit that can be error mitigated in polynomial time on most input states can also be classically simulated in polynomial time on most input states (Appendix I).

Corollary 1 (Lower bound on error mitigation). Given an ideal circuit $\mathcal{C}$, Pauli observable O, and a low-average ensemble $\mathcal{E} = \{\rho\}$. Suppose an error mitigation strategy proceeds by measuring O on any set of noisy circuits applied to ρ. For any $\varepsilon = 1/\mathrm{poly}(n)$, if error mitigation can estimate $\mathrm{tr}(\mathcal{C}\{\rho\}O)$ to root-mean-square error $\varepsilon/2$ in polynomial time, there is a classical algorithm to compute $\mathrm{tr}(\mathcal{C}\{\rho\}O)$ to root-mean-square error ε in polynomial time.

The reason is simple. If mitigation can recover an ideal expectation value from noisy circuits, then one can also compute the value classically, by substituting our algorithm for the circuit results. At an intuitive level, the properties that make error mitigation more efficient (having a signal dominated by low-weight Paulis) also make the signal easier to simulate. We emphasize that this discussion applies only for noise rates that are constant with respect to n; for smaller noise rates, $\gamma = \tilde{\mathcal{O}}(1/n)$, a circuit can be both complex and efficiently mitigable.

Noisy DQC1 $\subseteq$ BPP. Our results extend to circuits with a fixed input state within one scenario: when the state is highly mixed. This follows directly from Theorems 1 and 2 because the ensemble composed of a single highly mixed input state is low average (Appendix K). Experiments with highly mixed input states are commonplace in several quantum technologies, including nuclear magnetic resonance spectroscopy [91] and solid-state quantum simulators [92,93]. Formally, this scenario is captured by the deterministic quantum computing with one clean qubit (DQC1) model of quantum computation [94]. This model encompasses quantum experiments that are performed by (1) preparing an initial state, $|0\rangle\langle 0| \otimes \mathbb{1}^{n-1}/2^{n-1}$, which is clean on the first qubit and mixed on all others, (2) applying any quantum circuit to this initial state, and (3) measuring the first qubit in the Z basis. We prove that any *noisy* DQC1 circuit can be classically simulated in polynomial time.

Sampling from anticoncentrated circuits. Our algorithms can also be used to sample from noisy quantum circuits,

provided that the output distribution of the circuit *anticoncentrates*. Anticoncentration refers to a property of the output distribution that one is sampling from, namely that $\sum_x p(x)^2 = \text{poly}(n)/2^n$. Here, $p(x)$ is the output probability of a bit string $x \in \{0,1\}^n$. Intuitively, anticoncentration guarantees that the output distribution has relatively even support across all bit strings x. Anticoncentration is common in many output distributions [95–101], and is used frequently as an ingredient for proofs of both the classical hardness [95,102,103] and easiness [31] of sampling tasks. Our algorithm for sampling from noisy quantum circuits utilizes a standard reduction from noisy circuit sampling to computing noisy expectation values, which is valid in the presence of anticoncentration [26]. We achieve sampling to within small total variational distance in quasipolynomial time (Appendix E).

Random circuits with nonunital noise. Recent works have asked [36,67], can *random* quantum circuits with *nonunital* noise, such as spontaneous emission, escape the classical algorithms known for circuits with unital noise? This was answered in part by Ref. [36], which provides an algorithm for expectation values in random circuits with nonunital noise, that runs in quasipolynomial time for geometrically local circuits, and exponential time (in ε^{-1}) for general circuit architectures. Leveraging our proof method for Theorem 1, we show that Algorithm 1 computes expectation values in *polynomial* time for *any* random circuit with nonunital noise (Appendix H). We note that a direct extension of Theorem 1 to nonrandom circuits with nonunital noise is not possible, since for general quantum circuits with nonunital noise one can perform fault-tolerant computation [14].

Quantum advantage beyond sampling. While the focus of our work has been on noisy quantum circuits, our results also have broader implications in the search for near-term experiments that exhibit a quantum advantage. Namely, by inverting our Theorem 2, we show that any quantum circuit that is exponentially hard to simulate classically *must fail* in the presence of even a macroscopically small noise rate, $\gamma = \tilde{\Omega}(1/n)$ (Appendix J).

Corollary 2 (Classically hard quantum circuits must be highly sensitive to noise). Given any ideal quantum circuit $\mathcal{C}$ on n qubits with depth $d = \text{poly}(n)$, any normalized observable O, and any low-average ensemble $\mathcal{E} = \{\rho\}$. Suppose that the run-time of any classical algorithm to compute the expectation values, $\text{tr}(\mathcal{C}\{\rho\}O)$, to root-mean-square error ε is lower bounded by $\chi(n,\varepsilon)$. Then any noisy implementation of $\mathcal{C}$ *must only succeed* in computing $\text{tr}(\mathcal{C}\{\rho\}O)$ to root-mean-square error ε for noise rates:

$$\gamma \leq \mathcal{O}\left(\frac{\log(n)\log(n/\varepsilon)}{\log\left(\chi(n,\varepsilon)\right)}\right). \tag{13}$$

In particular, if the classical run-time is exponential in n, the noisy implementation must succeed only for macroscopically small noise rates, $\gamma = \mathcal{O}\left(\log^2(n)/n\right)$.

Our proof follows by an immediate triangle inequality: If a quantum circuit with noise rate γ produces an expectation value that is ε away from the ideal expectation value, then we can classically simulate the ideal expectation value to error 2ε by simulating the noisy circuit to precision ε. From Theorem 2, this can be done in time $\chi = n^{\mathcal{O}((1/\gamma)\log(d/\varepsilon))}$. Taking the logarithm of both sides, setting $d = \text{poly}(n)$, and solving for γ yields Corollary 2.

We propose that this criteria can function as a simple and rigorous "test" for whether a given quantum circuit can feature an advantage over classical computation. Thus far, demonstrations of quantum advantage have been restricted to sampling experiments [95,104], which are inherently highly sensitive to noise [105,106]. When moving beyond sampling, designing experiments whose expectation values are macroscopically sensitive to noise in a nontrivial manner is challenging. This is highlighted by recent experiments on IBM's quantum processor [89], which, despite involving many qubits and high gate fidelities, were rapidly simulated classically by follow-up works [107–112]. In light of our work, this can be understood as a consequence of the experiments' insensitivity to noise [113].

More broadly, our criteria are motivated by a recent surge of numerical [46,49,50,54–59] and theoretical [114–116] studies of many-body quantum dynamics, which show that local observables in strongly interacting quantum dynamics can often be efficiently simulated using Pauli-truncation-based methods. This occurs even when the dynamics appear difficult by all conventional complexity measures. When performing a quantum experiment, one might seek a simple benchmark for whether the experiment is susceptible to such classical simulations. Following Corollary 2, we propose that the experiments' sensitivity to noise provides precisely such a unifying benchmark.

Finally, it should be noted that the converse of Corollary 2 does not hold. Namely, one can design experiments that are highly sensitive to noise, such as those involving a many-body observable [89] or a Loschmidt echo [117], yet these are not guaranteed to be classically hard [111]. As a simple example, a Clifford circuit can evolve a local Pauli observable into a high-weight Z string, with a large expectation value on any initial computational basis state. This results in an apparent complexity according to our benchmark, since the string is high weight and therefore highly sensitive to noise. Nevertheless, the circuit is easily simulable with stabilizer methods. Analogous "spoofing" behavior occurs for nearly all established complexity metrics, including the entanglement entropy, magic, operator entanglement, out-of-time-ordered correlators (OTOCs), and others: It is always possible to construct circuits that appear complex by these metrics but remain classically simulable by other means [118]. With this in mind, we envision that our criteria are best used alongside existing complexity measures, much as current measures are used alongside one another.

E. Outlook

Our results provide the most extensive evidence yet that noisy quantum circuits without error correction can be efficiently classically simulated. In the long term, this emphasizes the necessity of quantum error correction for achieving a scalable quantum advantage over classical computation. For experiments without error correction, our results demonstrate the necessity of improving the noise rate proportional to the number of qubits. Moreover, even at sufficiently low noise rates, our results provide a rigorous test for whether a given circuit can yield a quantum advantage, by establishing, perhaps counterintuitively, that such circuits must be highly sensitive to noise.

Our work opens the door to a number of intriguing directions. First, is it possible to devise analogous classical algorithms for continuous-time dynamics instead of quantum circuits? Second, to what extent do our results carry over to more general errors, such as dephasing or thermal noise? Finally, can insights from our proof techniques be used to improve numerical algorithms for classically simulating quantum systems? In particular, we speculate that our approach may provide a fruitful starting point for simulating highly connected circuit architectures, where tensor network methods typically struggle.

ACKNOWLEDGMENTS

We are grateful to Dorit Aharonov, Zhenyu Cai, Matthias C. Caro, Andreas Elben, Bill Fefferman, Soumik Ghosh, Charvi Goyal, Greg Kahanamoku-Meyer, John Preskill, Dominik Wild, and Mike Zalatel for valuable discussions and insights. T. S. acknowledges support from the Walter Burke Institute for Theoretical Physics at Caltech. C. Y. is supported by the Department of Energy under Quantum Pathfinder Grant No. DE-SC0024324. X. G. acknowledges support from NSF PFC Grant No. PHYS 2317149 and start-up grants from CU Boulder. N. Y. Y acknowledges support from the NSF via the QLCI program (Grant No. OMA-2016245) and the STAQ II Program. The Institute for Quantum Information and Matter, with which T. S. is affiliated, is an NSF Physics Frontiers Center.

APPENDIX A: COMPARISON TO EXISTING RESULTS

In this appendix, we provide a more detailed comparison between the run-time of our classical algorithms and existing results. Let us first briefly summarize this comparison for the gate-based noise model, and then turn to the case of uniform noise.

(i) For quantum circuits with gate-based noise, the only existing classical algorithms are restricted to IQP circuits and random circuits (summarized in Table I of the main text). Thus, for general circuits, Theorem 2 provides an exponential improvement over the naive classical simulation time $2^{\mathcal{O}(n)}$.

This comparison becomes more involved for quantum circuits with uniform noise. As for the gate-based noise model, the only existing classical algorithms for uniform noise which explicitly take advantage of the noise are specific to IQP circuits and random circuits. However, since uniform noise limits the depth of any nontrivial quantum circuit [13], we should compare the scaling of our theorems to existing algorithms for low-depth quantum circuits. In particular, we show in Appendix C that expectation values in quantum circuits with uniform noise become ε-close to their values in the maximally mixed state at logarithmic depth, $d \geq \gamma^{-1}\log(1/\varepsilon)$ [77]. With this in mind, we compare the run-time of our classical algorithms to existing algorithms for circuits of such depth on a case-by-case basis below.

(1) Let us begin by considering general (i.e., all-to-all-connected) circuit architectures. In this setting, there are no existing classical algorithms for logarithmic depth circuits, and so Theorem 1 provides an exponential improvement over the naive classical simulation time $2^{\mathcal{O}(n)}$. This exponential improvement continues to hold even for more restricted classes of observables, as we discuss below.
   (a) For the specific case of local observables, we can moderately improve the naive classical simulation time by restricting to qubits within the light cone of the observable. For general circuits, the light cone contains 2^d qubits, leading to a run-time of $2^{\mathcal{O}(2^d)} = 2^{(1/\varepsilon)^{\mathcal{O}(\gamma^{-1})}}$. This is exponential in the inverse precision ε^{-1}, whereas our Algorithm 1 runs in polynomial time, $(1/\varepsilon)^{\tilde{\mathcal{O}}(\gamma^{-2})}$, for local observables.
   (b) For tensor-product observables in general circuits, Ref. [119] provides an algorithm with time $n^{\mathcal{O}(2^{2d}\log(n/\varepsilon))}$, i.e., $n^{(1/\varepsilon)^{\mathcal{O}(\gamma^{-1})}\log(n)}$, to estimate either (i) output probabilities, i.e., $O = |0^n\rangle\langle 0^n|$, or (ii) the magnitude of Pauli expectation values, $|\mathrm{tr}(\mathcal{C}\{\rho\}O)|$, where O is Pauli. The run-time is exponential in ε^{-1} when applied to noisy quantum circuits, due to the double exponential dependence of the original run-time on the depth d. In contrast, in case (i), Algorithm 1 computes output probabilities to an *exponentially smaller* error, $\varepsilon \cdot 2^{-n/2}$, in quasipolynomial time. In case (ii), Algorithm 1 computes Pauli expectation values (including their sign) to within the same error as Ref. [119] in polynomial time.

(2) We now turn to more restricted circuit architectures, beginning with the simplest case of 1D circuits. For tensor-product observables in 1D circuits, our algorithm provides no improvement because a trivial classical simulation via matrix product states is already very efficient.

(a) For tensor-product observables in 1D circuits, one can perform a classical simulation using matrix product states. This has run-time $2^{\mathcal{O}(d)}$, which translates to $(1/\varepsilon)^{\mathcal{O}(\gamma^{-1})}$ for quantum circuits with uniform noise. This is polynomial in ε^{-1} as is our algorithm.

(3) For 2D circuits, existing classical algorithms become less efficient, and our results generally provide a quasipolynomial to polynomial improvement.
(a) For local observables in 2D circuits, the best existing classical algorithm is a direct simulation restricted to the light cone of the observable. The light cone contains $\mathcal{O}(d^2)$ qubits, which leads to a run-time $2^{\mathcal{O}(d^2)} = (1/\varepsilon)^{\mathcal{O}(\gamma^{-2}\log(1/\varepsilon))}$ that is quasipolynomial in ε^{-1}. In contrast, our Algorithm 1 runs in polynomial time, $(1/\varepsilon)^{\tilde{\mathcal{O}}(\gamma^{-2})}$. We note that we can also utilize the same light-cone argument to improve the scaling of our Algorithm 2, by replacing n with $(\gamma^{-1}\log(1/\varepsilon))^2$ in Theorem 2. This leads to a run-time $(1/\varepsilon)^{\mathcal{O}(\gamma^{-1}\log\log(\varepsilon^{-1}))}$, which is nearly polynomial and has a better dependence on the noise rate γ.
(b) For tensor-product observables in 2D circuit architectures, Ref. [120] provides an algorithm to estimate expectation values in time $\mathcal{O}(nd^2 2^{d^2}/\varepsilon^2)$ for any tensor-product observable, $O = \otimes_{j=1}^{n} O_j$ with $||O_j||_\infty \le 1$. This translates to a run-time $n \cdot (1/\varepsilon)^{\mathcal{O}(\gamma^{-2}\log(1/\varepsilon))}$ for quantum circuits with uniform noise, which is quasipolynomial in ε^{-1}. For Pauli observables, our algorithm improves the run-time to polynomial in ε^{-1}. For general tensor-product observables, we note that our algorithm computes expectation values to within precision $\varepsilon \cdot ||O||_F$, where $||O||_F = \prod_{j=1}^{n} ||O_j||_F$ may be substantially smaller than 1. Thus, replacing $\varepsilon \to \varepsilon/||O||_F$ in Theorem 1 for the sake of comparison to Ref. [120], our algorithm has run-time $n^{\tilde{\mathcal{O}}(\gamma^{-1})+\mathcal{O}(\log(||O||_F/\varepsilon)/\gamma d)}(||O||_F/\varepsilon)^{\tilde{\mathcal{O}}(\gamma^{-2})}$ for general observables. This run-time may be better or worse than that of Ref. [120] depending on the scalings of n, d, ε, γ. For example, for large n but moderate ε and d, the algorithm in Ref. [120] is more efficient than our algorithm. On the other hand, for small ε or large d, our algorithm becomes more efficient. In particular, in the case $d = \Omega(\log(n))$, our algorithm runs in polynomial time whereas the algorithm in Ref. [120] runs in quasipolynomial time.

(4) Finally, for 3D circuit architectures, Theorem 1 provides a subexponential to polynomial improvement over existing algorithms for tensor-product observables, and a quasipolynomial to polynomial improvement for local observables.
(a) For tensor-product observables in 3D circuits, Ref. [120] provides an algorithm with run-time $\mathcal{O}(nd^3 2^{d^2 n^{1/3}}/\varepsilon^2)$, i.e., $n \cdot (1/\varepsilon)^{\mathcal{O}(\gamma^{-2}n^{1/3}\log(1/\varepsilon))}$ for quantum circuits with uniform noise. The run-time is subexponential in n, and is improved to either polynomial or quasipolynomial time by our algorithm.
(b) For local observables in 3D circuits, a direct light-cone simulation runs in quasipolynomial time, $2^{\mathcal{O}(d^3)} = (1/\varepsilon)^{\mathcal{O}(\gamma^{-3}\log(1/\varepsilon)^2)}$. In contrast, our Algorithm 1 runs in polynomial time, $(1/\varepsilon)^{\tilde{\mathcal{O}}(\gamma^{-2})}$, and our Algorithm 2 runs in nearly polynomial time, $(1/\varepsilon)^{\tilde{\mathcal{O}}(\gamma^{-1}\log\log(\varepsilon^{-1}))}$.

APPENDIX B: LOW-AVERAGE ENSEMBLES OF STATES

We now formally define our allowed ensembles of input states, which we call *low-average ensembles*.

Definition B1 (Low-average ensembles of states). We say that an ensemble of quantum states $\mathcal{E} = \{\rho\}$ is a low-average ensemble with purity c if $||(1/[|\mathcal{E}|]) \times \sum_\rho \rho||_\infty \le c/2^n$. For ensembles of pure states and $c = 1$, this coincides with the definition of a 1-design of states.

As an example, the ensemble of computational basis states with $n - m$ random bits and m fixed bits forms a low-average ensemble with purity $c = 2^m$. Any complete basis of pure states forms a low-average ensemble with $c = 1$.

We now state and prove the more general version of our Lemma 1 from the main text.

Lemma B1 (Extended version of Lemma 1 of the main text). Consider a low-average ensemble $\mathcal{E} = \{\rho\}$ with purity c and two observables O and $\tilde{O}$. The root-mean-square difference between the expectation value of O and $\tilde{O}$ is less than $\sqrt{c} \cdot ||O - \tilde{O}||_F$.

Proof of Lemma B1. For simplicity, we begin by assuming that each state is pure, and afterward generalize to mixed states. For a pure state $\rho = |\psi\rangle\langle\psi|$, the squared difference obeys

$$\mathrm{tr}(|\psi\rangle\langle\psi|\delta O)^2 = \langle\psi|\delta O^\dagger|\psi\rangle\langle\psi|\delta O|\psi_i\rangle \le \langle\psi|\delta O^\dagger \delta O|\psi\rangle, \tag{B1}$$

where we recognize the middle expression as the expectation value of $|\psi\rangle\langle\psi|$ in the state $\delta O|\psi\rangle$, which is upper bounded by the normalization of the state. Inserting this inequality into the mean-square error and applying the low-average condition gives our desired bound,

$$\begin{aligned}\frac{1}{|\mathcal{E}|}\sum_\psi \mathrm{tr}(|\psi\rangle\langle\psi|\delta O)^2 &\le \frac{1}{|\mathcal{E}|}\sum_\psi \langle\psi|\delta O^\dagger \delta O|\psi\rangle \\ &= \mathrm{tr}(\rho_{\mathcal{E}} \cdot \delta O^\dagger \delta O) \le ||\rho_{\mathcal{E}}||_\infty \cdot ||\delta O^\dagger \delta O||_1 = c \cdot ||\delta O||_F^2,\end{aligned} \tag{B2}$$

where we use $\rho_{\mathcal{E}} = (1/|\mathcal{E}|)\sum_{\psi} |\psi\rangle\langle\psi|$ to denote the mixture of the ensemble.

