

# Simulating Quantum Circuits with Arbitrary Local Noise Using Pauli Propagation

Armando Angrisani<sup>1,2,\*</sup> Antonio A. Mele<sup>3,4,†</sup> Manuel S. Rudolph<sup>1,2</sup> M. Cerezo<sup>1,5</sup>  
and Zoë Holmes<sup>1,2</sup>

<sup>1</sup>*Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland*

<sup>2</sup>*Centre for Quantum Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

<sup>3</sup>*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Berlin 14195, Germany*

<sup>4</sup>*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

<sup>5</sup>*Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*



(Received 14 May 2025; revised 23 December 2025; accepted 17 March 2026; published 22 April 2026)

We present a polynomial-time classical algorithm for estimating expectation values of arbitrary observables on typical quantum circuits under any incoherent local noise, including non-unital or dephasing. Although previous research demonstrated that some carefully designed quantum circuits affected by non-unital noise cannot be efficiently simulated, we show that this does not apply to average-case circuits, as these can be efficiently simulated using Pauli-path methods. Specifically, we prove that, with high probability over the circuit gates' choice, Pauli propagation algorithms with tailored truncation strategies achieve an inversely polynomially small simulation error. This result holds for arbitrary circuit topologies and for any local noise, under the assumption that the distribution of each circuit layer is invariant under single-qubit random gates. Under the same minimal assumptions, we also prove that most noisy circuits can be truncated to an effective logarithmic depth for the task of estimating expectation values of observables, thus generalizing prior results to a significantly broader class of circuit ensembles. We further numerically validate our algorithm with simulations on a  $6 \times 6$  lattice of qubits under the effects of amplitude damping and dephasing noise, as well as real-time dynamics on an  $11 \times 11$  lattice of qubits affected by amplitude damping.

DOI: [10.1103/fb28-wlv2](https://doi.org/10.1103/fb28-wlv2)

## I. INTRODUCTION

Recent years have seen a healthy back-and-forth between experimental teams attempting to implement nonclassically simulable quantum algorithms on quantum hardware and theorists attempting to demonstrate these experiments can be efficiently classically simulated [1–12]. Central to these discussions is the role of hardware noise. In particular, while there exist quantum circuits that are widely believed to be computationally hard to simulate classically, these cannot be implemented exactly on currently available devices. Rather, in the current pre-fault-tolerant era, all circuits are invariably subject

to hardware noise, and this issue forces researchers to make delicate compromises. On the one hand, circuits subject to noise are often simpler to classically simulate [2,4,13–16]. On the other hand, the search for circuits that are less affected by noise can push the user toward ones that are classically tractable [5–12]. Hence, understanding the subtle connections between noise and efficient classical simulation is key to the search for quantum advantage.

Given that hardware errors stem from complex, undesired interactions with surrounding environments, a variety of mathematical models have been developed to investigate these effects [17–23]. Nevertheless, the vast majority of works focused on noise models within the *depolarizing* class. This class includes noise channels that always increase the entropy of the system. Thus, if no measurements are performed and no fresh auxiliary qubits are supplied during the computation, the output of a quantum circuit composed of alternated layers of such noise and arbitrary unitary quantum gates is driven exponentially fast toward the maximally mixed state. This makes

\*Contact author: [armando.angrisani@epfl.ch](mailto:armando.angrisani@epfl.ch)

†Contact author: [a.mele@fu-berlin.de](mailto:a.mele@fu-berlin.de)

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

any nontrivial computation unfeasible beyond logarithmic depth, regardless of which gates are applied [14,24–27].

Another key feature of depolarizing noise models is that they have a strikingly simple representation in the basis of Pauli operators. Namely, under the action of *local* depolarizing noise, the contributions of global Pauli operators are suppressed exponentially more than those of local Pauli operators. This insight has been widely exploited in several recent works leveraging classical simulation techniques based on Pauli-path summations, which are discrete Feynman path integrals written in the Pauli basis [6,13,15,16,28–31]. While its remarkable simplicity makes the depolarizing class an appealing model for analysis, it fails to account for many dissipation phenomena that routinely arise in quantum computation experiments [1,32–34].