We now turn to mixed states. Let us write each mixed state in its orthonormal eigenbasis $\{|\psi_\rho\rangle\}$ as $\rho = \sum_{\psi_\rho} p_{\psi_\rho} |\psi_\rho\rangle\langle\psi_\rho|$. For each mixed state, we can define a quantum channel that dephases states in the eigenbasis of ρ, $\mathcal{D}_\rho\{|\psi_\rho\rangle\langle\psi_\rho|\psi'_\rho\} = \delta_{\psi_\rho,\psi'_\rho}|\psi_\rho\rangle\langle\psi_\rho|$. This allows us to write each mixed state ρ as its corresponding dephasing channel $\mathcal{D}_\rho$ applied to a pure state $|\phi_\rho\rangle$:

$$\rho = \mathcal{D}_\rho\{|\phi_\rho\rangle\langle\phi_\rho|\}, \qquad |\phi_\rho\rangle = \sum_{\psi_\rho} \sqrt{p_{\psi_\rho}}\,|\psi_\rho\rangle. \tag{B3}$$

Turning to the mean-square difference in expectation values, we first rewrite each trace as

$$\mathrm{tr}(\rho\cdot\delta O) = \mathrm{tr}(\mathcal{D}_\rho\{|\phi_\rho\rangle\langle\phi_\rho|\}\cdot\delta O) = \langle\phi_\rho|\mathcal{D}_\rho\{\delta O\}|\phi_\rho\rangle, \tag{B4}$$

since the dephasing channel is equal to its conjugate. Applying the same inequality as in Eq. (B1) gives

$$\mathrm{tr}(\rho\delta O)^2 \le \langle\phi_\rho|\mathcal{D}_\rho\{\delta O^\dagger\}\cdot\mathcal{D}_\rho\{\delta O\}|\phi_\rho\rangle. \tag{B5}$$

Since both operators within the expectation value are diagonal in the eigenbasis of ρ, we can compute the expectation value as

$$\begin{aligned}&\langle\phi_\rho|\mathcal{D}_\rho\{\delta O^\dagger\}\cdot\mathcal{D}_\rho\{\delta O\}|\phi_\rho\rangle\\ &= \sum_{\psi_\rho} \sqrt{p_{\psi_\rho}}\cdot(\delta O^\dagger)_{\psi_\rho\psi_\rho}\cdot(\delta O)_{\psi_\rho\psi_\rho}\cdot\sqrt{p_{\psi_\rho}}.\end{aligned} \tag{B6}$$

This is upper bounded by the analogous expression that includes the off-diagonal matrix elements of δO,

$$\begin{aligned}&\sum_{\psi_\rho} \sqrt{p_{\psi_\rho}}\cdot(\delta O^\dagger)_{\psi_\rho\psi_\rho}\cdot(\delta O)_{\psi_\rho\psi_\rho}\cdot\sqrt{p_{\psi_\rho}}\\ &\le \sum_{\psi_\rho,\psi'_\rho} p_{\psi_\rho}\cdot(\delta O^\dagger)_{\psi_\rho\psi'_\rho}\cdot(\delta O)_{\psi'_\rho\psi_\rho} = \mathrm{tr}(\rho\cdot\delta O^\dagger\delta O),\end{aligned} \tag{B7}$$

since each off-diagonal element contributes an amount $p_{\psi_\rho}(\delta O^\dagger)_{\psi_\rho\psi'_\rho}(\delta O)_{\psi'_\rho\psi_\rho} = p_{\psi_\rho}|(\delta O)_{\psi'_\rho\psi_\rho}|^2 \ge 0$. Putting it all together, we have

$$\begin{aligned}\frac{1}{|\mathcal{E}|}\sum_\rho \mathrm{tr}(\rho\delta O)^2 &\le \frac{1}{|\mathcal{E}|}\sum_\rho \mathrm{tr}(\rho\cdot\delta O^\dagger\delta O)\\ &= \mathrm{tr}(\rho_{\mathcal{E}}\cdot\delta O^\dagger\delta O) \le c\cdot||\delta O||_F^2,\end{aligned} \tag{B8}$$

where we denote $\rho_{\mathcal{E}} = (1/|\mathcal{E}|)\sum_\rho \rho$ once again. ∎

APPENDIX C: CLASSICAL ALGORITHM FOR UNIFORM NOISE (THEOREM 1)

We now provide our complete proof of Theorem 1 of the main text. We begin by establishing a short lemma that upper bounds the circuit depth of any nontrivial quantum circuit with uniform noise.

Lemma C1 (Upper bound on circuit depth for quantum circuits with uniform noise). Consider any observable O and any quantum circuit $\mathcal{C}$ with uniform noise. The Frobenius norm of the nonidentity component of $\mathcal{C}^\dagger\{O\}$ is upper bounded by $e^{-\gamma(d+1)}||O||_F$, where d is the circuit depth.

Proof. The Heisenberg time evolution of the nonidentity component of O has weight ≥ 1 at all layers. Thus, its Frobenius norm decreases by at least a factor $e^{-\gamma}$ from each noise channel. Since there are $d+1$ noise channels, we have $||\mathcal{C}^\dagger\{O\} - \mathrm{tr}(O)/2^n\cdot\mathbb{1}||_F \le e^{-\gamma(d+1)}||O||_F$. ∎

Lemma C1 can be viewed as an adaptation of Theorem 3 in Ref. [13] to the Frobenius norm of observables. Using our Lemma B1, it implies that one can trivially simulate expectation values (to within small root-mean-square error $\varepsilon\cdot\sqrt{c}\cdot||O||_F$ over a low-average ensemble of input states with purity c), in any quantum circuit with depth $d \ge \gamma^{-1}\log(1/\varepsilon)$ and uniform noise. To do so, one simply replaces the expectation value with its value in the maximally mixed state, $\mathrm{tr}(O)/2^n$. Thus, in what follows, we need only to analyze the performance of our Algorithm 1 on circuits of depth $d < \gamma^{-1}\log(1/\varepsilon)$.

Accuracy of algorithm. We establish the accuracy of Algorithm 1 by showing that our approximation $\tilde{O}$ is close to O in Frobenius norm:

$$||\tilde{O} - \mathcal{C}^\dagger\{O\}||_F \le \varepsilon||O||_F. \tag{C1}$$

From Lemma B1, this bounds the root-mean-square error in expectation values as desired. For simplicity, we assume that the operator O has no identity component; if it does, then this can be easily be included by adding a constant offset to the expectation values.

To show this, it is convenient to write operators as vectors in a doubled Hilbert space, with inner product $\langle\langle A||B\rangle\rangle = \mathrm{tr}(A^\dagger B)/2^n$ for operators A, B. The vector norm is equal to the squared Frobenius norm, $\langle\langle A||A\rangle\rangle = ||A||_F^2$. In this notation, each unitary layer U_t is represented by a superunitary $\mathcal{U}_t$ acting as $\mathcal{U}_t||A\rangle\rangle \equiv ||U_t^\dagger A U_t\rangle\rangle$. Meanwhile, the uniform noise channel acts on the basis of Pauli operators as $\mathcal{D}||P\rangle\rangle = e^{-\gamma w[P]}||P\rangle\rangle$. Finally, the projector $\mathcal{P}_\ell$ onto Pauli operators of weight w is given by $\mathcal{P}_w||P\rangle\rangle = \delta_{w[P],w}||P\rangle\rangle$. With this notation, our approximation $\tilde{O}$ in Eq. (6) of the main text can be written concisely as

$$||\tilde{O}\rangle\rangle = \sum_{w_0+\cdots+w_d\le\ell} e^{-\gamma(w_0+\cdots+w_d)}\mathcal{P}_{w_d}\mathcal{U}_d\cdots\mathcal{P}_{w_2}\mathcal{U}_2\mathcal{P}_{w_1}\mathcal{U}_1\mathcal{P}_{w_0}||O\rangle\rangle, \tag{C2}$$

where $w_t \ge 1$ is the weight at each layer. Each sequence of projectors in the sum corresponds to a single root-to-leaf sequence in the Pauli tree [Fig. 2(a) of the main text].

We would like to bound the Frobenius norm of $\tilde{O} - \mathcal{C}^\dagger\{O\} \equiv ||\tilde{O}\rangle\rangle - \mathcal{C}^\dagger||O\rangle\rangle$. As discussed in the main text, we do so by breaking up the sum over $w_0, \ldots, w_d$ layer by layer. We then use the fact that the $\mathcal{P}_{w_t}$ at each layer t project onto orthogonal subspaces for different w_t to upper bound the Frobenius norm. Let us illustrate this for the first layer. We can write

$$\begin{aligned}
||\tilde{O}\rangle\rangle - \mathcal{C}^\dagger||O\rangle\rangle &= \sum_{w_0+\cdots+w_d>\ell} e^{-\gamma(w_0+\cdots+w_d)}\mathcal{P}_{w_d}\mathcal{U}_d\cdots\mathcal{P}_{w_1}\mathcal{U}_1\mathcal{P}_{w_0}||O\rangle\rangle \\
&= \sum_{w_d=1}^{\infty} e^{-\gamma w_d}\mathcal{P}_{w_d}\mathcal{U}_d\left(\sum_{w_0+\cdots+w_{d-1}>\ell-w_d} e^{-\gamma(w_0+\cdots+w_{d-1})}\mathcal{P}_{w_{d-1}}\mathcal{U}_{d-1}\cdots\mathcal{P}_{w_1}\mathcal{U}_1\mathcal{P}_{w_0}||O\rangle\rangle\right) \\
&= \sum_{w_d=1}^{\ell-d+1} e^{-\gamma w_d}\mathcal{P}_{w_d}^{(\ell-d+1)}\mathcal{U}_d\left(\sum_{w_0+\cdots+w_{d-1}>\ell-w_d} e^{-\gamma(w_0+\cdots+w_{d-1})}\mathcal{P}_{w_{d-1}}\mathcal{U}_{d-1}\cdots\mathcal{P}_{w_1}\mathcal{U}_1\mathcal{P}_{w_0}||O\rangle\rangle\right).
\end{aligned} \tag{C3}$$

In the second line, we pull out the sum over w_d. In the third line, we group all terms with weight $w_d \ge \ell - d + 1$ at layer d together, by defining the modified "projectors":

$$\mathcal{P}_w^{(\ell')} \equiv \begin{cases} \mathcal{P}_w & w < \ell' \\ \mathcal{P}_{\ell'} + e^{-\gamma}\mathcal{P}_{\ell'+1} + \cdots + e^{-\gamma(n-\ell')}\mathcal{P}_n & w = \ell'. \end{cases} \tag{C4}$$

This grouping will reduce the number of terms in our error bound. The upper limit $\ell - d + 1$ is chosen to be just large enough such that the resulting sum of weights is always greater than ℓ, since there are d additional weights $w_{d-1}, \ldots, w_0$ and each weight is at least 1. Still working with only the first layer, we can now use the fact that $\mathcal{P}_w^{(\ell')}$ are orthogonal to remove them from the Frobenius norm:

$$\begin{aligned}
\left|\left|\,||\tilde{O}\rangle\rangle - \mathcal{C}^\dagger||O\rangle\rangle\,\right|\right|_F^2 &= \left|\left|\sum_{w_d=1}^{\ell-d+1} e^{-\gamma w_d}\mathcal{P}_{w_d}^{(\ell-d+1)}\mathcal{U}_d\left(\sum_{w_0+\cdots+w_{d-1}>\ell-w_d} e^{-\gamma(w_0+\cdots+w_{d-1})}\mathcal{P}_{w_{d-1}}\mathcal{U}_{d-1}\cdots\mathcal{P}_{w_1}\mathcal{U}_1\mathcal{P}_{w_0}||O\rangle\rangle\right)\right|\right|_F^2 \\
&= \sum_{w_d=1}^{\ell-d+1} e^{-2\gamma w_d}\left|\left|\mathcal{P}_{w_d}^{(\ell-d+1)}\mathcal{U}_d\left(\sum_{w_0+\cdots+w_{d-1}>\ell-w_d} e^{-\gamma(w_0+\cdots+w_{d-1})}\mathcal{P}_{w_{d-1}}\mathcal{U}_{d-1}\cdots\mathcal{P}_{w_1}\mathcal{U}_1\mathcal{P}_{w_0}||O\rangle\rangle\right)\right|\right|_F^2 \\
&\le \sum_{w_d=1}^{\ell-d+1} e^{-2\gamma w_d}\left|\left|\sum_{w_0+\cdots+w_{d-1}>\ell-w_d} e^{-\gamma(w_0+\cdots+w_{d-1})}\mathcal{P}_{w_{d-1}}\mathcal{U}_{d-1}\cdots\mathcal{P}_{w_1}\mathcal{U}_1\mathcal{P}_{w_0}||O\rangle\rangle\right|\right|_F^2.
\end{aligned} \tag{C5}$$

In the first line, we use that the Frobenius norm of a sum of orthogonal operators is equal to the sum of the operators' norms. In the second line, we use that $\mathcal{P}_w^{(\ell')}$ and unitary operation cannot increase the norm. This completes our analysis of the first layer.

To complete our proof, we repeat the steps above for each remaining layer in the quantum circuit. After doing so once more, for the second layer, we obtain

$$\left|\left|\,||\tilde{O}\rangle\rangle - \mathcal{C}^\dagger||O\rangle\rangle\,\right|\right|_F^2 \le \sum_{w_d=1}^{\ell-d+1}\sum_{w_{d-1}=1}^{\ell-d+2-w_d} e^{-2\gamma(w_d+w_{d-1})}\left|\left|\sum_{\substack{w_0+\cdots+w_{d-2}\\ >\ell-w_d-w_{d+1}}} e^{-\gamma(w_0+\cdots+w_{d-2})}\mathcal{P}_{w_{d-2}}\mathcal{U}_{d-2}\cdots\mathcal{P}_{w_1}\mathcal{U}_1\mathcal{P}_{w_0}||O\rangle\rangle\right|\right|_F^2.$$

After repeating for all layers, we obtain

$$\left|\left|\,|\tilde{O}\rangle\rangle - \mathcal{C}^\dagger|O\rangle\rangle\right|\right|_F^2 \le \sum_{w_d=1}^{\ell-d+1}\sum_{w_{d-1}=1}^{\ell-d+2-w_d}\cdots\sum_{w_1=1}^{\ell-w_d-\cdots-w_2} e^{-2\gamma(w_1+\cdots+w_d)}\left|\left|\sum_{w_0>\ell-w_d-\cdots-w_1} e^{-\gamma w_0}\mathcal{P}_{w_0}|O\rangle\rangle\right|\right|_F^2$$
$$= \sum_{w_d=1}^{\ell-d+1}\sum_{w_{d-1}=1}^{\ell-d+2-w_d}\cdots\sum_{w_1=1}^{\ell-w_d-\cdots-w_2}\sum_{w_0=\ell+1-w_d-\cdots-w_1}^{\ell+1-w_d-\cdots-w_1} e^{-2\gamma(w_0+\cdots+w_d)}\left|\left|\mathcal{P}_{w_0}^{(w_0)}|O\rangle\rangle\right|\right|_F^2$$
$$\le e^{-2\gamma(\ell+1)}||O||_F^2\cdot\left(\sum_{w_0+\cdots+w_d=\ell+1}1\right). \tag{C6}$$

In the second line, the final sum runs over a single value of w_0. It remains only to compute the final sum in the third line.

The sum in Eq. (6) counts the number of sequences $(w_0,\ldots,w_d)$ that sum to $\ell+1$. Each sequence corresponds to an individual truncation in our algorithm, as discussed and depicted in Fig. 2(a) of the main text. In mathematics, the number of such sequences is well known (referred to as "the number of compositions of $\ell+1$ into $d+1$ parts"), and is equal to $\binom{\ell}{d}$. Thus, we have

$$||\tilde{O}-\mathcal{C}^\dagger\{O\}||_F^2 \le e^{-2\gamma(\ell+1)}\cdot\binom{\ell}{d}\cdot||O||_F^2. \tag{C7}$$

To determine the required value of ℓ, we use a standard bound on the binomial coefficient, $\binom{\ell}{d}\le(e\ell/d)^d$. We then have

$$e^{-2\gamma(\ell+1)}\cdot\binom{\ell}{d}\le e^{-2\gamma(\ell+1)}\cdot\left(\frac{e\ell}{d}\right)^d\le e^{-\gamma(\ell+1)}, \tag{C8}$$

where the second inequality holds as long as $d\log(e\ell/d)\le\gamma(\ell+1)$. Thus, we can ensure that the desired error bound Eq. (C1) holds by choosing

$$\ell = c_\gamma d + 2\gamma^{-1}\log(1/\varepsilon), \tag{C9}$$

where c_γ is the solution of $c_\gamma=\gamma^{-1}\log(ec_\gamma)$; at small γ, we have $c_\gamma=\gamma^{-1}\log(e/\gamma)+\mathcal{O}(\gamma^{-1}\log\log(1/\gamma))$.

Run-time of algorithm. The run-time of Algorithm 1 is determined by the number of Pauli paths in our approximation of $\tilde{O}$. From Lemma 8 in Ref. [31], the number of such paths is upper bounded by

$$\min\left(n^{\ell/d},m\right)\cdot 2^{\mathcal{O}(\ell)}$$
$$=\min\left(n^{\mathcal{O}(c_\gamma+(2/\gamma d)\log(1/\varepsilon))},m\right)\cdot 2^{\mathcal{O}(c_\gamma d)}\cdot(1/\varepsilon)^{\mathcal{O}(1/\gamma)} \tag{C10}$$

if the observable O is a sum of m Pauli operators. To derive the second term in the minimum, we observe that in the second part of the proof of Lemma 8 of Ref. [31] we can choose the layer t to be the final circuit layer, which by definition contains at most m Pauli operators. Otherwise we can choose the circuit layer with the minimum weight, which gives the first term in the minimum, $n^{\ell/d}$, as in Ref. [31]. Inserting the value of c_γ and invoking Lemma C1 to restrict attention to $d<\gamma^{-1}\log(1/\varepsilon)$, we find a run-time

$$\min\left(n^{\mathcal{O}(\log(1/\gamma)/\gamma)+\mathcal{O}(\log(1/\varepsilon)/\gamma d)},m\right)\cdot(1/\varepsilon)^{\mathcal{O}(\log(1/\gamma)/\gamma^2)}. \tag{C11}$$

Taking the first term in the minimum leads to the first scaling in Theorem 1 of the main text. Taking the second term with $m=\mathrm{poly}(n)$ leads to the second scaling. ∎

APPENDIX D: CLASSICAL ALGORITHM FOR GATE-BASED NOISE (THEOREM 2)

We now turn to our Algorithm 2 for gate-based noise. We begin in Appendix D 1 with a proof of Lemma 2 of the main text, and analyze the remaining aspects of the algorithm in Appendix D 2.

1. Bounding noisy operator growth

We now prove Lemma 2 of the main text, which upper bounds the cumulative operator weight distribution,

$$Q^{(t)}(w)=\sum_{w'=w+1}^{n}P^{(t)}(w'), \tag{D1}$$

in any noisy quantum circuit. We also include a slightly stronger bound on the cumulative weight distribution when considering the approximate time-evolved operator $\tilde{O}$.

Lemma D1 (Extended version of Lemma 2 of the main text). Consider the Heisenberg time evolution $O^{(t)}$ of an observable O in a quantum circuit with gate-based noise. The cumulative weight distribution is less than

$$Q^{(t)}(w)\le e^{-2\gamma(w+1)}\cdot||O||_F^2 \tag{D2}$$

for all γ, w, t. We can also extend this to our classical approximation $\tilde{O}$ of O, which truncates Pauli operators of

weight greater than ℓ between circuit layers. In this case, the cumulative operator weight distribution $\tilde{Q}^{(t)}(\ell)$ obeys

$$\sum_{t=0}^{d} \tilde{Q}^{(t)}(\ell) \le e^{-2\gamma(\ell+1)} \cdot ||O||_F^2. \tag{D3}$$

Proof of Lemma D1. We begin by proving the first statement; the second will follow shortly after. We keep track of the Heisenberg time evolution of O gate by gate. We denote the operator evolved through the first g gates and noise channels of the circuit as $O^{(g)}$; if g_t denotes the final gate within a circuit layer t, then $O^{(g_t)}$ corresponds to $O^{(t)}$ in our layer-by-layer notation. We again use $O^{(0)}$ to denote the operator evolved through the readout noise channel.

The initial observable O will have some operator weight distribution $P(w)$ with norm $||O||_F^2$. After the application of readout noise, the weight distribution is damped to $P^{(0)}(w) = e^{-2\gamma w} P(w)$. This leads to a trivial upper bound on the cumulative operator weight distribution at time zero:

$$\begin{aligned} Q^{(0)}(w) &= \sum_{w'=w+1}^{N} e^{-2\gamma w'} P(w') \le e^{-2\gamma(w+1)} \cdot Q(w) \\ &\le e^{-2\gamma(w+1)} \cdot ||O||_F^2. \end{aligned} \tag{D4}$$

Our aim is to show that this bound in fact holds for all times.

We proceed gate by gate. The evolution during each gate consists of two steps: First, we apply the unitary gate U_g, then we apply local depolarizing noise to the two qubits in the gate. In the first step, application of U_g can transfer operator norm between adjacent values of ℓ. We capture this by defining the *current* $J^{(g)}(w)$ via

$$Q'^{(g)}(w-1) = Q^{(g-1)}(w-1) + J^{(g)}(w), \tag{D5}$$

where $Q'^{(g)}(w-1)$ denotes the operator norm at weight above $w-1$ after applying U_g but before applying the noise channel $\mathcal{D}_g$. Similarly, we have

$$P'^{(g)}(w) = P^{(g-1)}(w) + J^{(g)}(w) - J^{(g)}(w+1), \tag{D6}$$

where $P'^{(g)}(w)$ is also defined after U_g but before $\mathcal{D}_g$. The current $J^{(g)}(w)$ quantifies the net transfer in operator norm from weight $w-1$ to w by U_g. Writing it out explicitly, we have

$$\begin{aligned} J^{(g)}(w) = &\langle\langle O_{w-1,w} + O_{w,w} || O_{w-1,w} + O_{w,w} \rangle\rangle \\ &- \langle\langle O_{w,w} || O_{w,w} \rangle\rangle - \langle\langle O_{w,w-1} || O_{w,w-1} \rangle\rangle, \end{aligned} \tag{D7}$$

where we use the doubled bra-ket notation to denote inner products of operators, $\langle\langle A||B\rangle\rangle \equiv (1/2^n)\mathrm{tr}(AB)$, and abbreviate $O_{w,w'} \equiv \mathcal{P}_{w'}\{U_g^\dagger \mathcal{P}_w\{O^{(g-1)}\}U_g\}$ for the component of the operator that starts at weight w and ends at weight w'. The first term is the contribution from weights $w-1$ and w before the gate to the norm at weight w after the gate. The second and third term arise when taking the difference with $Q^{(g-1)}(w-1)$. Although it does not immediately appear so, the flow is in fact antisymmetric upon exchanging $w-1$ and w, as it must be in order to preserve the total operator norm. This can be verified by expanding the first term and using $\langle\langle O_{w,w} || O_{w-1,w} \rangle\rangle = -\langle\langle O_{w-1,w-1} || O_{w,w-1} \rangle\rangle$. This equality is guaranteed because $U_g^\dagger \mathcal{P}_w\{O^{(g-1)}\}U_g$ and $U_g^\dagger \mathcal{P}_{w-1}\{O^{(g-1)}\}U_g$ are orthogonal.

To incorporate the noise channel $\mathcal{D}_g$, we use the fact that any operator that increases in weight under U_g (in particular, the operator $O_{w-1,w}$) must have support on both qubits in U_g after the gate is applied. Thus, the norm of such an operator is damped by a factor of $e^{-2\gamma}$ by the noise channel. In full detail, we have

$$\begin{aligned} P^{(g)}(w) &= \langle\langle O_{w-1,w} + O_{w,w} + O_{w+1,w} || \mathcal{D}_g \cdot \mathcal{D}_g || O_{w-1,w} + O_{w,w} + O_{w+1,w} \rangle\rangle \\ &= \langle\langle O_{w-1,w} + O^1_{w,w} || \mathcal{D}_g \cdot \mathcal{D}_g || O_{w-1,w} + O^1_{w,w} \rangle\rangle + \langle\langle O_{w+1,w} + O^2_{w,w} || \mathcal{D}_g \cdot \mathcal{D}_g || O_{w+1,w} + O^2_{w,w} \rangle\rangle \\ &\quad + \langle\langle O^3_{w,w} || \mathcal{D}_g \cdot \mathcal{D}_g || O^3_{w,w} \rangle\rangle \le e^{-4\gamma} \langle\langle O_{w-1,w} + O^1_{w,w} || O_{w-1,w} + O^1_{w,w} \rangle\rangle \\ &\quad + e^{-2\gamma} \langle\langle O_{w+1,w} + O^2_{w,w} || O_{w+1,w} + O^2_{w,w} \rangle\rangle + \langle\langle O^3_{w,w} || O^3_{w,w} \rangle\rangle \\ &= P^{(g-1)}(w) + e^{-4\gamma} \langle\langle O_{w-1,w} + O^1_{w,w} || O_{w-1,w} + O^1_{w,w} \rangle\rangle - \langle\langle O^1_{w,w} || O^1_{w,w} \rangle\rangle - \langle\langle O_{w,w-1} || O_{w,w-1} \rangle\rangle \\ &\quad + e^{-2\gamma} \langle\langle O_{w+1,w} + O^2_{w,w} || O_{w+1,w} + O^2_{w,w} \rangle\rangle - \langle\langle O^2_{w,w} || O^2_{w,w} \rangle\rangle - \langle\langle O_{w,w+1} || O_{w,w+1} \rangle\rangle \\ &\le P^{(g-1)}(w) + \{1, e^{-4\gamma}\} \cdot (\langle\langle O_{w-1,w} + O^1_{w,w} || O_{w-1,w} + O^1_{w,w} \rangle\rangle - \langle\langle O^1_{w,w} || O^1_{w,w} \rangle\rangle - \langle\langle O_{w,w-1} || O_{w,w-1} \rangle\rangle) \\ &\quad + \{1, e^{-2\gamma}\} \cdot (\langle\langle O_{w+1,w} + O^2_{w,w} || O_{w+1,w} + O^2_{w,w} \rangle\rangle - \langle\langle O^2_{w,w} || O^2_{w,w} \rangle\rangle - \langle\langle O_{w,w+1} || O_{w,w+1} \rangle\rangle) \\ &= P^{(g-1)}(w) + \{1, e^{-4\gamma}\} \cdot (\langle\langle O_{w-1,w} + O_{w,w} || O_{w-1,w} + O_{w,w} \rangle\rangle - \langle\langle O_{w,w} || O_{w,w} \rangle\rangle - \langle\langle O_{w,w-1} || O_{w,w-1} \rangle\rangle) \\ &\quad + \{1, e^{-2\gamma}\} \cdot (\langle\langle O_{w+1,w} + O_{w,w} || O_{w+1,w} + O_{w,w} \rangle\rangle - \langle\langle O_{w,w} || O_{w,w} \rangle\rangle - \langle\langle O_{w,w+1} || O_{w,w+1} \rangle\rangle). \end{aligned} \tag{D8}$$

In the second line, we perform the orthogonal decomposition $O_{w,w} = O^1_{w,w} + O^2_{w,w} + O^3_{w,w}$, where $O^1_{w,w}$ is proportional to $O_{w-1,w}$, and $O^2_{w,w}$ is proportional to $O_{w+1,w}$, and $O^3_{w,w}$ is orthogonal to both. This is an orthogonal decomposition because $O_{w-1,w}$ and $O_{w+1,w}$ are orthogonal, which follows since $U_g^\dagger \mathcal{P}_{w-1}\{O^{(g-1)}\}U_g$ and $U_g^\dagger \mathcal{P}_{w+1}\{O^{(g-1)}\}U_g$ are orthogonal. The three operators remain orthogonal after application of $\mathcal{D}_g$, since $O_{w-1,w}$ and $O_{w+1,w}$ are eigenstates of the noise channel $\mathcal{D}_g$ (with eigenvalues $e^{-2\gamma}$ and $e^{-\gamma}$, respectively). In the second to last line, we use that

$$\begin{aligned} P^{(g-1)}(w) &= \langle\langle O_{w,w-1} + O_{w,w} + O_{w,w+1} || O_{w,w-1} + O_{w,w} + O_{w,w+1}\rangle\rangle \\ &= \langle\langle O_{w,w-1} || O_{w,w-1}\rangle\rangle + \langle\langle O_{w,w+1} || O_{w,w+1}\rangle\rangle + \langle\langle O^1_{w,w} || O^1_{w,w}\rangle\rangle + \langle\langle O^2_{w,w} || O^2_{w,w}\rangle\rangle + \langle\langle O^3_{w,w} || O^3_{w,w}\rangle\rangle, \end{aligned} \tag{D9}$$

which follows since the five operators in the second line are orthogonal to one another. Finally, in the last line, we use the notation $\{1, e^{-4\gamma}\}$ and $\{1, e^{-2\gamma}\}$ to denote that the bound holds for either choice of constant, where one can make the choice independently between the two terms. Recognizing the first three terms in parentheses as $J^{(g)}(w)$ and the latter three terms as $-J^{(g)}(w+1)$, we have the simpler bound,

$$\begin{aligned} P^{(g)}(w) \le P^{(g-1)}(w) &+ \{1, e^{-4\gamma}\} \cdot J^{(g)}(w) \\ &- \{1, e^{-2\gamma}\} \cdot J^{(g)}(w+1), \end{aligned} \tag{D10}$$

where again the bound holds for any choice of the constants. When $J^{(g)}(w)$ is positive, the stronger bound is obtained by taking the second constant in the first term, and when it is negative, the first constant. The reverse holds for the $J^{(g)}(w+1)$ term. To summarize, increases in $P^{(g)}(w)$ are discounted by a factor $e^{-4\gamma}$ or $e^{-2\gamma}$ due to gate-based noise, while decreases are not.

We would like to utilize our update rule, Eq. (D10), to bound the cumulative norm above a threshold weight w. To begin, we can iterate Eq. (D10) over each gate $h = 1, \ldots, g$ in the circuit to obtain

$$\begin{aligned} P^{(g)}(k) \le P^{(0)}(k) &+ \left(\sum_{h=1}^{g} \{1, e^{-4\gamma}\} \cdot J^{(h)}(k)\right) \\ &- \sum_{h=1}^{g} J^{(h)}(k+1), \end{aligned} \tag{D11}$$

where we choose the constant 1 for each term in the second sum, and for convenience change the variable from w to k. To bound the cumulative weight distribution above weight w, we can sum over $k = w+1, \ldots, N$. Taking the constant $e^{-4\gamma}$ for $k = w+1$, and the constant 1 for all other k, we find

$$Q^{(g)}(w) \le Q^{(0)}(w) + e^{-4\gamma}\left(\sum_{h=1}^{g} J^{(h)}(w+1)\right). \tag{D12}$$

We would like to bound the sum within the parentheses above. Intuitively, the sum cannot be too large, since the net current from w to $w+1$ is sourced from the operator's support on weight w. To make this precise, let us observe Eq. (D11) with $k = w$. The rightmost term is precisely the sum we would like to bound. Inverting the inequality, taking the constant $e^{-4\gamma}$ for each term in the first sum, and using that $P^{(g)}(w-1) \ge 0$, we have

$$\left(\sum_{h=1}^{g} J^{(h)}(w+1)\right) \le P^{(0)}(w) + e^{-4\gamma}\left(\sum_{h=1}^{g} J^{(h)}(w)\right). \tag{D13}$$

This bounds the net current from w to $w+1$ by the initial norm on w, plus the net current from $w-1$ to w discounted by $e^{-4\gamma}$. We can iterate over the weight to obtain

$$\left(\sum_{h=1}^{g} J^{(h)}(w+1)\right) \le \sum_{k=1}^{w} e^{-4\gamma(k-1)} \cdot P^{(0)}(w+1-k). \tag{D14}$$

Inserting into Eq. (D12), we find

$$Q^{(g)}(w) \le Q^{(0)}(w) + \sum_{k=1}^{w} e^{-4\gamma k} \cdot P^{(0)}(w+1-k). \tag{D15}$$

The kth term in the sum corresponds to operators that begin at weight $w+1-k$ and evolve to weight greater than w. We find that such operators' contribution to $Q^{(g)}(w)$ is discounted by at least $e^{-4\gamma k}$, since they have participated in at least k gates. Expressing $P^{(0)}(w+1-k)$ and $Q^{(0)}(w)$ in terms of their values before readout noise, we have

$$\begin{aligned} Q^{(g)}(w) &\le e^{-2\gamma w} \cdot Q(w) + \sum_{k=1}^{w} e^{-4\gamma k} \cdot e^{-2\gamma(w+1-k)} \cdot P(w+1-k) \\ &\le e^{-2\gamma(w+1)} \cdot ||O||_F^2, \end{aligned} \tag{D16}$$

as desired. The right-hand side can be improved to $e^{-4\gamma(w+1/2)} \cdot ||O||_F^2$ if the initial observable has weight one.

We now turn to the second inequality in the lemma. Our classical algorithm truncates high-weight components of

$\tilde{O}^{(t)}$ after each layer of the circuit. As aforementioned, we denote the final gate in layer t as g_t, so that the truncation occurs in between gates g_t and g_t+1. To begin, note that the truncation in our classical algorithm cannot increase the operator weight distribution. This means that the upper bounds Eqs. (D4) and (D10) continue to apply to noisy quantum circuit with truncation. Moreover, for the specific weight $w=\ell$, our upper bound on the cumulative weight distribution can be strengthened to

$$\tilde{Q}^{(t)}(\ell) \le e^{-4\gamma}\left(\sum_{h=g_{t-1}+1}^{g_t} \tilde{J}^{(h)}(\ell+1)\right), \tag{D17}$$

since the norm above ℓ is reset to zero by the truncation after layer $t-1$. Here, $\tilde{J}^{(g)}(\ell)$ is defined analogous to $J^{(g)}(\ell)$, but with $\tilde{O}$ instead of O. Summing over the layers t, we have

$$\sum_{t=0}^{d} \tilde{Q}^{(t)}(\ell) \le \tilde{Q}^{(0)}(\ell) + e^{-4\gamma}\left(\sum_{h=1}^{g_d} \tilde{J}^{(h)}(\ell+1)\right), \tag{D18}$$

where the right-hand side is now the same quantity we bounded for the first inequality. Repeating the steps of that proof, we find

$$\sum_{t=0}^{d} \tilde{Q}^{(t)}(\ell) \le e^{-2\gamma(\ell+1)} \cdot ||O||_F^2, \tag{D19}$$

as desired. ∎

2. Analysis of algorithm

We now apply our bound on operator weight distributions to prove Theorem 2. As for Theorem 1, we break our proof into two parts, which address the accuracy and runtime of Algorithm 2, respectively.

Accuracy of algorithm. Our classical algorithm truncates operator components with weight greater than ℓ in between each layer of the circuit. Each truncation can increase the distance between the time-evolved operator $O^{(t)}$ and our approximation $\tilde{O}^{(t)}$ by at most

$$||O^{(t)} - \tilde{O}^{(t)}||_F \le ||O^{(t-1)} - \tilde{O}^{(t-1)}||_F + ||\mathcal{P}_{>\ell}\{\mathcal{D}_t\{U_t^\dagger \tilde{O}^{(t-1)} U_t\}\}||_F, \tag{D20}$$

by the triangle inequality. The latter term is precisely the square root of $\tilde{Q}^{(t)}(\ell)$, the cumulative weight distribution for our approximation $\tilde{O}^{(t)}$. Summing over time steps $t=0,\ldots,d$, we have

$$\begin{aligned} ||O^{(d)} - \tilde{O}^{(d)}||_F &\le \sum_{t=0}^{d} \sqrt{\tilde{Q}^{(t)}(\ell)} \\ &\le \sqrt{d+1}\left(\sum_{t=0}^{d} \tilde{Q}^{(t)}(\ell)\right)^{1/2} \\ &\le \sqrt{d+1}\, e^{-\gamma(\ell+1)} ||O||_F, \end{aligned} \tag{D21}$$

where in the second line we apply the Cauchy-Schwarz inequality, and in the third line we apply Lemma D1.

From Lemma B1, the above bound on the Frobenius norm of $O^{(d)} - \tilde{O}^{(d)}$ leads to an upper bound on the root-mean-square error for any low-average ensemble:

$$\begin{aligned} &\left(\frac{1}{|\mathcal{E}|}\sum\nolimits_{\rho} |\mathrm{tr}(\rho O^{(d)}) - \mathrm{tr}(\rho \tilde{O}^{(d)})|^2\right)^{1/2} \\ &\le \sqrt{c}\sqrt{d+1} \cdot e^{-\gamma(\ell+1)} \cdot ||O||_F. \end{aligned} \tag{D22}$$

For a desired error $\varepsilon\sqrt{c}||O||_F$, it suffices to take

$$\ell \ge \frac{1}{\gamma}\log\left(\frac{\sqrt{d+1}}{\varepsilon}\right) - 1. \tag{D23}$$

Run-time of algorithm. We now turn to the complexity of classically computing the expectation value $\mathrm{tr}(\rho_i \tilde{O}^{(d)})$. As described in the main text, we do so by simulating the time evolution of $\tilde{O}^{(t)} = \sum_{P:w[P]\le\ell} c_P^{(t)} P$ layer by layer in the Pauli basis. This requires keeping track of the real-valued coefficients $c_P^{(t)}$ for all Pauli operators with weight less than ℓ. There are

$$D_\ell = \sum_{k=0}^{\ell} \binom{n}{k} \cdot 3^k \le \binom{n}{\ell} \cdot 3^\ell \cdot \frac{n-\ell+1}{n-\frac{4}{3}\ell+1} \le \frac{3}{2}\frac{(3n)^\ell}{\ell!} \tag{D24}$$

such coefficients. The sum counts the number of Pauli operators with weight k by first counting the number of sets of k qubits, $\binom{n}{k}$, on which the Pauli has support, and then multiplying by the number of Pauli operators on those k qubits, 3^k. The second inequality upper bounds the sum when $\ell \le 3n/4$, and the final inequality upper bounds the fraction by $3/2$, which applies for $\ell \le n/2$.