Alongside the depolarizing class, realistic quantum noise can also be described by the *dephasing* class and the *non-unital* class, which both present qualitatively different features, as they do not necessarily increase the entropy of the quantum system. Specifically, non-unital noise can decrease the entropy of the system, potentially driving it toward a pure state. Moreover, as demonstrated in Ref. [35], under some circumstances one can take advantage of non-unital noise by exploiting it as a form of quantum refrigerator, allowing for fault-tolerant quantum computation for exponential depth without fresh auxiliary qubits or mid-circuit measurements. On the other hand, dephasing noise never decreases the entropy of the systems, but it leaves the entropy of computational basis states unchanged. In both cases, investigating the properties of these noisy quantum circuits is arguably more challenging than analyzing those with depolarizing noise, as standard entropy accumulation tools [14,24] cannot be used.

As the behavior of *carefully designed* quantum circuits drastically differs under different sources of noise, it is natural to investigate the effect of arbitrary noise on *typical* quantum circuits, i.e., circuits randomly sampled from suitable circuit distributions. On one hand, random circuits affected by non-unital noise exhibit qualitatively

different behavior from circuit with solely depolarizing noise, as shown in the context of random circuit sampling [36] and variational quantum algorithms, where non-unital noise has been demonstrated to induce absence of barren plateaus [37,38], in contrast to the depolarizing noise scenario [26]. On the other hand, noisy random circuits exhibit an effective logarithmic depth [37], echoing the logarithmic-depth barrier [14,24] associated with depolarizing noise.

In this work we leverage the insight that circuit randomness and generic local noise behaves analogously to depolarizing noise to develop an efficient classical simulation algorithm for simulating typical quantum circuits in the presence of arbitrary local noise. Concretely, we show that expectation values of observables can be estimated with inversely polynomial precision in polynomial time on *most* circuits and for *any* incoherent local noise, possibly non-unital or dephasing, provided that each circuit layer is sampled from a distribution invariant under single-qubit random gates. Our analysis combines the Pauli-path simulation algorithms also employed in Refs. [13,16,30] with the normal form representation of local noise, previously employed in Refs. [18,35,37]. Our results, sketched in Fig. 1, improve the state-of-the-art in the following ways:

- (1) **General noise model.** We make minimal assumptions about the noise model, considering circuit layers interspersed with single-qubit noise channels acting on all qubits. Rather than adopting a specific noise channel, as in Refs. [13,16], we only require these single-qubit channels to be non-unital with a constant noise rate. Our results also encompass dynamic non-unitary operations such as the ones considered in Ref. [40], provided that they are applied with a constant rate.
- (2) **Polynomial runtime for any geometry.** While Ref. [37] previously analyzed random circuits with arbitrary local noise, it demonstrated efficient classical simulation for arbitrary circuit connectivity, but