At the beginning of our algorithm, we initialize the coefficients $c_P^{(0)}$ by computing the Pauli coefficients of O. As stated in the theorem, we assume that each coefficient can be computed efficiently, i.e., in $\mathcal{O}(1)$ time [121], and thus the entire vector in $\mathcal{O}(D_\ell)$ time. At the end of our algorithm, we compute the trace $\mathrm{tr}(\rho_i \tilde{O}^{(d)})$ by computing the Pauli coefficients $\mathrm{tr}(\rho P)$ of ρ and taking their inner product with $c_P^{(d)}$. Again, if each coefficient can be computed in $\mathcal{O}(1)$ time, then the expectation value can be computed in $\mathcal{O}(D_\ell)$ time.

We update the coefficients from layer to layer via Eq. (8) of the main text, where the transition amplitudes $a_{QP}^{(t)}$ are computed from the depth-1 unitary $U_t = \prod_g U_g$ as

$$a_{PQ}^{(t)} \equiv \frac{1}{2^n}\mathrm{tr}\left(QU_tPU_t^\dagger\right) = \prod_g \frac{1}{4}\mathrm{tr}_{\mathrm{supp}(g)}\left(Q_{\mathrm{supp}(g)} \cdot \mathcal{D}_g\{U_g^\dagger P_{\mathrm{supp(g)}} U_g\}\right). \quad \text{(D25)}$$

The product runs over all gates g in U_t, where $\mathrm{supp}(g)$ denotes the pair of qubits in gate g. We see that the n-qubit transition amplitudes factorize into a product of two-qubit transition amplitudes. Each two-qubit amplitude can be computed in $\mathcal{O}(1)$ time, and thus the n-qubit amplitude in $\mathcal{O}(n)$ time.

A naive algorithm to update the coefficients would compute each transition amplitude $a_{QP}^{(t)}$ individually and thus require $\mathcal{O}(n \cdot D_\ell^2)$ time. We can moderately improve this scaling by leveraging the sparsity of the transition amplitudes. Note that the support of $U_tPU_t^\dagger$ consists of at most 2ℓ qubits: the qubits in $\mathrm{supp}(P)$ and those that they couple to under U_t. Thus, for each P, we only need to compute transitions to Q whose support lies within this set. There are at most $(3/2)3^\ell 2\binom{\ell}{\ell} \leq (3/2)12^\ell$ such Q, by a similar computation to Eq. (D24). Moreover, each transition amplitude can be computed from a product of only ℓ two-qubit transition amplitudes, in contrast to the naive number $n/2$. Together, these lead to a run-time of $\mathcal{O}(D_\ell \cdot 12^\ell \cdot \ell)$ per circuit layer.

Inserting our bound on D_ℓ and summing over the d circuit layers, our algorithm requires

$$\text{time} = \mathcal{O}\left(d \cdot \ell \cdot \frac{(36n)^\ell}{\ell!}\right) = \mathcal{O}(d \cdot n^\ell), \quad \text{(D26)}$$

where d is the depth of the circuit. On the right, we simplify the expression by using $\ell \cdot 36^\ell/\ell! = \mathcal{O}(1)$. (For moderate values of $\ell \lesssim 36$, this may incur a large constant prefactor in practice, although it will still be much smaller than the additional factor of n^ℓ in the naive update algorithm.) The algorithm uses space

$$\text{space} = \mathcal{O}\left(\frac{(3n)^\ell}{\ell!}\right) = \mathcal{O}(n^\ell), \quad \text{(D27)}$$

proportional to the number of coefficients $c_P^{(t)}$. These requirements are comparable to performing exact time evolution in a system of $n_{\mathrm{eff}} \sim \ell \log(n)$ qubits.

For a desired error ε, we take ℓ according to Eq. (D23). Plugging into the above, this gives

$$\text{time} = \mathcal{O}\left(d \cdot n^{(1/\gamma)\log(\sqrt{d+1}/\varepsilon)-1}\right) \quad \text{(D28)}$$

and

$$\text{space} = \mathcal{O}\left(n^{(1/\gamma)\log(\sqrt{d+1}/\varepsilon)-1}\right). \quad \text{(D29)}$$

This concludes our proof. ∎

APPENDIX E: CLASSICAL ALGORITHMS FOR SAMPLING FROM NOISY QUANTUM CIRCUITS

In select cases, our classical algorithms can also be leveraged to sample from noisy quantum circuits. In sampling, we are interested in reproducing the statistics when the final state of the circuit is measured in the computational basis. From Born's rule, each measurement produces an n-bit string z with probability $p_{\rho,\mathcal{C}}(z) = \langle z|\mathcal{C}\{\rho\}|z\rangle$. In general, sampling is a strictly harder task than computing expectation values, since the distribution $p_{\rho,\mathcal{C}}(z)$ contains information about exponentially many expectation values in the computational basis.

We provide three algorithms for sampling from noisy quantum circuits. The first two are direct extensions of Theorems 1 and 2, and apply to any circuit on most input states. Similar to previous works [26,31], we require one additional assumption: that the output distribution of the circuit *anticoncentrates*, in the sense that $(1/|\mathcal{E}|)\sum_\rho \sum_z p_{\rho,\tilde{\mathcal{C}}}(z)^2 = \mathcal{O}(2^{-n})$ for the ensemble $\mathcal{E} = \{\rho\}$. Here, $\tilde{\mathcal{C}}$ denotes the circuit $\mathcal{C}$ with the readout noise channel omitted. Anticoncentration guarantees that most expectation values in the computational basis are close to zero. In the presence of readout noise, this enables one to sample from $p_{\rho,\mathcal{C}}(z)$ by computing only a small number of low-weight expectation values [26], which we show can be done efficiently using our Algorithms 1 and 2 (see Appendix E for details).

Theorem E1 (Sampling from noisy quantum circuits; informal). Consider a noisy quantum circuit $\mathcal{C}$ and a low-average ensemble $\mathcal{E} = \{\rho\}$ whose Pauli coefficients can be efficiently computed. Assume the output distribution anticoncentrates, $(1/|\mathcal{E}|)\sum_\rho \sum_z p_{\rho,\tilde{\mathcal{C}}}(z)^2 = \mathcal{O}(2^{-n})$. Then for either uniform or gate-based noise, there is a classical algorithm to sample from $p_{\rho,\mathcal{C}}(z)$ within root-mean-square total variational distance ε in quasipolynomial time.

Our third sampling algorithm specializes to quantum circuits with low depth. In such circuits, one can exactly compute low-weight expectation values by restricting to the light cone of the observable of interest. Using this property, we show that one can classically sample from any low-depth circuit with *readout* noise, as long as the output distribution of the circuit anticoncentrates. This holds even when the circuit gates themselves are noiseless.

Theorem E2 (Sampling from low-depth quantum circuits with readout noise; informal). Consider a quantum circuit $\mathcal{C}$ with depolarizing readout noise and a state ρ whose Pauli coefficients can be efficiently computed. Assume the output

distribution anticoncentrates, $\sum_z p_{\rho,\tilde{\mathcal{C}}}(z)^2 = \mathcal{O}(2^{-n})$. If the circuit has depth $d = \mathcal{O}(\log n)$ and is geometrically local, or has depth $d = \mathcal{O}(\log\log n)$ more generally, then there is classical algorithm to sample from $p_{\rho,\mathcal{C}}(z)$ within total variational distance ε in quasipolynomial time.

Quantum advantages for sampling from low-depth circuits are highly sought, since low-depth circuits are generally less susceptible to experimental noise. Nonetheless, our result shows that in many cases even a small amount of readout noise can preclude any such advantage [122].

In the following subsections, we first establish the reduction from noisy circuit sampling to expectation values in the presence of anticoncentration, and then prove the theorems above.

1. Reduction from noisy circuit sampling with anticoncentration to expectation values

To extend our classical algorithms for expectation values to sampling, we leverage the following lemma. The lemma follows almost entirely from Ref. [26], by adapting their Theorem 4 to state ensembles and neglecting the components of the theorem that are specific to IQP circuits.

Lemma E1. Consider any ensemble of states $\mathcal{E} = \{\rho\}$, sampled in the computational basis with readout noise γ per qubit. Let

$$\bar{\alpha} = \frac{1}{|\mathcal{E}|}\sum_{\rho} 2^n \sum_{\mathbf{s}} p_\rho(\mathbf{s})^2 \tag{E1}$$

denote the mean collision probability of the output distribution before readout noise, multiplied by 2^n. Suppose that, for some ℓ_s, one can compute expectation values in ρ of Pauli operators with weight less than ℓ_s to within root-mean-square error ε', in time χ. Then there exists a classical algorithm to simulate sampling to within root-mean-square total variational distance:

$$\varepsilon = 6\sqrt{2}\sqrt{\varepsilon'^2(n^{\ell_s}+1) + \bar{\alpha}e^{-2\gamma\ell_s}}. \tag{E2}$$

The classical algorithm takes time $\mathcal{O}(\chi \cdot n^{\ell_s+1})$ per sample.

The lemma also holds for any fixed state ρ, by taking the ensemble $\mathcal{E} = \{\rho\}$ to be composed of only a single state.

Proof of Lemma E1. Our goal is to sample from the probability distribution:

$$p_{\mathcal{D}\{\rho\}}(\mathbf{s}) = \langle \mathbf{s}|\mathcal{D}\{\rho\}|\mathbf{s}\rangle. \tag{E3}$$

Let us decompose the projector $|\mathbf{s}\rangle\langle\mathbf{s}|$ as a sum of Pauli operators,

$$|\mathbf{s}\rangle\langle\mathbf{s}| = \frac{1}{2^n}\sum_{\mathbf{t}\in\{0,1\}^n} (-1)^{\mathbf{s}\cdot\mathbf{t}} Z_{\mathbf{t}}, \tag{E4}$$

where $\mathbf{s}\cdot\mathbf{t} \equiv \sum_i s_i t_i$, and $Z_{\mathbf{t}} = \bigotimes_{i=0}^n (Z_i)^{t_i}$ denotes the Pauli operator with identity support where $t_i = 0$ and Z support where $t_i = 1$. Since $w[Z_{\mathbf{t}}] = |\mathbf{t}|$, we have

$$\begin{aligned} p_{\mathcal{D}\{\rho\}}(\mathbf{s}) &= \frac{1}{2^n}\sum_{\mathbf{t}} (-1)^{\mathbf{s}\cdot\mathbf{t}} \mathrm{tr}(Z_{\mathbf{t}}\mathcal{D}\{\rho\}) \\ &= \frac{1}{2^n}\sum_{\mathbf{t}} e^{-\gamma|\mathbf{t}|}(-1)^{\mathbf{s}\cdot\mathbf{t}}\mathrm{tr}(Z_{\mathbf{t}}\rho). \end{aligned} \tag{E5}$$

This can be viewed as a Fourier transform over $\mathbb{Z}_2^n$ of the coefficients $e^{-\gamma|\mathbf{t}|}\mathrm{tr}(Z_{\mathbf{t}}\rho)$.

Following Ref. [26], our first step is to form a classical approximation $q_{\mathcal{D}\{\rho\}}(\mathbf{s})$ of the probability distribution $p_{\mathcal{D}\{\rho\}}(\mathbf{s})$. We do so via two steps. First, we truncate all terms in Eq. (E5) with high-weight, $|\mathbf{t}| > \ell_s$. Second, we replace each remaining expectation value $\mathrm{tr}(Z_{\mathbf{t}}\mathcal{D}\{\rho\})$ with its classical estimate. We denote the classical estimate as $a_{\mathbf{t}}$; we have $(1/|\mathcal{E}|)\sum_\rho |a_{\mathbf{t}} - \mathrm{tr}(Z_{\mathbf{t}}\mathcal{D}\{\rho\})|^2 \le \varepsilon'^2$ by assumption. Together, these two steps give a classical approximation:

$$q_{\mathcal{D}\{\rho\}}(\mathbf{s}) \equiv \frac{1}{2^n}\sum_{\mathbf{t}:|\mathbf{t}|\le\ell_s} (-1)^{\mathbf{s}\cdot\mathbf{t}} a_{\mathbf{t}}. \tag{E6}$$

The approximation may take small negative values, and is thus not guaranteed to be a probability distribution. However, one can show that it is close to $p_{\mathcal{D}\{\rho\}}(\mathbf{s})$ in mean-square total variational distance:

$$\begin{aligned} \frac{1}{|\mathcal{E}|}\sum_\rho \Big(\sum_{\mathbf{s}} |p_{\mathcal{D}\{\rho\}(\mathbf{s})} - q_{\mathcal{D}\{\rho\}}(\mathbf{s})|\Big)^2 &\le \frac{2^n}{|\mathcal{E}|}\sum_\rho\sum_{\mathbf{s}} (p_{\mathcal{D}\{\rho\}}(\mathbf{s}) - q_{\mathcal{D}\{\rho\}}(\mathbf{s}))^2 \\ &= \frac{1}{|\mathcal{E}|}\sum_\rho \left(\sum_{|\mathbf{t}|\le\ell_s} \left[\mathrm{tr}(Z_{\mathbf{t}}\mathcal{D}\{\rho\})^2 - a_{\mathbf{t}}\right]^2 + \sum_{|\mathbf{t}|>\ell_s} e^{-2\gamma|\mathbf{t}|}\mathrm{tr}(Z_{\mathbf{t}}\rho)^2\right) \\ &\le \sum_{|\mathbf{t}|\le\ell_s} \varepsilon'^2 + e^{-2\gamma(\ell_s+1)}\cdot\frac{1}{|\mathcal{E}|}\sum_\rho\sum_{|\mathbf{t}|>\ell_s}\mathrm{tr}(Z_{\mathbf{t}}\rho)^2 \le \sum_{|\mathbf{t}|\le\ell_s}\varepsilon'^2 + e^{-2\gamma(\ell_s+1)}\cdot\frac{1}{|\mathcal{E}|}\sum_\rho\sum_{\mathbf{t}}\mathrm{tr}(Z_{\mathbf{t}}\rho)^2 \\ &\le \sum_{|\mathbf{t}|\le\ell_s}\varepsilon'^2 + e^{-2\gamma(\ell_s+1)}\cdot\frac{1}{|\mathcal{E}|}\sum_\rho 2^n\sum_{\mathbf{s}} p_\rho(\mathbf{s})^2 = (n^{\ell_s}+1)\varepsilon'^2 + e^{-2\gamma(\ell_s+1)}\cdot\bar{\alpha}. \end{aligned} \tag{E7}$$

In going from the first to second line, we use Eqs. (E5) and (E6) to expand the first and second copy of $p_{\mathcal{D}\{\rho\}}(\mathbf{s})$, $q_{\mathcal{D}\{\rho\}}(\mathbf{s})$ as a sum over variables $\mathbf{t}$ and $\mathbf{t}'$. Because of the sign factor $(-1)^{\mathbf{s}\cdot\mathbf{t}}(-1)^{\mathbf{s}\cdot\mathbf{t}'}$, the sum over $\mathbf{s}$ gives a δ function $2^n\delta_{\mathbf{t},\mathbf{t}'}$. Using the δ function to eliminate the sum over $\mathbf{t}'$ gives the second line. In going from the fourth to fifth lines, we reverse these steps to convert the second sum over $\mathbf{t}$ to a sum over $\mathbf{s}$. The final line uses the fact that there are at most $n^{\ell_s}+1$ bit strings with $|\mathbf{t}|\le\ell_s$, as well as the definition of the mean collision probability $\bar{\alpha}$.

To translate this approximation into a sampling algorithm, we use Lemma 10 in Ref. [26]. The lemma shows that one can construct a true probability distribution $\tilde{q}_{\mathcal{D}\{\rho\}}(\mathbf{s})$ from $q_{\mathcal{D}\{\rho\}}(\mathbf{s})$, at the cost of only a slightly larger total variational distance,

$$\sum_{\mathbf{s}}|p_{\mathcal{D}\{\rho\}}(\mathbf{s})-\tilde{q}_{\mathcal{D}\{\rho\}}(\mathbf{s})|\le 4\eta_\rho/(1-\eta_\rho),$$
$$\text{where } \eta_\rho\equiv\sum_{\mathbf{s}}|p_{\mathcal{D}\{\rho\}}(\mathbf{s})-q_{\mathcal{D}\{\rho\}}(\mathbf{s})|, \tag{E8}$$

for each ρ. We can translate this to the mean-square total variational distance using

$$\begin{aligned}&\frac{1}{|\mathcal{E}|}\sum_{\rho}\left(\sum\nolimits_{\mathbf{s}}|p_{\mathcal{D}\{\rho\}}(\mathbf{s})-\tilde{q}_{\mathcal{D}\{\rho\}}(\mathbf{s})|\right)^2\\&\le\frac{1}{|\mathcal{E}|}\sum_{\rho:\eta_\rho>1/3}2^2+\frac{1}{|\mathcal{E}|}\sum_{\rho:\eta_\rho\le1/3}\left(\frac{4\eta_\rho}{1-\eta_\rho}\right)^2\\&\le 2^2\frac{\varepsilon^2}{(1/3)^2}+6^2\frac{1}{|\mathcal{E}|}\sum_{\rho:\eta_\rho\le1/3}\eta_\rho^2\le 72\varepsilon^2,\end{aligned} \tag{E9}$$

where we denote the right-hand side of the final line of Eq. (E7) as ε^2. In the first line, we use that the total variational distance is upper bounded by 2 for any two probability distributions, which allows us to replace $4\eta_\rho/(1-\eta_\rho)$ by 2 when $\eta_\rho>1/3$. In the second line, we use Markov's inequality to bound the probability that $\eta_\rho>1/3$.

To sample from $\tilde{q}_{\mathcal{D}\{\rho\}}(\mathbf{s})$, we invoke Sec. 3.2 of Ref. [26]. The section provides a simple sampling algorithm, given the ability to compute marginals of the original approximation $q_{\mathcal{D}\{\rho\}}(\mathbf{s})$. Any such marginal can be computed exactly by expanding $q_{\mathcal{D}\{\rho\}}(\mathbf{s})$ as in Eq. (E6):

$$\begin{aligned}\sum_{\mathbf{s}:\mathbf{s}_{1\cdots k}=\mathbf{y}}q_{\mathcal{D}\{\rho\}}(\mathbf{s})&=\frac{1}{2^n}\sum_{|\mathbf{t}|\le\ell_s}\sum_{\mathbf{s}:\mathbf{s}_{1\cdots k}=\mathbf{y}}(-1)^{\mathbf{s}\cdot\mathbf{t}}a_P\\&=\frac{1}{2^n}\sum_{|\mathbf{t}|\le\ell_s}\delta_{\mathbf{t}_{k+1\cdots n},\mathbf{0}}\cdot(-1)^{\mathbf{y}\cdot\mathbf{t}_{1\cdots k}}\cdot a_P.\end{aligned} \tag{E10}$$

Here, the left-hand side is the marginal distribution on the first k qubits, where $\mathbf{y}\in\{0,1\}^k$. To compute the marginal, one can enumerate each low-weight $\mathbf{t}$ and compute the sum on the right-hand side exactly term by term. This requires time $\chi(n^{\ell_s}+1)$. Generating a single sample requires that we compute one marginal for each value of $k=1,\ldots,n$ [26], and thus requires time $\chi(n^{\ell_s}+1)n$. ∎

2. Proofs of sampling algorithms

We now provide the detailed statements and proofs of our theorems on sampling from noisy quantum circuits. As before, we let $\tilde{\mathcal{C}}$ denote the circuit $\mathcal{C}$ with the readout noise channel omitted. For general circuits, we have the following.