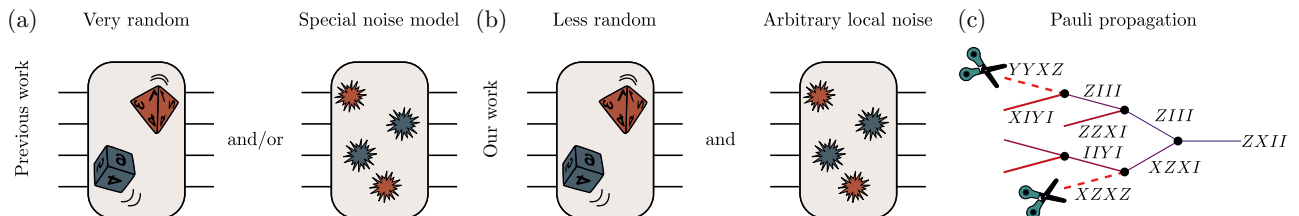


FIG. 1. Schematic summary of our main results. (a) Prior work has provided guarantees for simulating a quantum circuit either for highly random quantum circuits or assuming a special noise model. (b) Here we show that the combination of a very general noise model (arbitrary local incoherent noise) with some randomness allows quantum circuits to be efficiently simulated. (c) Our algorithm combines Pauli propagation with a truncation scheme based on the so-called *path weight*, which is the *total cumulative weight* of Pauli operators on a given branch [13,16,30] in contrast to *current* Pauli weight [6,16,39].

crucially only for constant precision. For inversely polynomial precision—which, in many cases, corresponds to the desired accuracy (e.g., for estimating extensive quantities such as a molecule’s ground-state energy [41])—polynomial-time simulation was guaranteed only under the restriction of one-dimensional connectivity. In contrast, our results establish that any inversely polynomial error can be efficiently achieved for all circuit topologies, thereby matching the accuracy required in many physically motivated scenarios.

- (3) **Less random gate-sets.** Our assumptions about the circuit ensemble are also minimal, enabling efficient classical simulation for a broader class of circuits than those considered in previous works, such as Refs. [13,37]. Specifically, we assume that each layer is independently sampled from a distribution invariant under different families of random single-qubit gates. In particular, for non-unital or depolarizing-like noise, we assume that the random single-qubit gates are sampled from local unitary 1-designs. For dephasing noise, we assume the gates satisfy an approximate local unitary 2-design property (e.g., they implement two random Pauli rotations occurring in orthogonal directions).

Under the same minimal assumptions on circuit structure and noise type, we demonstrate that for the task of estimating expectation values of observables, any incoherent noise reduces the circuit to an effective logarithmic depth. That is, any gates outside of this logarithmic depth window do not have a meaningful effect on expectation value estimates and thus can be ignored while classically simulating the circuit. This result significantly generalizes the main finding of Ref. [37], which applies specifically to noisy circuits composed of two-qubit gates sampled from local unitary 2-designs.

Finally, we complement our theoretical investigation with numerical simulations of a periodic  $6 \times 6$  lattice evolving under a transverse-field Ising variational ansatz affected by amplitude-damping noise and dephasing noise. For our error estimates, we employ a Monte Carlo certification approach that can be used beyond exactly simulable regimes. The resulting analysis shows that Pauli propagation equipped with path-weight truncation substantially outperforms our analytic accuracy guarantees. Finally, we provide a large-scale example of a dynamical simulation of a noisy rotated transverse-field Ising model on an  $11 \times 11$  square lattice with 121 qubits subject to amplitude-damping noise. These structured circuits do not resemble typical random circuits and thus break the assumptions of our theoretical analysis. Despite this fact, we find that Pauli propagation remains reasonably efficient, thus showing its promise in regimes beyond our theoretical guarantees.

## II. FRAMEWORK

**Scope.** Before defining the circuit ensemble and noise models below, we clarify the computational task considered here. Our algorithms estimate  $\text{Tr}[OC(\rho)]$  for a given observable  $O$ , noisy circuit  $\mathcal{C}$  and initial state  $\rho$ , to prescribed additive accuracy  $\varepsilon$  and with high probability over the circuit randomness. This is an *expectation-value estimation* guarantee: we do not provide an efficient procedure to sample from the full output distribution of  $\mathcal{C}(\rho)$ . We pose efficient output-distribution simulation as an open problem in Sec. VI.

**General local noise model.** In general, any single-qubit channel  $\mathcal{N}$  can be written in the following *normal form* [18,35,37]:

$$\mathcal{N}(\cdot) = U\mathcal{N}'(V(\cdot)V^\dagger)U^\dagger, \quad (1)$$

where  $U, V$  are arbitrary single-qubit unitaries. The action of  $\mathcal{N}'$  on the single-qubit Pauli matrices is given by

$$\mathcal{N}'(I) = I + t_X X + t_Y Y + t_Z Z, \quad (2)$$

$$\mathcal{N}'(X) = D_X X, \quad (3)$$

$$\mathcal{N}'(Y) = D_Y Y, \quad (4)$$

$$\mathcal{N}'(Z) = D_Z Z, \quad (5)$$

where  $\mathbf{D} = (D_X, D_Y, D_Z) \in [-1, 1]^3$  and  $\mathbf{t} = (t_X, t_Y, t_Z) \in [-1, 1]^3$  are two vectors, which we refer to as *normal form parameters*. We say that the channel  $\mathcal{N}$  has constant noise rate if  $\frac{1}{3} \|\mathbf{D}\|_2^2$  is a constant strictly smaller than one. Moreover, we say that  $\mathcal{N}$  is *unital* if  $\mathcal{N}(I) = I$  and *non-unital* otherwise.

There are in broad terms three different families of noise models, which we schematically sketch in Fig. 2.

- (1) **Depolarizing-like noise:** This corresponds to unital channels satisfying  $\|\mathbf{D}\|_\infty < 1$ , i.e., to noise channels that drive *any* input state toward the maximally mixed state. The most prominent example of depolarizing-like noise is the depolarizing channel, where  $D_X = D_Y = D_Z$ .
- (2) **Dephasing-like noise:** This corresponds to unital channels satisfying  $|D_P| = 1$  for exactly one  $P \in \{X, Y, Z\}$ . Such channels leave invariant the quantum states that lie along a specific diameter of the Bloch sphere, while driving all other states toward such diameter.
- (3) **Non-unital noise:** Finally, non-unital noise channels are those that do not preserve the identity, i.e.,  $\mathcal{N}(I) \neq I$ . Repeatedly applying a non-unital channel leads the quantum state toward a *fixed point* different than maximally mixed state. A common example of non-unital noise is the amplitude-damping channel, characterized by

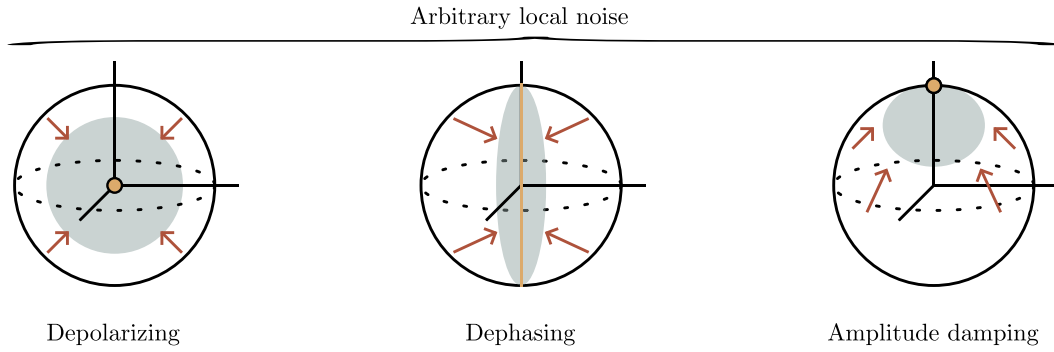


FIG. 2. Overview of noise models. Here we provide a geometric sketch of three commonly considered families of local noise: (a) Depolarizing noise, which drives any input state toward the maximally mixed state, (b) Dephasing-like noise, which drives states to  $z$ -axis connecting poles of the Bloch sphere, and (c) Non-unital noise, which drives states to a given fixed point on the Bloch sphere. While prior work considered special cases of these families our analysis applies to any local noise channel with a constant noise rate.

$\mathbf{D} = (\sqrt{1-\gamma}, \sqrt{1-\gamma}, 1-\gamma)$  and  $\mathbf{t} = (0, 0, \gamma)$  for some  $\gamma \in (0, 1]$ . The fixed point of the amplitude-damping channel is the computational zero state.