Theorem E3 (Sampling from noisy quantum circuits; formal). Consider a noisy quantum circuit $\mathcal{C}$ and a low-average ensemble $\mathcal{E}=\{\rho\}$, where the Pauli coefficients of each ρ can be efficiently computed. Let $\bar{\alpha}$ denote the mean collision probability of $\tilde{\mathcal{C}}\{\rho\}$ multiplied by 2^n. For circuits with uniform noise, there is a classical algorithm to sample from $\mathcal{C}\{\rho\}$ to within root-mean-square total variational distance ε that runs in time:

$$n^{\mathcal{O}(\gamma^{-3}\log(1/\gamma)\log(\bar{\alpha}/\varepsilon))}. \tag{E11}$$

For circuits with gate-based noise, there is a classical algorithm that runs in time:

$$n^{\mathcal{O}(\gamma^{-2}\log(n)\log(\bar{\alpha}/\varepsilon)+\gamma^{-1}\log(d))}. \tag{E12}$$

In both cases, the run-time is quasipolynomial so long as $\bar{\alpha}=\mathrm{poly}(n)$. For simplicity of presentation, in the main text we take $\bar{\alpha}=\mathcal{O}(1)$, which is referred to as *anticoncentration* [26,31].

Proof of Theorem E3. We take $\ell_s=(1/\gamma)\log(\sqrt{2\bar{\alpha}}/\varepsilon)$ in Lemma E1 to ensure that the second term in the error is small. This requires $2\varepsilon'^2\le\varepsilon^2/(n^{\ell_s}+1)$ for the first term. For circuits with uniform noise, we can compute the expectation value of any Pauli operator in time $\chi=(1/\varepsilon')^{\mathcal{O}(\log(1/\gamma)/\gamma^2)}$ (Theorem 1). From Lemma E1, this gives a total run-time

$$\mathcal{O}(\chi\cdot n^{\ell_s+1})=(n^{\ell_s}/\varepsilon^2)^{\mathcal{O}(\log(1/\gamma)/\gamma^2)}=n^{\mathcal{O}(\gamma^{-3}\log(1/\gamma)\log(\bar{\alpha}/\varepsilon))}, \tag{E13}$$

which establishes Eq. (E11). Similarly, for circuits with gate-based noise, setting χ according to Eq. (D28) gives a total run-time

$$n^{\mathcal{O}((1/\gamma)\log(dn^{\ell_s}/\varepsilon))}=n^{\mathcal{O}(\gamma^{-2}\log(n)\log(\bar{\alpha}/\varepsilon)+\gamma^{-1}\log(d))}, \tag{E14}$$

which establishes Eq. (E12). ∎

For low-depth circuits, we have the following.

Theorem E4 (Sampling from low-depth quantum circuits with readout noise; formal). Consider any circuit $\mathcal{C}$ that is either depth $d=\mathcal{O}(\log(n))$, and geometrically local, or

depth $d = \mathcal{O}\log(\log(n))$, and contains readout noise γ per qubit. Let ρ be any state whose Pauli coefficients can be efficiently computed, and α denote the collision probability of $\tilde{\mathcal{C}}\{\rho\}$ multiplied by 2^n. There is classical algorithm to sample from $\mathcal{C}\{\rho\}$ within total variational distance ε in time,

$$2^{\mathcal{O}(\gamma^{-1}\log(\bar{\alpha}/\varepsilon)(\log n)^c)}, \tag{E15}$$

where $c = D$ for geometrically local circuits in spatial dimension D, and $c = \mathcal{O}(1)$ for general circuits.

Proof of Theorem E4. The theorem follows from a simple observation: that expectation values of low-weight Pauli operators can be computed exactly in any circuit of sufficiently low depth. More precisely, any Pauli operator with initial weight at most ℓ_s can have support on at most $\mathcal{O}(\ell_s \cdot d^D)$ or $\mathcal{O}(\ell_s \cdot 2^d)$ qubits after Heisenberg time evolution. The first scaling applies to geometrically local circuits, and the second to any circuit. The expectation value of the Pauli operator can therefore be computed exactly in time $2^{\mathcal{O}(\ell_s \cdot d^D)}$ or $2^{\mathcal{O}(\ell_s \cdot 2^d)}$. Therefore, setting $\varepsilon' = 0$ and $\ell_s = \gamma^{-1}\log(6\sqrt{2\bar{\alpha}}/\varepsilon)$ in Lemma E1, one can classically sample from the circuit in time $2^{\mathcal{O}(\ell_s \cdot d^D)+\mathcal{O}(\ell_s \cdot \log(n))}$ or $2^{\mathcal{O}(\ell_s \cdot 2^d)+\mathcal{O}(\ell_s \cdot \log(n))}$. Setting $d = \mathcal{O}(\log(n))$ for geometrically local circuits, the first scaling becomes $2^{\mathcal{O}(\ell_s \cdot \log(n)^D)} = 2^{\mathcal{O}(\gamma^{-1}\log(\bar{\alpha}/\varepsilon)\cdot\log(n)^D)}$. Meanwhile, setting $d = \mathcal{O}(\log(\log(n)))$ for general circuits, the second scaling becomes $2^{\ell_s \cdot \log(n)^{\mathcal{O}(1)}} = 2^{\gamma^{-1}\log(\bar{\alpha}/\varepsilon)\cdot\log(n)^{\mathcal{O}(1)}}$. This concludes our proof. ∎

APPENDIX F: EXTENSION TO ANY SINGLE-QUBIT NOISE CHANNEL IN THE DEPOLARIZING CLASS

In this appendix, we briefly discuss the extension of our results to any single-qubit noise channel in the depolarizing class. The depolarizing class of noise channels corresponds to all channels whose sole steady state is the maximally mixed state. Any single-qubit noise channel $\mathcal{N}$ in the depolarizing class can be written as a composition of unitary operations and a Pauli noise channel [14], $\mathcal{N}(\cdot) = U\mathcal{N}_{\text{Pauli}}(V(\cdot)V^\dagger)U^\dagger$, where $U, V \in U(2)$ and $\mathcal{N}_{\text{Pauli}}(P) = e^{-\gamma_P}P$ for each Pauli $P \in \{X, Y, Z\}$. Here, $\mathcal{N}_{\text{Pauli}}(\mathbb{1}) = \mathbb{1}$ and $\gamma_P > 0$.

We consider any quantum circuit with either uniform or gate-based noise, in which each single-qubit noise channel is in the depolarizing class. The noise channel, $\mathcal{N}_{i,t}$ (specified by $U_{i,t}, V_{i,t}, \gamma_P^{i,t}$), may depend arbitrarily on the qubit and circuit layer. We assume only that each Pauli noise rate is at least constant, $\gamma_P^{i,t} = \Omega(1)$, with respect to the system size n.

Our classical algorithms generalize as follows. We set $\gamma = \min_{i,t}\gamma_P^{i,t} = \Omega(1)$ to be the smallest Pauli noise rate. We can then rewrite each noise channel as $\mathcal{N}_{i,t} = \mathcal{D}_\gamma \circ \tilde{\mathcal{N}}_{i,t}$, where $\tilde{\mathcal{N}}_{i,t}$ is the same noise channel with each Pauli noise rate $\gamma_P^{i,t}$ replaced by $\gamma_P^{i,t} - \gamma$. Note that $\tilde{\mathcal{N}}_{i,t}$ is a valid quantum channel, since $\gamma_P^{i,t} - \gamma \geq 0$ for all i, t, P. Since $\tilde{\mathcal{N}}_{i,t}$ is a single-qubit channel, we can absorb its action into the action of the gate preceding it. In this way, one can rewrite the original circuit with arbitrary noise channels in the depolarizing class as a new circuit with only depolarizing noise channels. Our classical algorithms immediately apply to the new circuit, and hence to the original circuit as well.

APPENDIX G: EXTENSION TO CIRCUITS WITH NOISE ONLY IN NON-CLIFFORD GATES

In this appendix, we show that both of our classical algorithms naturally extend to settings where noise occurs only in non-Clifford gates, and Clifford gates are implemented noiselessly. Intuitively, this follows because our algorithms track operators in the Pauli basis. Clifford gates permute the Pauli operators, and thus do not increase the complexity of our classical approximation. We formalize this in the following extension of our main result.

Extension G1 (Classical algorithm for noisy quantum circuits with non-Clifford noise). Consider the settings of Theorems 1 and 2 of the main text but where (i) for the uniform noise model, noise occurs only in circuit layers that contain non-Clifford gates and during readout, and (ii) for the gate-based noise model, noise occurs only in non-Clifford gates and during readout. Then all of the stated results continue to hold, with the scaling in Theorem 2 replaced by

$$\mathcal{O}\big(d \cdot m^2 \cdot n^{(2/\gamma)\log(\sqrt{d+1}/\varepsilon)}\big), \tag{G1}$$

where d is the number of circuit layers containing non-Clifford gates, and m is the number of Pauli operators with nonzero coefficients in O and weight less than or equal to ℓ.

The result continues to hold even if the readout is noiseless, provided that O is a sum of polynomially many Pauli operators, $m = \text{poly}(n)$.

Proof of Extension G1. We provide separate proofs for our algorithms for uniform and gate-based noise

Uniform noise. Let us begin with our algorithm for uniform noise. We use the same approximation $\tilde{O}$ of the Heisenberg time-evolved operator as in Eq. (C2), but where $\mathcal{U}_t$ now corresponds to the unitary of the tth non-Clifford layer combined with all of the ensuing Clifford layers before the $(t+1)$th non-Clifford layer. One can verify that Eq. (C2) still holds, because it relies only on the super-unitarity of $\mathcal{U}_t$. Therefore, as before, the approximation $\tilde{O}$ is accurate with small Frobenius error [Eq. (C6)] provided that we choose the truncation weight ℓ as in Eq. (C9). Viewing Eq. (C9), the depth d now corresponds to the number of non-Clifford layers, and, as before, we have

$d < \gamma^{-1}\log(1/\varepsilon)$ before the system approaches the maximally mixed state.

It remains to show that our algorithm remains efficient—i.e., that the scaling of the number of Pauli paths in Eq. (C2) is unchanged. Following Ref. [31], we count the number of Pauli paths for each fixed sequence $(\ell_0, \ldots, \ell_d)$ individually, where $\ell_0 + \cdots + \ell_d \leq \ell$, since there are at most 2^ℓ such sequences. One starts from the non-Clifford layer with the smallest weight ℓ_t, at the beginning of which there are at most $(3n)^{\ell/d}$ possible Pauli operators (since one must have $\ell_t \leq \ell/d$). If O is a sum of polynomially many Pauli operators, we start instead at the final circuit layer $t = 0$, in which case this factor is replaced by m. The ensuing non-Clifford layer can "split" each Pauli operator to at most $2^{\mathcal{O}(\ell_t)}$ new Pauli operators, whose number is unchanged by the following Clifford layers. Thus, there are at most $\min\left((3n)^{\ell/d}, m\right) \cdot 2^{\mathcal{O}(\ell_t)}$ possible Pauli operators at the beginning of layer $t+1$. Among these, we count only those of weight ℓ_{t+1}. At the next non-Clifford layer, each of these Pauli operators can split to at most $2^{\mathcal{O}(\ell_{t+1})}$ new Pauli operators, and so on. Iterating this argument, our upper bound on the number of Pauli paths is exactly the same as before [Eq. (C10)], which leads to the same run-time as in Theorem 1 of the main text.

Gate-based noise. We now turn to our algorithm for gate-based noise. We incorporate the non-Clifford noise model into our algorithm by modifying our definition of the Pauli weight. In our original algorithm, the Pauli weight served as a lower bound on the number of *two-qubit gates* required to grow our initial operator into a Pauli operator of interest. Since each two-qubit gate damped the operator amplitude by a factor $e^{-\gamma}$ due to noise, we formed an accurate classical approximation of our time-evolved operator by keeping track only of low-weight Pauli operators. Applying this intuition to the non-Clifford noise model, we should introduce a modified weight that lower bounds the number of *non-Clifford gates* required to transform our initial operator into some Pauli of interest. This modification will depend on the circuit architecture (i.e., the placement of the Clifford and non-Clifford gates) as well as the choice of Clifford gates.

To write this down explicitly, for each observable O, circuit $\mathcal{C}$, and circuit layer t, we define the *non-Clifford distance* $w_{\mathrm{nc}}[P; \mathcal{C}, t]$ as the minimal number of non-Clifford gates in any Pauli path from Q at layer zero to P at layer t, where Q is any Pauli operator with a nonzero coefficient in O. For example, at layer zero, the set of Pauli operators in O have non-Clifford distance zero and all other Pauli operators have non-Clifford distance ∞. At later layers, any Pauli operator that can be reached from Q in O via solely Clifford operations also has non-Clifford distance zero, at the layers at which it can be reached. We emphasize that the definition of the non-Clifford distances changes as we move through the circuit; for example, a Pauli operator Q in O may have distance zero at time zero but distance > 0 at time one, depending on the Clifford gates between time zero and one.

From this definition, we can also define a non-Clifford distance *distribution*, $P_{\mathrm{nc}}^{(g)}(w)$, and cumulative non-Clifford distance distribution, $Q_{\mathrm{nc}}^{(g)}(w)$, for the time-evolved operator O. These measure the norm of $O^{(g)}$ on Pauli operators with distance w at circuit gate g, and are analogous to the weight distribution and cumulative weight distribution defined previously. (One can think of this as working in a "rotating frame" determined by the action of the Clifford gates.) From the definition of the non-Clifford distance, the distribution $P_{\mathrm{nc}}^{(g)}(w)$ begins as a δ function at zero. The distributions do not change under the action of any Clifford gate, and shift by at most one under each non-Clifford gate.

With these definitions, one can easily repeat our proof of Lemma D1 with non-Clifford gate-based noise instead of two-qubit gate-based noise, and Clifford distance instead of Pauli weight. The only change is the factor of $e^{-4\gamma}$ is replaced with $e^{-2\gamma}$ throughout the proof, where the replacement arises because an operator can increase in non-Clifford distance despite having support only on one qubit in the gate and not two. The end result is that at any layer t, the cumulative non-Clifford distance distribution is less than $Q_{\mathrm{nc}}^{(t)}(w) \leq e^{-2\gamma(w+1)}||O||_F^2$ for all w. If we form a low-distance approximation,

$$\tilde{O}_{\mathrm{nc}}^{(t)} = \sum_{P:\, w_{\mathrm{nc}}[P;\mathcal{C},t] \leq \ell} \tilde{c}_P^{(t)} P, \tag{G2}$$

at each layer t, then we similarly have $\sum_{t=0}^{d} \tilde{Q}_{\mathrm{nc}}^{(t)}(\ell) \leq e^{-2\gamma(\ell+1)}||O||_F^2$.

It remains only to bound the run-time of our classical approximation. This is determined by the number of low-non-Clifford-distance Pauli operators that we must keep track of. We begin with the operator O acted on by the readout noise channel. Because of the readout noise, we can truncate all Pauli operators in O with weight above ℓ. From this starting point, each non-Clifford gate can increase the number of accessible Pauli operators by at most a constant factor C: for T gates by a factor of 2, for general two-qubit non-Clifford gates by a factor of 15. There are $\binom{n_{\mathrm{nc}}}{\ell}$ sets of ℓ non-Clifford gates. Thus, there are at most $m\binom{n_{\mathrm{nc}}}{\ell}C^\ell \leq mn_{\mathrm{nc}}^\ell C^\ell/\ell! = \mathcal{O}(mn_{\mathrm{nc}}^\ell)$ possible Pauli operators with non-Clifford distance $\leq \ell$ at any circuit layer, where m is the number of Pauli operators with nonzero coefficients in O and weight $\leq \ell$. The coefficients can be updated one by one in time $\mathcal{O}(m^2 n_{\mathrm{nc}}^{2\ell})$. Repeating d times for each circuit layer, we arrive at the scaling in Eq. (G1). ∎

APPENDIX H: EXTENSION TO RANDOM NONUNITAL NOISE

We now consider the extension of our classical algorithms to quantum circuits with nonunital noise. Unital noise corresponds to any noise process that maps the maximally mixed state to itself, a primary example being depolarizing

noise. Nonunital noise refers to noise processes that do not preserve the maximally mixed state, such as spontaneous emission noise (also referred to as amplitude damping).

Interestingly, the arguments in Ref. [14] prevent any immediate extension of our results to nonunital noise. For any local nonunital noise model, Ref. [14] provides a scheme to perform fault-tolerant quantum computation for exponentially long times. Their scheme assumes that the circuit begins in the zero state, in contrast to the input state ensembles that we consider. However, for noise models that have a product state fixed point, such as spontaneous emission, this can easily be circumvented. One simply needs to wait for some time $\mathcal{O}(\gamma^{-1}\log(n/\varepsilon))$ before applying the circuit in Ref. [14], since after this time any input state will have become close to the fixed point.

While these arguments exclude classical algorithms for arbitrary quantum circuits with nonunital noise, recent work has shown that this may not be the case for restricted classes of circuits [36]. In particular, Ref. [36] introduces a classical algorithm for expectation values in *random circuits* with nonunital noise. By taking each gate in the circuit to be independently random, one explicitly excludes the scheme in Ref. [14] which opens the door to an efficient classical simulation. The algorithm in Ref. [36] runs in time $\exp(\mathcal{O}\{(\gamma^{-1}\log(\varepsilon^{-1}))^D\})$ for local observables in geometrically local circuits in dimension D. This is polynomial in n for $D = 1$, and quasipolynomial for larger D. For nonlocal observables, the run-time is instead exponential in n, and for circuits without geometric locality, it is exponential in ε^{-1}.

Our classical algorithms improve upon this performance in two ways. First, we show that, in fact, a much milder form of randomness is sufficient to achieve an efficient classical simulation. We elaborate on this in the following paragraph. Second, we show that, under this mild randomness, our classical algorithms achieve run times matching those in Theorems 1 and 2. Namely, our algorithm runs in *polynomial time* for *any* circuit architecture for large classes of observables, and quasipolynomial time otherwise.

To elaborate on the first point, we consider quantum circuits with *arbitrary* gates, and require only that each nonunital noise channel occurs in a *random direction*. Considering spontaneous emission specifically, our requirement translates to choosing each local spontaneous emission channel to emit from either the 1 to 0 state,

$$\mathcal{A}_+\{\rho\} = \rho + \gamma_s\langle 1|\rho|1\rangle|0\rangle\langle 0| - \frac{\gamma_s}{2}|1\rangle\langle 1|\rho - \frac{\gamma_s}{2}\rho|1\rangle\langle 1|, \tag{H1}$$

or the 0 to 1 state,

$$\mathcal{A}_-\{\rho\} = \rho + \gamma_s\langle 0|\rho|0\rangle|1\rangle\langle 0| - \frac{\gamma_s}{2}|0\rangle\langle 0|\rho - \frac{\gamma_s}{2}\rho|0\rangle\langle 0|, \tag{H2}$$

with equal probability. This randomization is performed implicitly in random circuits, via the random gates before and after each noise channel. We emphasize that we are considering simulating a specific instance of the random nonunital noise (with high success probability over the random instances), and not a mixture of the instances. Finally, as with our previous results, we consider the performance of our algorithm over an ensemble of input states.

This mild form of randomness neatly avoids the no-go results of Ref. [14], by ensuring that the gates in the circuit are decoupled from the structure of the nonunital noise. Thus, even though nonunital noise is present, the circuit cannot "take advantage" of it. Within this setting, we show that our classical algorithms in Theorems 1 and 2 of the main text apply equally well to circuits with nonunital noise.

Extension H1 (Classical algorithm for random nonunital noise). Consider the settings of Theorems 1 and 2 of the main text, but with each local depolarizing channel $\mathcal{D}_i$ replaced by spontaneous emission $\mathcal{A}_{i,\pm}$ in independently random directions with rate $\gamma_s \le 4/7$. Replace each error metric with its root-mean-square value over the random direction. Then all of the stated results continue to hold, with $\gamma = \log\left(1/(1-\gamma_s/2)\right) \approx \gamma_s/2$ and the factor of d in Eq. (3) of the main text replaced by 1.

If one considers instead a random quantum circuit with uniform spontaneous emission noise, for any circuit architecture, then the run-time of our classical algorithm improves to

$$\mathrm{poly}(n)\cdot(1/\varepsilon)^{\mathcal{O}(1/\gamma)}, \tag{H3}$$

if O is a sum of polynomially many Paulis operator,

$$n^{(1/\gamma)\log(1/\varepsilon)}\cdot(1/\varepsilon)^{\mathcal{O}(1/\gamma)}, \tag{H4}$$

for any O.

We make a few additional remarks. First, for simplicity of analysis, we restrict attention to spontaneous emission noise as the prototypical example of nonunital noise. We expect our results to generalize straightforwardly to any strictly contractive nonunital noise channel [123]. Second, the restriction to noise rates below a large threshold, $\gamma_s \le 4/7$, is also chosen as a convenient simplification for our proof. We do not expect larger noise rates to lead to harder classical simulations. Finally, as a stepping stone in our proof, we prove a general lemma (Lemma H1) showing that expectation values in circuits with random nonunital noise depend only on the final $\mathcal{O}(\gamma^{-1}\log(n/\varepsilon))$ layers of the circuit for most input states. This extends a result of Ref. [36], and follows extremely quickly after we introduce the general framework of our proof.

Proof of Extension H1. The strategy of our proof is to show that, while nonunital noise can increase the Frobenius norm of Heisenberg time-evolved operators in certain

instances, it decreases the norm *on average* over the random directions of the noise channel. Thus, typical instances of nonunital noise share key properties in common with depolarizing noise.

To show this, let us begin by writing down the action of the spontaneous emission channel in the Heisenberg picture:

$$\begin{aligned}\mathcal{A}_{\pm}^{\dagger}\{\mathbb{1}\} &= \mathbb{1},\\ \mathcal{A}_{\pm}^{\dagger}\{X\} &= (1-\gamma_s/2)X,\\ \mathcal{A}_{\pm}^{\dagger}\{Y\} &= (1-\gamma_s/2)Y,\\ \mathcal{A}_{\pm}^{\dagger}\{Z\} &= (1-\gamma_s)Z \pm \gamma_s\mathbb{1}.\end{aligned} \tag{H5}$$

The key idea we use is to decompose the spontaneous emission channel into the product,

$$\mathcal{A}_{\pm}^{\dagger} = \tilde{\mathcal{A}}_{\pm}^{\dagger} \bigcirc \mathcal{D}, \tag{H6}$$

where the superoperator $\tilde{\mathcal{A}}$ is defined via

$$\begin{aligned}\tilde{\mathcal{A}}_{\pm}^{\dagger}\{\mathbb{1}\} &= \mathbb{1},\\ \tilde{\mathcal{A}}_{\pm}^{\dagger}\{X\} &= X,\\ \tilde{\mathcal{A}}_{\pm}^{\dagger}\{Y\} &= Y,\\ \tilde{\mathcal{A}}_{\pm}^{\dagger}\{Z\} &= \frac{1-\gamma_s}{1-\gamma_s/2}Z \pm \frac{\gamma_s}{1-\gamma_s/2}\mathbb{1},\end{aligned} \tag{H7}$$

and $\mathcal{D}$ is the depolarizing channel of strength $\gamma = -\log(1-\gamma_s/2) \approx \gamma_s/2$.

We will now show that, in expectation, $\tilde{\mathcal{A}}_{\pm}^{\dagger}$ does not increase the Frobenius norm of any operator. Suppose that $\tilde{\mathcal{A}}_{\pm}^{\dagger}$ acts on one qubit of a larger system. Let us decompose an arbitrary operator O in terms of its support on that qubit, $O = \mathbb{1}\otimes O_{\mathbb{1}} + X\otimes O_X + Y\otimes O_Y + Z\otimes O_Z$. Considering the mean-square Frobenius norm over the two possible directions of the noise channel, we have

$$\begin{aligned}\mathbb{E}_{\pm}\left[||\tilde{\mathcal{A}}_{\pm}^{\dagger}\{O\}||_F^2\right] &= ||O_{\mathbb{1}}||_F^2 + ||O_X||_F^2 + ||O_Y||_F^2 + \left(\frac{(1-\gamma_s)^2}{(1-\gamma_s/2)^2} + \frac{\gamma_s^2}{(1-\gamma_s/2)^2}\right)||O_Z||_F^2\\ &\le ||O||_F^2 + \left(\frac{(1-\gamma_s)^2}{(1-\gamma_s/2)^2} + \frac{\gamma_s^2}{(1-\gamma_s/2)^2} - 1\right)||O_Z||_F^2 = ||O||_F^2 - \gamma_s\left(\frac{1-7\gamma_s/4}{(1-\gamma_s/2)^2}\right)||O_Z||_F^2 \le ||O||_F^2,\end{aligned} \tag{H8}$$

where the final inequality holds if $\gamma_s \le 4/7$. In the first line, we use that the cross terms between Z and 1 in $\tilde{\mathcal{A}}_{\pm}^{\dagger}\{Z\}$ vanish in expectation.

To connect this to the accuracy of our classical algorithms, let us absorb each superoperator $\tilde{\mathcal{A}}_{\pm}$ into its adjacent unitary U_t, leaving only the depolarizing channels $\mathcal{D}$ as the noise channels. We then analyze each setting in Extension H1 as follows.

Gate-based noise. For our Algorithm 2 for gate-based noise, the only properties of the unitaries U_t that we used in our proofs were that (i) each gate in U_t cannot increase the weight of an operator by more than 1 and (ii) each gate cannot increase the Frobenius norm of an operator. Since the superoperators $\tilde{\mathcal{A}}_{\pm}$ do not increase the operator weight, we are clearly allowed to absorb them into the U_t with respect to condition (i). In regards to condition (ii), by viewing our previous proof, one sees that the mean-square error over the input state ensemble depends on only the Frobenius norms of the truncated operators. Therefore, when taking the mean-square error over the random nonunital noise in addition, the error will depend only on the mean-square Frobenius norm over the random noise channels. We have just shown that the superoperators $\tilde{\mathcal{A}}_{\pm}$ do not increase the mean-square Frobenius norm. Thus, we are free to absorb the superoperators $\tilde{\mathcal{A}}_{\pm}$ into the unitaries U_t. Our classical algorithm thus succeeds with run-time as in Theorem 2 of the main text, with $\gamma = -\log(1-\gamma_s/2) \approx \gamma_s/2$.

Uniform noise. For our Algorithm 1 for uniform noise, we require three small additional modifications. The first two modifications arise because the presence of nonunital noise changes our counting of the number of valid Pauli paths, since nonunital noise can transform nonidentity Pauli operators to identity Pauli operators (but not vice versa) in the Heisenberg picture. This requires us to take Pauli paths with sequences of weights $(\ell_0,\ldots,\ell_t,0,\ldots,0)$ into consideration, where $\ell_0+\cdots+\ell_t \le \ell$ and $t>0$. As our first modification, when bounding the error in Algorithm 1, these new Pauli paths lead to an increased prefactor of $\sum_{t=0}^{d}\binom{\ell}{t}$ instead of $\binom{\ell}{d}$. To derive this, note that in the first sum of Eq. (C6) we must now sum from $w_d = 0$ to $\ell - d + 1$ instead of $w_d = 1$ to $\ell - d + 1$. If $w_d > 0$, the entire bound proceeds as is and contains $\binom{\ell}{d}$ individual truncations. If $w_d = 0$, then we are left with a bound resembling our original quantity of interest, but now with d layers instead of $d+1$. One can then apply the same argument to w_{d-1} and so on. For instance, the Pauli paths with $w_d = 0$ and $w_{d-1} > 0$ can be grouped into $\binom{\ell}{d} - 1$ individual truncations, and the Pauli paths with $w_d = w_{d-1} = 0$ and $w_{d-2} > 0$ can be grouped into $\binom{\ell}{d} - 2$ individual truncations, and so on. In the end, this increased prefactor does not change the stated run-time of our algorithm, since when determining the required value of ℓ we used the inequality $\binom{\ell}{d} \le (e\ell/d)^d$ [Eq. (C8)]. This inequality is in

fact somewhat weak, and is also obeyed by the partial sum of binomial coefficients, $\sum_{t=0}^{d}\binom{\ell}{t} \le (e\ell/d)^d$.

Our second modification consists of a slight increase in the algorithm run-time due to the need to keep track of the new Pauli paths. When O is a sum of polynomially many m Pauli operators, nonunital noise increases the number of Pauli paths by at most a factor of d, corresponding to the extra sum over $t = 1, \ldots, d$. For more general O, the sum over t (with ℓ fixed) replaces the factor $n^{\ell/d}$ with $\sum_{t=1}^{d} n^{\ell/t} \le d \cdot n^{\ell}$.

Finally, we must slightly modify our argument (Lemma C1) for restricting attention to circuits of depth $d < \gamma^{-1}\log(1/\varepsilon)$. To do so, we provide a simple proof that observables in any circuit with random nonunital noise depend only on the final $\mathcal{O}[\gamma^{-1}\log(1/\varepsilon)]$ layers of the circuit, for most input states. This was proven recently for random quantum circuits using other techniques [36].

Lemma H1 (Extension of Lemma C1 to random nonunital noise). Consider any observable O and any circuit $\mathcal{C}$ with uniform noise, where each noise channel corresponds to spontaneous emission in a random direction with rate $\gamma_s \le 4/7$. The root-mean-square Frobenius norm of the nonidentity component of $\mathcal{C}^{\dagger}\{O\}$ is upper bounded by $e^{-\gamma(d+1)}||O||_F$, where d is the circuit depth and $\gamma = \log\left(1/(1-\gamma_s/2)\right) \approx \gamma_s/2$.

Proof. We decompose each noise channel as $\mathcal{A}_{\pm}^{\dagger} = \tilde{\mathcal{A}}_{\pm}^{\dagger} \bigcirc \mathcal{D}$, and apply Eq. (H8) to the proof of Lemma C1. ∎

The lemma implies that we can approximate $\mathcal{C}^{\dagger}\{O\}$ for depths $d \ge d^* = \gamma^{-1}\log(1/\varepsilon)$ by its identity component, $\mathrm{tr}(O^{(d^*)})/2^n \cdot \mathbb{1}$, at layer d^*. This follows because the nonidentity components have small root-mean-square Frobenius norm, and the layers after d^* map the identity operator to itself in the Heisenberg picture. From Lemma B1, this approximation suffices to compute expectation values to within root-mean-square error $\varepsilon \cdot \sqrt{c} \cdot ||O||_F$ over any low-average ensemble of input states with purity c.

Random quantum circuits with uniform noise. Finally, we demonstrate the improved run-time of our Algorithm 1 when applied to random quantum circuits. Compared to the general quantum circuits that we consider, the main simplification that occurs for random quantum circuits is that our error bound for Algorithm 1 improves to $e^{-\gamma(\ell+1)}||O||_F$ instead of $e^{-\gamma(\ell+1)}\sqrt{\sum_{t=0}^{d}\binom{\ell}{t}}||O||_F$. This follows because different Pauli paths, on average, do not coherently interfere in random quantum circuits (we refer to Ref. [31] for a comprehensive discussion). This improved error bound allows us to choose a much smaller threshold $\ell = \gamma^{-1}\log(1/\varepsilon)$. This gives a total number of low-weight Pauli paths:

$$\min(d\cdot n^{\ell}, d\cdot m)\cdot 2^{\mathcal{O}(\ell)} = d\cdot\min(n^{(1/\gamma)\log(1/\varepsilon)}, m)\cdot(1/\varepsilon)^{\mathcal{O}(1/\gamma)}. \tag{H9}$$

The factor of d can be absorbed into the factor $(1/\varepsilon)^{\mathcal{O}(1/\gamma)}$, giving our stated run-time. ∎

APPENDIX I: IMPLICATIONS FOR QUANTUM ERROR MITIGATION

In this appendix, we prove Corollary 1 of the main text. We note that the corollary captures nearly all error mitigation strategies of interest, including zero-noise extrapolation [60,61], probabilistic error cancellation [60], and Clifford data regression [85]. One notable exception is *virtual distillation* protocols [86,87], which evade the settings of our work by performing experiments involving multiple copies of each input state. Another (partial) exception is noise *suppression* strategies, such as dynamical decoupling [124,125], which serve to decrease the effective noise rate γ (and thus increase the potential classical complexity), by addressing the microscopic, unitary source of the noise. We remark that, in the corollary, the root-mean-square error for an error mitigation strategy is necessarily taken over both the ensemble of input states as well as the probabilistic measurement outcomes of the noisy quantum experiments themselves.

Proof of Corollary 1. We denote the set of noisy quantum circuits as $\{\mathcal{C}'\}$ and the number of circuits as $M = |\{\mathcal{C}'\}|$. To prove the corollary, we adapt our classical algorithms to allow sampling from the measurement outcomes of O in each of the M noisy circuits $\mathcal{C}'$. Without loss of generality, we can assume that the measurements occur in the eigenbasis of O, so that each outcome is some eigenvalue λ. The outcomes are received with probabilities $p_{\mathcal{C}'\{\rho\}}(\lambda) = \mathrm{tr}(P_\lambda \mathcal{C}'\{\rho\})$, where P_λ is the projector onto the λ eigenspace of O. Thus, to classically reproduce the result of the error mitigation strategy, we need to sample once from $p_{\mathcal{C}'\{\rho\}}(\lambda)$ for each $\mathcal{C}'$. We will show that this can be done with minimal overhead for Pauli observables, since each λ takes only two possible eigenvalues ± 1.

To sample from each individual $p_{\mathcal{C}'\{\rho\}}(\lambda)$, we first classically compute the expectation value $\mathrm{tr}(O\mathcal{C}'\{\rho\})$ to within error ε'. Denote our classical approximation as $a_{\mathcal{C}'}$. From this, we approximate the true outcome distribution $p_{\mathcal{C}'\{\rho\}}(\pm 1)$ via $q_{\mathcal{C}'\{\rho\}}(\pm 1) = (1 \pm a_{\mathcal{C}'})/2$. The approximation has total variational distance $\sum_{\pm}|p_{\mathcal{C}'\{\rho\}}(\pm 1) - q_{\mathcal{C}'\{\rho\}}(\pm 1)| \le \varepsilon'$. We can sample from $q_{\mathcal{C}'\{\rho\}}(\pm 1)$ efficiently since it takes only two values.

We now analyze how the individual sampling errors propagate when we sample from all M noisy circuits $\{\mathcal{C}'\}$ in succession. The ideal outcome distribution is the tensor product $p(\mathbf{r}) = \bigotimes_{m=1}^{M} p_{\mathcal{C}'_r\{\rho\}}(r_i)$. Here, $\mathbf{r} = (r_1, \ldots, r_M)$ for $r_m \in \{-1, 1\}$ is the sequence of measurement outcomes, and $p(\mathbf{r})$ is the probability for a set of M experiments to receive the sequence. Our approximation is the tensor product $q(\mathbf{r}) = \bigotimes_{m=1}^{M} q_{\mathcal{C}'_r\{\rho\}}(r_i)$. We will bound the total variational distance between $p(\mathbf{r})$ and $q(\mathbf{r})$. Denote the total variational distance as $\mathrm{TVD}(p, q) = \sum_{\mathbf{r}}|p(\mathbf{r}) - q(\mathbf{r})|$. We

use the fact that the total variational distance is at most additive under the tensor product:

$$\mathrm{TVD}(p_1 \otimes p_2, q_1 \otimes q_2) \leq \mathrm{TVD}(p_1, q_1) + \mathrm{TVD}(p_2, q_2). \tag{I1}$$

Applied M times, this gives $\mathrm{TVD}(p, q) \leq M\varepsilon'$.

To conclude our proof, we note that any error mitigation strategy must assign some estimate $a(\mathbf{r})$ of $\mathrm{tr}(O\mathcal{C}\{\rho\})$, for every possible sequence $\mathbf{r}$ of measurement outcomes. Without loss of generality, we can assume that $|a(\mathbf{r})| \leq 1$, since $||O||_\infty = 1$ for any Pauli operator. Now, we can utilize the same assignment $a(\mathbf{r})$ for our classical algorithm. To bound the mean-square error in our classical estimate, we use

$$\begin{aligned}\sum_{\mathbf{r}} q(\mathbf{r})(O(\mathbf{r}) - \mathrm{tr}(O\mathcal{C}\{\rho\}))^2 &= \sum_{\mathbf{r}} p(\mathbf{r})(O(\mathbf{r}) - \mathrm{tr}(O\mathcal{C}\{\rho\}))^2 + \sum_{\mathbf{r}} (p(\mathbf{r}) - q(\mathbf{r})) \cdot (O(\mathbf{r}) - \mathrm{tr}(O\mathcal{C}\{\rho\}))^2 \\ &\leq \sum_{\mathbf{r}} p(\mathbf{r})(O(\mathbf{r}) - \mathrm{tr}(O\mathcal{C}\{\rho\}))^2 + 4\mathrm{TVD}(p, q) \leq \sum_{\mathbf{r}} p(\mathbf{r})(O(\mathbf{r}) - \mathrm{tr}(O\mathcal{C}\{\rho\}))^2 + 4M\varepsilon'. \end{aligned} \tag{I2}$$

The first term is the root-mean-square error in the error mitigation strategy. Thus, if error mitigation can compute the observable to error $\varepsilon/2$, our classical algorithm can do so to root-mean-square error ε by taking $\varepsilon' = \varepsilon/8M$. By assumption, ε^{-1} and M are both polynomial in n. Hence, ε'^{-1} is also polynomial in n. Thus, by Theorem 1, the computation of $\mathrm{tr}(O\mathcal{C}'\{\rho\})$ above can be done in polynomial time, and so the ideal expectation value can also be computed in polynomial time. ∎

APPENDIX J: TEST FOR THE HARDNESS OF A GIVEN QUANTUM CIRCUIT

Let us now turn to Corollary 2. In general, classical simulation of noiseless quantum circuits will require exponential resources. However, this is not guaranteed for every such circuit. Motivated by modern quantum experiments, in the main text we proposed Corollary 2 as a simple test for whether a given quantum circuit has the potential to achieve a quantum advantage.