While our results encompass any local noise with a constant noise rate, i.e.,  $\|\mathbf{D}\|_2 \in \Theta(1)$ ,  $\frac{1}{3}\|\mathbf{D}\|_2^2 < 1$ , previous works have considered special cases. For instance, Refs. [13,15,16,30] consider noise within the depolarizing class. Additionally, Ref. [16] investigates randomized non-unital noise, which models spontaneous emissions occurring in random directions. This noise is characterized by normal form parameters given by  $\mathbf{D} = (1-\gamma/2, 1-\gamma/2, 1-\gamma)$  and  $\mathbf{t} = (0, 0, r\gamma)$ , where  $r$  is randomly set to either 1 or  $-1$  with equal probability.

**Random local noise circuit models.** In this work, we focus on the task of estimating expectation values such as  $\text{Tr}[O\mathcal{C}(\rho)]$ . Here,  $\rho$  is an arbitrary initial  $n$ -qubit quantum state,  $O$  is a bounded Hermitian observable such that  $\|O\| \leq 1$  (where  $\|\cdot\|$  denotes the spectral norm), and  $\mathcal{C}$  is an  $L$ -layered noisy random circuit of the form

$$\mathcal{C} := \mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_L \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_{L-1} \circ \dots \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_1. \quad (6)$$

Above, the final layer  $\mathcal{V}^{\text{single}}$  consists of single-qubit gates, and the layers  $\mathcal{U}_j$  consist of non-overlapping gates, each acting on  $\mathcal{O}(1)$  qubits. Then, we assume for simplicity that each  $\mathcal{N}$  is an arbitrary single-qubit noise channel as defined in Eq. (1). The extension to different local noise channels acting on different qubits is straightforward. Finally, we study the case when all random layers  $\mathcal{U}_1, \dots, \mathcal{U}_L$  and  $\mathcal{V}^{\text{single}}$  are sampled independently from some distributions invariant (up to their second moment) under right-multiplication of random single-qubit gates. In particular, we focus on two broad classes of random circuits, characterized by the distribution of these single-qubit gates.

- (1) **Locally unbiased circuits:** If the single-qubit gates are sampled from local unitary 1-designs, then we say that the layers are sampled from *locally unbiased* distributions. This generalizes the ensembles of circuits considered in Ref. [13], which are composed by layers invariant under right-multiplication of random Pauli unitaries. In this work, we demonstrate that circuits with locally unbiased layers affected by non-unital noise can be efficiently classically simulated.
- (2) **Approximately locally scrambling circuits:** If the single-qubit gates are sampled from local unitary 2-designs, then we say that the layers are sampled from *locally scrambling* distributions [39,42,43]. In Appendix A 2 we introduce a generalization of this notion termed *approximately locally scrambling* distributions. This broader framework encompasses, for example, arbitrary unitaries preceded by two consecutive random Pauli rotations along orthogonal directions. We further show that circuits composed of approximately locally scrambling layers, affected by arbitrary local noise, possibly dephasing-like, can be efficiently simulated classically.

**Effective depolarizing rate.** A key conceptual point behind our techniques is that *any* incoherent local noise suppresses the contribution of high-weight Pauli paths in random noisy circuits, analogously to local depolarizing noise. To formalize this idea, we associate to each ensemble of noisy circuits an *effective depolarizing rate*  $p(\mathcal{N}, \mathcal{D})$ . We develop this concept in detail in Definition C2 in Appendix C.

The value of  $p(\mathcal{N}, \mathcal{D})$  depends on the interplay between the noise channel  $\mathcal{N}$  and the circuit distribution  $\mathcal{D}$ . Importantly, we show that if the circuit is locally unbiased and  $\mathcal{N}$  is depolarizing-like or non-unital, then  $p$  is strictly positive, leading to an average-case contraction of nonidentity Pauli terms akin to that induced by depolarizing noise.