Proof of Corollary 2. We denote the noiseless quantum circuit as $\mathcal{C}$, and the same circuit with noise as $\mathcal{C}_\gamma$ for a noise rate γ. Let $O^{(d)} = \mathcal{C}^\dagger\{O\}$ denote the Heisenberg time-evolved observable under the noiseless circuit, $O_\gamma^{(d)} = \mathcal{C}_\gamma^\dagger\{O\}$ under the noisy circuit, and $\tilde{O}^{(d)}$ be the approximation obtained by applying our classical algorithm to $\mathcal{C}_\gamma$. For each input state ρ, we have

$$\begin{aligned}|\mathrm{tr}(\rho O^{(d)}) - \mathrm{tr}(\rho \tilde{O}^{(d)})| &\leq |\mathrm{tr}(\rho O^{(d)}) - \mathrm{tr}(\rho O_\gamma^{(d)})| \\ &\quad + |\mathrm{tr}(\rho O_\gamma^{(d)}) - \mathrm{tr}(\rho \tilde{O}^{(d)})|. \end{aligned} \tag{J1}$$

Suppose that the noisy quantum experiment succeeds in estimating the noiseless expectation value to within root-mean-square error ε, for some noise rate γ. This implies that the root mean square of the first term on the rhs is less than ε. Meanwhile, the root mean square of the second term on the rhs is less than ε if we take ℓ as in Theorem 2 of the main text. Thus, the classical algorithm in Theorem 2 can compute expectation values within root-mean-square error 4ε in time as stated in Theorem 2. Imposing the lower bound $\chi(n)$ on the run-time of the classical algorithm, setting $d, \varepsilon^{-1} = \mathrm{poly}(n)$, and solving for the noise rate γ gives our desired result. ∎

APPENDIX K: NOISY QUANTUM CIRCUITS ON ANY HIGHLY MIXED INPUT STATE

As mentioned in the main text, our results in Theorems 1 and 2 immediately extend to circuits on any highly mixed input state. This follows because the ensemble composed of a single state ρ is low average whenever $||\rho||_\infty = c/2^n$ with $c = \mathcal{O}(1)$. In fact, we can prove a slightly stronger statement, in which this condition on the spectral norm is replaced by a condition on the Renyi-2 entropy, $S^{(2)}(\rho) = \log_2(1/||\rho||_2^2)$.

Corollary K1 (Classical algorithms for noisy quantum circuits on highly mixed states). Consider a noisy circuit $\mathcal{C}$ and a state ρ with Renyi-2 entropy $S^{(2)}(\rho) = n - \log_2(c)$. Assume the Pauli coefficients of ρ can be efficiently computed. There exists a classical algorithm to compute the state $\mathcal{C}\{\rho\}$ to within trace norm error $\varepsilon \cdot \sqrt{c}$ in time as in Theorems 1 and 2.

Proof of Corollary K1. The corollary follows by applying our Algorithms 1 and 2 to the density matrix ρ instead of an observable. This produces a classical approximation $\tilde{\rho}$ of the density matrix. The approximation has unit trace but is not guaranteed to be positive. We quantify the error, $\delta\rho = \tilde{\rho} - \mathcal{C}\{\rho\}$, in the approximation via the trace norm $||\delta\rho||_1$. We have

$$||\delta\rho||_1 \leq \sqrt{2^n} \cdot ||\delta\rho||_2 \leq \varepsilon \cdot \sqrt{2^n} \cdot ||\rho||_2 = \varepsilon \cdot \sqrt{c}, \tag{K1}$$

where the second inequality follows from the proofs of Theorems 1 and 2, and the final inequality uses the definition of the Renyi-2 entropy. ∎

APPENDIX L: NOISY QUANTUM CIRCUITS WITH SPATIAL DISORDER

We can also exchange the ensemble of input states in Theorems 1 and 2 with an ensemble of *quantum circuits* applied to any fixed input state. Intuitively, this follows by absorbing the preparation of the input state ensemble into the circuit itself. We illustrate this with a simple example. Consider estimating the expectation value of a Pauli operator Q over the ensemble of computational basis states. We can write each state $|\mathbf{s}\rangle$ as a Pauli operator applied to the zero state, $|\mathbf{s}\rangle = X_{\mathbf{s}}|0^n\rangle$, with $X_{\mathbf{s}} = \otimes_{i=0}^{n} X_i^{s_i}$. The expectation values are

$$\begin{aligned}\mathrm{tr}(\mathcal{C}\{|\mathbf{s}\rangle\langle\mathbf{s}|\}Q) &= \mathrm{tr}(\mathcal{C}\{X_{\mathbf{s}}|0^n\rangle\langle 0^n|X_{\mathbf{s}}^\dagger\}\cdot Q)\\ &= \mathrm{tr}(\mathcal{C}^{X_{\mathbf{s}}}\{|0^n\rangle\langle 0^n|\}\cdot X_{\mathbf{s}}^\dagger Q X_{\mathbf{s}})\\ &= (-1)^{a[X_{\mathbf{s}},Q]}\cdot\mathrm{tr}(\mathcal{C}^{X_{\mathbf{s}}}\{|0^n\rangle\langle 0^n|\}\cdot Q),\end{aligned} \tag{L1}$$

where $a[X_{\mathbf{s}}, Q]$ is 0 if $X_{\mathbf{s}}$ and Q commute and 1 if they anticommute. Here, we define the conjugated circuit $\mathcal{C}^P\{(\cdot)\} = P^\dagger\mathcal{C}\{P(\cdot)P^\dagger\}P$, which replaces each unitary in the original circuit with its conjugation by P, $U_t \to P^\dagger U_t P$. We note that multiplying by an overall phase does not change the estimation error. Thus, since our classical algorithm can estimate expectation values for the fixed circuit $\mathcal{C}$ applied to the ensemble of computational basis states, it can also estimate expectation values for the ensemble of circuits $\{\mathcal{C}^{X_{\mathbf{s}}}\}$ applied to the fixed input state $|0^n\rangle\langle 0^n|$.

We generalize this example in two ways. First, we can extend beyond Pauli observables by decomposing an arbitrary observable in the Pauli basis, $O = \sum_Q c_Q Q$. Since each Pauli can be estimated within error ε, the operator O can be estimated within error $\varepsilon\cdot||O||_{\mathrm{Pauli},1}$, where $||O||_{\mathrm{Pauli},1} = \sum_Q |c_Q|$ is the Pauli 1-norm. For most observables of interest, the Pauli 1-norm is upper bounded by the Frobenius and spectral norms. In particular, we have $||O||_{\mathrm{Pauli},1} \le \sqrt{m}||O||_F$ for observables that are sums of m Pauli operators, and $||O||_{\mathrm{Pauli},1} \le C(d,k)||O||_\infty$ [72] for k-local observables of degree d, where $C(d,k) = 3\sqrt{d}\exp(\Theta(k\log k))$.

Second, we can replace the specific circuit ensemble $\{\mathcal{C}^{X_{\mathbf{s}}}\}$ with any circuit ensemble that is invariant under conjugation by $X_{\mathbf{s}}$, i.e., $\{\mathcal{C}^{X_{\mathbf{s}}}\} = \{\mathcal{C}\}$ for all $\mathbf{s}$. Since the single-qubit Pauli-X operators generate the full set of $X_{\mathbf{s}}$, this is equivalent to the ensemble being invariant under conjugation by single-qubit Pauli-X operators. Hence, we refer to this as *spatial disorder* in the circuit ensemble. In the case that the ensemble is invariant under *any* single-qubit Pauli operator, we can extend this result to arbitrary input states. This follows since $\{P\rho P^\dagger|\mathcal{P}_n\}$ is a low-average ensemble with $c = 1$ for any fixed ρ when P is drawn from the n-qubit Pauli group $\mathcal{P}_n$. We capture both of these settings with the following definition.

Definition L1 (Spatial disorder in quantum circuit ensembles). We say that an ensemble of quantum circuits $\mathcal{E}_{\mathcal{C}} = \{\mathcal{C}\}$ has *spatial disorder* if the ensemble is invariant under conjugation by any single-qubit Pauli operator P:

$$\{\mathcal{C}^P\} = \{\mathcal{C}\}. \tag{L2}$$

As a weaker condition, we say that an ensemble $\mathcal{E}_{\mathcal{C}}$ has *spatial disorder on a quantum state* ρ if the same condition holds for a subset $\mathcal{E}_P$ of Pauli operators such that $(1/|\mathcal{E}_P|)\sum_P P\rho P^\dagger = \mathbb{1}/2^n$.

Our definition encompasses numerous circuits of interest for the quantum simulation of disordered spin models, including multiple recent experiments [73,126–130]. For example, a circuit composed of unitaries of the form

$$U_t = \exp\left(i\sum_i g_i X_i\right)\cdot\exp\left(i\sum_{(i,j)\in U_t} J_{ij}Z_iZ_j\right) \tag{L3}$$

has spatial disorder if g_i, J_{ij} are drawn at random from any distributions that are symmetric about zero. We can also allow, for instance, the g_i to be nonrandom if the input state is diagonal in the y or z basis, or the J_{ij} to be nonrandom for the x or y basis.

Leveraging our Theorems 1 and 2, we show that noisy quantum circuits with spatial disorder can be efficiently classically simulated.

Corollary L1 (Classical algorithm for noisy quantum circuits with spatial disorder). Consider a state ρ, an observable O, and an ensemble $\mathcal{E}_{\mathcal{C}} = \{\mathcal{C}\}$ of noisy quantum circuits with spatial disorder on ρ. Assume the Pauli coefficients of ρ and O can be efficiently computed. There exists a classical algorithm to compute the expectation values $\mathrm{tr}(\mathcal{C}\{\rho\}O)$ to within root-mean-square error $\varepsilon\cdot||O||_{\mathrm{Pauli},1}$ in time as in Theorems 1 and 2 of the main text.

Proof of Corollary L1. We use the same classical algorithms as in Theorems 1 and 2. The truncations in our classical algorithms respect additivity, so that our approximation for O under circuit $\mathcal{C}$ is equal to a sum of the corresponding approximations for each Pauli operator,

$$\tilde{O}_{\mathcal{C}}^{(d)} = \sum_Q c_Q\cdot\tilde{Q}_{\mathcal{C}}^{(d)}, \tag{L4}$$

where the coefficients c_Q are obtained from the Pauli decomposition of O before time evolution. If we consider instead the conjugated quantum circuit $\mathcal{C}^P$ for some Pauli operators P, we have

$$\tilde{O}_{\mathcal{C},P}^{(d)} = \sum_Q (-1)^{a[P,Q]}\cdot c_Q\cdot P^\dagger\tilde{Q}_{\mathcal{C}}^{(d)}P. \tag{L5}$$

Thus, the difference between the exact and approximate operator is

$$O_{\mathcal{C},P}^{(d)} - \tilde{O}_{\mathcal{C},P}^{(d)} = \sum_{Q} (-1)^{a[P,Q]} \cdot c_Q \cdot P^\dagger \left(Q_{\mathcal{C}}^{(d)} - \tilde{Q}_{\mathcal{C}}^{(d)} \right) P. \tag{L6}$$

By assumption, the ensemble $\mathcal{E}_{\mathcal{C}}$ is invariant under conjugation by any Pauli operator, since this follows from it being invariant under local Pauli conjugations. Thus, we can replace the average over $\mathcal{E}_{\mathcal{C}}$ with an identical average over the ensemble $\mathcal{E}_{\mathcal{C}^P} \equiv \{\mathcal{C}^P | \mathcal{C} \in \mathcal{E}_{\mathcal{C}}, P \in \mathcal{P}_n\}$, where $\mathcal{P}_n$ is the set of all n-qubit Pauli operators. The mean-square error over $\mathcal{E}_{\mathcal{C}^P}$ is

$$\begin{aligned}
\frac{1}{4^n |\mathcal{E}_{\mathcal{C}}|} \sum_{\mathcal{C},P} \mathrm{tr}\left(\rho \cdot (O_{\mathcal{C},P}^{(d)} - \tilde{O}_{\mathcal{C},P}^{(d)}) \right)^2 &\leq \frac{1}{4^n |\mathcal{E}_{\mathcal{C}}|} \sum_{\mathcal{C},P} \left(\sum\nolimits_Q |c_Q| \cdot |\mathrm{tr}\left(P\rho P^\dagger \cdot (Q_{\mathcal{C}}^{(d)} - \tilde{Q}_{\mathcal{C}}^{(d)}) \right)| \right)^2 \\
&\leq \frac{1}{4^n |\mathcal{E}_{\mathcal{C}}|} \sum_{\mathcal{C},P,Q,Q'} |c_Q||c_{Q'}||\mathrm{tr}(P\rho P^\dagger \cdot \delta Q_{\mathcal{C}})||\mathrm{tr}(P\rho P^\dagger \cdot \delta Q'_{\mathcal{C}})| \\
&\leq \frac{1}{|\mathcal{E}_{\mathcal{C}}|} \sum_{\mathcal{C},Q,Q'} |c_Q||c_{Q'}| \left(\frac{1}{4^n} \sum\nolimits_P \mathrm{tr}(P\rho P^\dagger \cdot \delta Q_{\mathcal{C}})^2 \right)^{1/2} \left(\frac{1}{4^n} \sum\nolimits_P \mathrm{tr}(P\rho P^\dagger \cdot \delta Q'_{\mathcal{C}})^2 \right)^{1/2},
\end{aligned} \tag{L7}$$

where the last line uses the Cauchy-Schwarz inequality, and we abbreviate $\delta Q_{\mathcal{C}} = Q_{\mathcal{C}}^{(d)} - \tilde{Q}_{\mathcal{C}}^{(d)}$. Now, note that the terms within parentheses correspond to mean-square error of Q and Q' under the *state ensemble* $\mathcal{E} = \{P\rho P^\dagger | P \in \mathcal{P}_n\}$. By assumption, this is a zero-average ensemble. Thus, by the results of Theorems 1 and 2, the term is upper bounded ε^2. We have

$$\frac{1}{4^n |\mathcal{E}_{\mathcal{C}}|} \sum_{\mathcal{C},P} \mathrm{tr}\left(\rho \cdot (O_{\mathcal{C},P}^{(d)} - \tilde{O}_{\mathcal{C},P}^{(d)}) \right)^2 \leq \frac{1}{|\mathcal{E}_{\mathcal{C}}|} \sum_{\mathcal{C}} \sum_{Q,} \sum_{Q'} |c_Q||c'_Q| \cdot \varepsilon^2 = \varepsilon^2 \left(\sum_Q |c_Q| \right)^2 = \varepsilon^2 \cdot ||O||_{\mathrm{Pauli},1}^2, \tag{L8}$$

as desired. ∎

[1] R. P. Feynman, *Simulating physics with computers*, in *Feynman and Computation* (CRC Press, 2018), pp. 133–153.

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

[3] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, *Quantum algorithms for quantum chemistry and quantum materials science*, Chem. Rev. **120**, 12685 (2020).

[4] A. M. Dalzell, S. McArdle, M. Berta, P. Bienias, C.-F. Chen, A. Gilyén, C. T. Hann, M. J. Kastoryano, E. T. Khabiboulline, A. Kubica *et al.*, *Quantum algorithms: A survey of applications and end-to-end complexities*, in *Quantum Algorithms* (Cambridge University Press, Cambridge, England, 2025).

[5] D. Aharonov and M. Ben-Or, *Polynomial simulations of decohered quantum computers*, in *Proceedings of 37th Conference on Foundations of Computer Science* (IEEE, 1996), pp. 46–55.

[6] D. Aharonov, *Quantum to classical phase transition in noisy quantum computers*, Phys. Rev. A **62**, 062311 (2000).

[7] A. W. Harrow and M. A. Nielsen, *Robustness of quantum gates in the presence of noise*, Phys. Rev. A **68**, 012308 (2003).

[8] J. Kempe, O. Regev, F. Unger, and R. De Wolf, *Upper bounds on the noise threshold for fault-tolerant quantum computing*, in *Proceedings of the International Colloquium on Automata, Languages, and Programming* (Springer, New York, 2008), pp. 845–856.

[9] P. W. Shor, *Fault-tolerant quantum computation*, in *Proceedings of 37th Conference on Foundations of Computer Science* (IEEE, 1996), pp. 56–65.

[10] D. Aharonov and M. Ben-Or, *Fault-tolerant quantum computation with constant error*, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* (1997), pp. 176–188.

[11] A. Y. Kitaev, *Quantum computations: Algorithms and error correction*, Russ. Math. Surv. **52**, 1191 (1997).

[12] E. Knill, R. Laflamme, and W. H. Zurek, *Resilient quantum computation*, Science **279**, 342 (1998).

[13] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, *Limitations of noisy reversible computation*, arXiv:quant-ph/9611028.

[14] M. Ben-Or, D. Gottesman, and A. Hassidim, *Quantum refrigerator*, arXiv:1301.1995.

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

[16] 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).
[17] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, *The complexity of NISQ*, arXiv:2210.07234.
[18] C. Huang, F. Zhang, M. Newman, J. Cai, X. Gao, Z. Tian, J. Wu, H. Xu, H. Yu, B. Yuan *et al.*, *Classical simulation of quantum supremacy circuits*, arXiv:2005.06787.
[19] Y. Liu, X. Liu, F. Li, H. Fu, Y. Yang, J. Song, P. Zhao, Z. Wang, D. Peng, H. Chen *et al.*, *Closing the "quantum supremacy" gap: Achieving real-time simulation of a random quantum circuit using a new sunway supercomputer*, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (2021), pp. 1–12.
[20] F. Pan and P. Zhang, *Simulating the Sycamore quantum supremacy circuits*, arXiv:2103.03074.
[21] B. Villalonga, D. Lyakh, S. Boixo, H. Neven, T. S. Humble, R. Biswas, E. G. Rieffel, A. Ho, and S. Mandrà, *Establishing the quantum supremacy frontier with a 281 Pflop/s simulation*, Quantum Sci. Technol. **5**, 034003 (2020).
[22] S. Cheng, C. Cao, C. Zhang, Y. Liu, S.-Y. Hou, P. Xu, and B. Zeng, *Simulating noisy quantum circuits with matrix product density operators*, Phys. Rev. Res. **3**, 023005 (2021).
[23] K. Noh, L. Jiang, and B. Fefferman, *Efficient classical simulation of noisy random quantum circuits in one dimension*, Quantum **4**, 318 (2020).
[24] Z. Chen, Y. Bao, and S. Choi, *Optimized trajectory unraveling for classical simulation of noisy quantum dynamics*, Phys. Rev. Lett. **133**, 230403 (2024).
[25] Z. Cheng and M. Ippoliti, *Efficient sampling of noisy shallow circuits via monitored unraveling*, PRX Quantum **4**, 040326 (2023).
[26] M. J. Bremner, A. Montanaro, and D. J. Shepherd, *Achieving quantum supremacy with sparse and noisy commuting quantum computations*, Quantum **1**, 8 (2017).
[27] J. Rajakumar, J. D. Watson, and Y.-K. Liu, *Polynomial-time classical simulation of noisy IQP circuits with constant depth*, in *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, arXiv:2403.14607.
[28] X. Gao and L. Duan, *Efficient classical simulation of noisy quantum computation*, arXiv:1810.03176.
[29] B. Barak, C.-N. Chou, and X. Gao, *Spoofing linear cross-entropy benchmarking in shallow quantum circuits*, arXiv:2005.02421.
[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 (2021).
[31] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, *A polynomial-time classical algorithm for noisy random circuit sampling*, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (2023), pp. 945–957.
[32] M.-H. Yung and X. Gao, *Can chaotic quantum circuits maintain quantum supremacy under noise?*, arXiv:1706.08913.
[33] E. Fontana, M. S. Rudolph, R. Duncan, I. Rungger, and C. Cîrstoiu, *Classical simulations of noisy variational quantum circuits*, arXiv:2306.05400.
[34] Y. Shao, F. Wei, S. Cheng, and Z. Liu, *Simulating quantum mean values in noisy variational quantum algorithms: A polynomial-scale approach*, Phys. Rev. Lett. **133**, 120603 (2024).
[35] A. Bouland, B. Fefferman, Z. Landau, and Y. Liu, *Noise and the frontier of quantum supremacy*, in *Proceedings of the 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022), pp. 1308–1317.
[36] 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.
[37] R. Movassagh and J. Schenker, *Theory of ergodic quantum processes*, Phys. Rev. X **11**, 041001 (2021).
[38] K. Fujii, *Noise threshold of quantum supremacy*, arXiv:1610.03632.
[39] G. Kalai and G. Kindler, *Gaussian noise sensitivity and bosonsampling*, arXiv:1409.3093.
[40] R. García-Patrón, J. J. Renema, and V. Shchesnovich, *Simulating boson sampling in lossy architectures*, Quantum **3**, 169 (2019).
[41] H. Qi, D. J. Brod, N. Quesada, and R. García-Patrón, *Regimes of classical simulability for noisy Gaussian boson sampling*, Phys. Rev. Lett. **124**, 100502 (2020).
[42] C. Oh, L. Jiang, and B. Fefferman, *On classical simulation algorithms for noisy boson sampling*, arXiv:2301.11532.
[43] C. Oh, M. Liu, Y. Alexeev, B. Fefferman, and L. Jiang, *Tensor network algorithm for simulating experimental Gaussian boson sampling*, Nat. Phys. **20**, 1461 (2024).
[44] D. Hangleiter and M. J. Gullans, *Bell sampling from quantum circuits*, Phys. Rev. Lett. **133**, 020601 (2024).
[45] A. Tanggara, M. Gu, and K. Bharti, *Classically spoofing system linear cross entropy score benchmarking*, arXiv:2405.00789.
[46] I. Ermakov, O. Lychkovskiy, and T. Byrnes, *Unified framework for efficiently computable quantum circuits*, arXiv:2401.08187.
[47] I. Kuprov, N. Wagner-Rundell, and P. Hore, *Polynomially scaling spin dynamics simulation algorithm based on adaptive state-space restriction*, J. Magn. Reson. **189**, 241 (2007).
[48] A. Karabanov, I. Kuprov, G. Charnock, A. van der Drift, L. J. Edwards, and W. Köckenberger, *On the accuracy of the state space restriction approximation for spin dynamics simulations*, J. Chem. Phys. **135** (2011).
[49] T. Rakovszky, C. W. von Keyserlingk, and F. Pollmann, *Dissipation-assisted operator evolution method for capturing hydrodynamic transport*, Phys. Rev. B **105**, 075131 (2022).
[50] Y. Yoo, C. D. White, and B. Swingle, *Open-system spin transport and operator weight dissipation in spin chains*, Phys. Rev. B **107**, 115118 (2023).
[51] T. Vovk and H. Pichler, *Entanglement-optimal trajectories of many-body quantum Markov processes*, Phys. Rev. Lett. **128**, 243601 (2022).
[52] A. M. Dalzell, N. Hunter-Jones, and F. G. Brandão, *Random quantum circuits transform local noise into global white noise*, Commun. Math. Phys. **405**, 78 (2024).
[53] L. Paletta, A. Leverrier, A. Sarlette, M. Mirrahimi, and C. Vuillot, *Robust sparse IQP sampling in constant depth*, Quantum **8**, 1337 (2024).

[54] C. D. White, M. Zaletel, R. S. K. Mong, and G. Refael, *Quantum dynamics of thermalizing systems*, Phys. Rev. B **97**, 035127 (2018).
[55] B. Ye, F. Machado, C. D. White, R. S. K. Mong, and N. Y. Yao, *Emergent hydrodynamics in nonequilibrium quantum systems*, Phys. Rev. Lett. **125**, 030601 (2020).
[56] C. Von Keyserlingk, F. Pollmann, and T. Rakovszky, *Operator backflow and the classical simulation of quantum transport*, Phys. Rev. B **105**, 245101(R) (2022).
[57] C. D. White, *Effective dissipation rate in a Liouvillian-graph picture of high-temperature quantum hydrodynamics*, Phys. Rev. B **107**, 094311 (2023).
[58] T. Klein Kvorning, L. Herviou, and J. H. Bardarson, *Time-evolution of local information: Thermalization dynamics of local observables*, SciPost Phys. **13**, 080 (2022).
[59] C. Artiaco, C. Fleckenstein, D. Aceituno Chávez, T. K. Kvorning, and J. H. Bardarson, *Efficient large-scale many-body quantum dynamics via local-information time evolution*, PRX Quantum **5**, 020352 (2024).
[60] K. Temme, S. Bravyi, and J. M. Gambetta, *Error mitigation for short-depth quantum circuits*, Phys. Rev. Lett. **119**, 180509 (2017).
[61] Y. Li and S. C. Benjamin, *Efficient variational quantum simulator incorporating active error minimization*, Phys. Rev. X **7**, 021050 (2017).
[62] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, *Quantum error mitigation*, Rev. Mod. Phys. **95**, 045005 (2023).
[63] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, *Exponentially tighter bounds on limitations of quantum error mitigation*, Nat. Phys. **20**, 1648 (2024).
[64] R. Takagi, S. Endo, S. Minagawa, and M. Gu, *Fundamental limits of quantum error mitigation*, npj Quantum Inf. **8**, 114 (2022).
[65] R. Takagi, H. Tajima, and M. Gu, *Universal sampling lower bounds for quantum error mitigation*, Phys. Rev. Lett. **131**, 210602 (2023).
[66] K. Tsubouchi, T. Sagawa, and N. Yoshioka, *Universal cost bound of quantum error mitigation based on quantum estimation theory*, Phys. Rev. Lett. **131**, 210601 (2023).
[67] B. Fefferman, S. Ghosh, M. Gullans, K. Kuroiwa, and K. Sharma, *Effect of non-unital noise on random circuit sampling*, arXiv:2306.16659.
[68] As a consequence, our algorithms for uniform noise can be thought of as working explicitly in the regime $d = \mathcal{O}(\gamma^{-1}\log(n/\varepsilon))$. After this depth, one can approximate any expectation value with its value in the maximally mixed state [13].
[69] In particular, threshold theorems typically assume two ingredients [9–12]: midcircuit measurement and feedforward operations, and the ability to insert "fresh qubits" unaffected by idle errors. In effect, both noise models allow midcircuit measurement and feedforward operations, since one can implement the feedforward operations via controlled quantum gates. (For a fixed input state, this can be made fault tolerant by encoding the state of the "measured" qubit in a classical error correcting code at the layer at which it is "measured.") Crucially however, the gate-based noise model allows the insertion of fresh qubits, while the uniform noise model does not.
[70] See Ref. [31], where this is referred to as "orthogonality." This requirement is satisfied in random circuits, where averaged properties of operator time evolution are governed by an incoherent classical Markov process, but not more generally.
[71] M. C. Caro, H.-Y. Huang, N. Ezzell, J. Gibbs, A. T. Sornborger, L. Cincio, P. J. Coles, and Z. Holmes, *Out-of-distribution generalization for learning quantum dynamics*, Nat. Commun. **14**, 3751 (2023).
[72] H.-Y. Huang, S. Chen, and J. Preskill, *Learning to predict arbitrary quantum processes*, arXiv:2210.14894.
[73] X. Mi, M. Ippoliti, C. Quintana, A. Greene, Z. Chen, J. Gross, F. Arute, K. Arya, J. Atalaya, R. Babbush *et al.*, *Time-crystalline eigenstate order on a quantum processor*, Nature (London) **601**, 531 (2022).
[74] M. Dupont and J. E. Moore, *Universal spin dynamics in infinite-temperature one-dimensional quantum magnets*, Phys. Rev. B **101**, 121106 (2020).
[75] E. Rosenberg, T. Andersen, R. Samajdar, A. Petukhov, J. Hoke, D. Abanin, A. Bengtsson, I. Drozdov, C. Erickson, P. Klimov *et al.*, *Dynamics of magnetization at infinite temperature in a Heisenberg spin chain*, Science **384**, 48 (2024).
[76] At this halting step, we also group together all weights above the threshold in order to achieve a tighter bound on the error. See Appendix C for full details.
[77] We note that the maximum nontrivial depth for quantum circuits with random input states, $d = \mathcal{O}(\gamma^{-1}\log(1/\varepsilon))$, is slightly shorter than the maximum depth for quantum circuits with a fixed input state, $d = \mathcal{O}(\gamma^{-1}\log(n/\varepsilon))$ [13]. See Appendix C for further details.
[78] D. A. Roberts, D. Stanford, and A. Streicher, *Operator growth in the SKY model*, J. High Energy Phys. 01 (2018) 122.
[79] A. R. Brown, H. Gharibyan, S. Leichenauer, H. W. Lin, S. Nezami, G. Salton, L. Susskind, B. Swingle, and M. Walter, *Quantum gravity in the lab. I. Teleportation by size and traversable wormholes*, PRX Quantum **4**, 010320 (2023).
[80] T. Schuster, B. Kobrin, P. Gao, I. Cong, E. T. Khabiboulline, N. M. Linke, M. D. Lukin, C. Monroe, B. Yoshida, and N. Y. Yao, *Many-body quantum teleportation via operator spreading in the traversable wormhole protocol*, Phys. Rev. X **12**, 031013 (2022).
[81] T. Schuster and N. Y. Yao, *Operator growth in open quantum systems*, Phys. Rev. Lett. **131**, 160402 (2023).
[82] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, *Low-cost error mitigation by symmetry verification*, Phys. Rev. A **98**, 062339 (2018).
[83] S. McArdle, X. Yuan, and S. Benjamin, *Error-mitigated digital quantum simulation*, Phys. Rev. Lett. **122**, 180501 (2019).
[84] N. C. Rubin, R. Babbush, and J. McClean, *Application of fermionic marginal constraints to hybrid quantum algorithms*, New J. Phys. **20**, 053020 (2018).
[85] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, *Error mitigation with Clifford quantum-circuit data*, Quantum **5**, 592 (2021).
[86] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R.

McClean, *Virtual distillation for quantum error mitigation*, Phys. Rev. X **11**, 041036 (2021).
[87] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, *Error mitigation via verified phase estimation*, PRX Quantum **2**, 020317 (2021).
[88] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala, *Scalable error mitigation for noisy quantum circuits produces competitive expectation values*, Nat. Phys. **19**, 752 (2023).
[89] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme *et al.*, *Evidence for the utility of quantum computing before fault tolerance*, Nature (London) **618**, 500 (2023).
[90] T. L. Scholten, C. J. Williams, D. Moody, M. Mosca, W. Hurley, W. J. Zeng, M. Troyer, J. M. Gambetta *et al.*, *Assessing the benefits and risks of quantum computers*, arXiv:2401.16317.
[91] J. A. Jones, *Quantum computing with NMR*, Prog. NMR Spectrosc. **59**, 91 (2011).
[92] M. W. Doherty, N. B. Manson, P. Delaney, F. Jelezko, J. Wrachtrup, and L. C. Hollenberg, *The nitrogen-vacancy colour centre in diamond*, Phys. Rep. **528**, 1 (2013).
[93] C. Zu, F. Machado, B. Ye, S. Choi, B. Kobrin, T. Mittiga, S. Hsieh, P. Bhattacharyya, M. Markham, D. Twitchen *et al.*, *Emergent hydrodynamics in a strongly interacting dipolar spin ensemble*, Nature (London) **597**, 45 (2021).
[94] E. Knill and R. Laflamme, *Power of one bit of quantum information*, Phys. Rev. Lett. **81**, 5672 (1998).
[95] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandão, D. A. Buell *et al.*, *Quantum supremacy using a programmable superconducting processor*, Nature (London) **574**, 505 (2019).
[96] D. Hangleiter, J. Bermejo-Vega, M. Schwarz, and J. Eisert, *Anticoncentration theorems for schemes showing a quantum speedup*, Quantum **2**, 65 (2018).
[97] A. M. Dalzell, N. Hunter-Jones, and F. G. S. L. Brandão, *Random quantum circuits anticoncentrate in log depth*, PRX Quantum **3**, 010333 (2022).
[98] T. Schuster, J. Haferkamp, and H.-Y. Huang, *Random unitaries in extremely low depth*, Science **389**, 92 (2025).
[99] A. Sauliere, B. Magni, G. Lami, X. Turkeshi, and J. De Nardis, *Universality in the anticoncentration of chaotic quantum circuits*, arXiv:2503.00119.
[100] L. Grevink, J. Haferkamp, M. Heinrich, J. Helsen, M. Hinsche, T. Schuster, and Z. Zimborás, *Will it glue? On short-depth designs beyond the unitary group*, arXiv:2506.23925.
[101] L. Cui, T. Schuster, F. Brandao, and H.-Y. Huang, *Unitary designs in nearly optimal depth*, arXiv:2507.06216.
[102] R. Movassagh, *The hardness of random quantum circuits*, Nat. Phys. **19**, 1719 (2023).
[103] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, *On the complexity and verification of quantum random circuit sampling*, Nat. Phys. **15**, 159 (2019).
[104] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan *et al.*, *Strong quantum computational advantage using a superconducting quantum processor*, Phys. Rev. Lett. **127**, 180501 (2021).
[105] 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).
[106] A. Morvan, B. Villalonga, X. Mi, S. Mandra, A. Bengtsson, P. Klimov, Z. Chen, S. Hong, C. Erickson, I. Drozdov *et al.*, *Phase transition in random circuit sampling*, Nature (London) **634**, 328 (2024).
[107] J. Tindall, M. Fishman, M. Stoudenmire, and D. Sels, *Efficient tensor network simulation of IBM's kicked Ising experiment*, PRX Quantum **5**, 010308 (2024).
[108] T. Begušić, J. Gray, and G. K. Chan, *Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance*, Sci. Adv. **10**, eadk4321 (2024).
[109] K. Kechedzhi, S. Isakov, S. Mandrà, B. Villalonga, X. Mi, S. Boixo, and V. Smelyanskiy, *Effective quantum volume, fidelity and computational cost of noisy quantum processing experiments*, Future Gener. Comput. Syst. **153**, 431 (2024).
[110] S. Anand, K. Temme, A. Kandala, and M. Zaletel, *Classical benchmarking of zero noise extrapolation beyond the exactly-verifiable regime*, arXiv:2306.17839.
[111] M. S. Rudolph, E. Fontana, Z. Holmes, and L. Cincio, *Classical surrogate simulation of quantum systems with LOWESA*, arXiv:2308.09109.
[112] S. Patra, S. S. Jahromi, S. Singh, and R. Orús, *Efficient tensor network simulation of IBM's largest quantum processors*, Phys. Rev. Res. **6**, 013326 (2024).
[113] A similar but less precise connection was made in Ref. [109].
[114] A. Nahum, S. Roy, S. Vijay, and T. Zhou, *Real-time correlators in chaotic quantum many-body systems*, Phys. Rev. B **106**, 224310 (2022).
[115] T. Yoshimura, S. J. Garratt, and J. Chalker, *Operator dynamics in Floquet many-body systems*, Phys. Rev. B **111**, 094316 (2025).
[116] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, *Classically estimating observables of noiseless quantum circuits*, arXiv:2409.01706.
[117] A. Goussev, R. A. Jalabert, H. M. Pastawski, and D. Wisniacki, *Loschmidt echo*, Scholarpedia **7**, 11687 (2012).
[118] For example, for the entanglement entropy and OTOC, one can consider a deep Clifford circuit. This generates high entanglement despite being classically simulable via the Gottesman-Knill theorem. For the magic, one can consider a 1D circuit with many T gates but low circuit depth, or a Gaussian fermionic state. Both have an extensive number of T gates, while remaining simulable via matrix product state and match gate algorithms, respectively. For the operator entanglement entropy (and all aforementioned metrics as well), one can consider a generic highly ergodic many-body system. Time-evolved operators in such systems typically feature a linear growth in operator entanglement entropy over time, despite the operators' expectation values often being efficiently simulated via Pauli truncation methods [46,49–51,54–59].
[119] S. Bravyi, D. Gosset, and Y. Liu, *Classical simulation of peaked shallow quantum circuits*, arXiv:2309.08405.

[120] S. Bravyi, D. Gosset, and R. Movassagh, *Classical algorithms for quantum mean values*, Nat. Phys. **17**, 337 (2021).

[121] If the individual coefficients can be computed in time τ instead of time $\mathcal{O}(1)$, then the time to read in all coefficients will be $\mathcal{O}(\tau \cdot D_\ell)$ instead of $\mathcal{O}(D_\ell)$. Our final run-time will be unchanged as long as $\tau = \mathcal{O}(d \cdot \ell!)$, and will be quasipolynomial in n as long as τ is quasipolynomial as well.

[122] To suppress the effect of read-out noise, one can consider implementing classical error correction prior to measurement (e.g., via a repetition code on each qubit). In this case, the output distribution does not anticoncentrate and so our theorem does not apply.

[123] M. Raginsky, *Strictly contractive quantum channels and physically realizable quantum computers*, Phys. Rev. A **65**, 032306 (2002).

[124] L. Viola, E. Knill, and S. Lloyd, *Dynamical decoupling of open quantum systems*, Phys. Rev. Lett. **82**, 2417 (1999).

[125] K. Khodjasteh and D. A. Lidar, *Fault-tolerant quantum dynamical decoupling*, Phys. Rev. Lett. **95**, 180501 (2005).

[126] S. Choi, J. Choi, R. Landig, G. Kucsko, H. Zhou, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, V. Khemani *et al.*, *Observation of discrete time-crystalline order in a disordered dipolar many-body system*, Nature (London) **543**, 221 (2017).

[127] J. Randall, C. Bradley, F. van der Gronden, A. Galicia, M. Abobeih, M. Markham, D. Twitchen, F. Machado, N. Yao, and T. Taminiau, *Many-body–localized discrete time crystal with a programmable spin-based quantum simulator*, Science **374**, 1474 (2021).

[128] R. Harris, Y. Sato, A. J. Berkley, M. Reis, F. Altomare, M. Amin, K. Boothby, P. Bunyk, C. Deng, C. Enderud *et al.*, *Phase transitions in a programmable quantum spin glass simulator*, Science **361**, 162 (2018).

[129] A. D. King, J. Raymond, T. Lanting, R. Harris, A. Zucca, F. Altomare, A. J. Berkley, K. Boothby, S. Ejtemaee, C. Enderud *et al.*, *Quantum critical dynamics in a 5,000-qubit programmable spin glass*, Nature (London) **617**, 61 (2023).

[130] A. D. King, A. Nocera, M. M. Rams, J. Dziarmaga, R. Wiersema, W. Bernoudy, J. Raymond, N. Kaushal, N. Heinsdorf, R. Harris *et al.*, *Computational supremacy in quantum simulation*, arXiv:2403.00910.