Moreover, as we show in Supplemental Lemma B2 in Appendix B, for a locally scrambling circuit distribution  $\mathcal{D}_{\text{LS}}$ , the effective depolarizing rate admits a remarkably simple expression in terms of the normal form parameters of  $\mathcal{N}$ ,

$$p(\mathcal{N}, \mathcal{D}_{\text{LS}}) = 1 - \sqrt{\frac{\|\mathbf{D}\|_2^2 + \|\mathbf{t}\|_2^2}{3}}, \quad (7)$$

which is strictly positive unless  $\mathcal{N}$  is unitary. For the important cases of amplitude-damping noise with rate  $\gamma$  and dephasing noise with rate  $q$ , this yields

$$p(\mathcal{N}_\gamma^{(\text{amp})}, \mathcal{D}_{\text{LS}}) = 1 - \sqrt{\frac{1 + 2(\gamma - 1)^2}{3}} \approx \frac{2}{3}\gamma, \quad (8)$$

$$p(\mathcal{N}_q^{(\text{deph})}, \mathcal{D}_{\text{LS}}) = 1 - \sqrt{\frac{1 + 2(1 - 2q)^2}{3}} \approx \frac{4}{3}q, \quad (9)$$

where the approximations hold respectively in the regimes  $\gamma, q \ll 1$ . Hence, the effective depolarizing rates  $p(\mathcal{N}_\gamma^{(\text{amp})}, \mathcal{D}_{\text{LS}})$  and  $p(\mathcal{N}_q^{(\text{deph})}, \mathcal{D}_{\text{LS}})$  scale linearly with physical noise rates  $\gamma$  and  $q$ .

### III. RELATED WORKS

This work combines several technical tools and insights from the previous literature. Here, we briefly review some previous works on noisy quantum circuits and on classical simulation of quantum circuits via Pauli-path methods.

**Classical simulation of quantum circuits with noise beyond the “depolarizing assumption.”** Previous works have significantly advanced the understanding of quantum noise beyond the depolarizing model. Under certain circumstances, different noise models can lead to qualitatively different scenarios. A prominent example is the so-called “quantum refrigerator” [35], which demonstrates that non-unital noise can be turned into a resource for fault tolerance, removing the need for fresh auxiliary qubits or the mid-circuit measurements required for implementing error-correcting schemes. In particular, for a noise rate below the error-correcting threshold, quantum computation under non-unital noise remains possible for  $\Theta(\exp(n))$  time [35,44]. An analogous, yet weaker, construction is also possible under dephasing noise, which allows for arbitrary quantum computation for polynomial time [35]. As there exist polynomial depth quantum circuits that are widely believed to be hard to classically simulate, these results imply that there exist at least some (carefully designed) circuits subject to non-unital noise and dephasing noise that cannot be efficiently simulated classically.

More recently, a series of works investigated the impact of arbitrary local noise on random quantum circuits [36, 37,45,46]. Even in this setting, some stark differences arise between unital and non-unital noise. In the latter case,

noisy random circuits do not exhibit anti-concentration of the output distribution [36] and exponential concentration of local observables [37,38]. Loosely speaking, these effects occur because non-unital noise effectively cools down the qubits throughout the computation. On the other hand, when averaged over random gates, any local noise induces an *effective depth* [37], that is any noise “truncates” most quantum circuits to effectively logarithmic depth, for the task of estimating observable expectation values. In particular, as discussed in Ref. [37], the existence of this noise-induced effective depth allows one to classically estimate expectation values of observables with inverse-polynomial precision in quasi-polynomial time if the circuit has constant geometric locality. The runtime improves to polynomial if the circuit architecture is one-dimensional. In the present paper, we conduct a more fine-grained analysis of this phenomenon, by extending it to a more general class of random circuits on one hand, and by employing a more sophisticated simulation algorithm on the other hand, achieving inversely polynomial precision in polynomial time across all circuit architectures.

**Classical simulation of noisy circuits via Pauli propagation algorithms.** Pauli propagation algorithms have gained momentum in recent years as powerful tools for simulating quantum circuits in a variety of settings. For circuits interspersed with noise within the depolarizing class, polynomial-time classical simulation with inverse-polynomial precision has been established, both for sampling from the output distribution of random circuits [13] and for estimating expectation values [15,29]. Recently, it was shown that arbitrary circuits under depolarizing noise, when applied to a random input state, are also classically simulable in polynomial time with inverse-polynomial precision [16].

While the results mentioned above rely on average-case assumptions, worst-case results for simulating noisy circuits using Pauli propagation have also been demonstrated in specific regimes [16,24,28,30]. Namely, any circuit under depolarizing noise is classically simulable at depths larger than logarithmic in the system size, as the output state is inversely polynomially close to the maximally mixed state [24]. Conversely, at sublogarithmic depth, circuits dominated by Clifford gates [28,30] or those with input states that are nearly maximally mixed, such as in the one clean-qubit DQC1 model of computation [16], are also classically simulable under local depolarizing noise.

Beyond circuits with depolarizing noise, Ref. [16] investigates a *specific* example of non-unital noise that effectively acts as depolarizing noise. This enables the estimation of expectation values of observables with inverse-polynomial precision in polynomial time across all circuit architectures, given a mild form of randomness in the choice of the input state. The noise model in question involves spontaneous emissions occurring in random directions. As the authors discuss, such randomization

TABLE I. Comparison of runtimes for previous and our results established under different noise models and average-case assumptions. The goal is to estimate the expectation value of an observable  $O$  with inversely polynomial precision and inversely polynomial failure probability for a polynomial depth circuit with an arbitrary topology. We assume that the observable  $O$  satisfies  $\|O\| \leq 1$  and that it contains at most  $M$  different Pauli terms. The parameter  $p$  denotes the noise rate, i.e., the physical noise rate in prior literature and, for our work, the *effective depolarizing rate*  $p(\mathcal{N}, \mathcal{D})$ , formally defined in Definition C2 in Appendix C. As discussed in Sec. II, for typical circuit distributions and noise models,  $p(\mathcal{N}, \mathcal{D})$  scales linearly with the physical noise rate.

| Local noise model                                 | Average-case assumptions                                     | Runtime                             |
|---------------------------------------------------|--------------------------------------------------------------|-------------------------------------|
| Noiseless                                         | Locally scrambling layers                                    | $n^{\mathcal{O}(\log(n))}$ [39]     |
| Depolarizing                                      | Random input state                                           | $n^{\mathcal{O}(1/p)}$ [16]         |
| Spontaneous emission                              | Emissions occurring in random directions, random input state | $Mn^{\mathcal{O}(1/p)}$ [16]        |
| Depolarizing-like [15,29], amplitude damping [47] | Circuit with Clifford gates and random Pauli rotations       | $Mn^{\mathcal{O}(1/p)}$             |
| Arbitrary non-unital noise                        | Locally unbiased layers                                      | $Mn^{\mathcal{O}(1/p)}$ (Theorem 1) |
| Arbitrary unital noise (including dephasing-like) | Approximately locally scrambling layers                      | $n^{\mathcal{O}(1/p)}$ (Theorem 2)  |

arises naturally in random circuits as an effect of the random gates applied before and after each noise channel. In our work, we extend this analysis by explicitly identifying several classes of noisy random circuits which behave effectively as subject to depolarizing noise, and providing a unified treatment for *any* arbitrary local noise, possibly non-unital or dephasing-like. Another extension of Pauli-path simulation to non-unital noise is presented in Ref. [47], which focuses on circuits consisting of Clifford gates and random Pauli rotations interspersed with amplitude-damping noise. This work introduces a randomized variant of the Pauli Propagation algorithm specifically designed for circuits affected by non-unital noise. We summarize the state-of-the-art results for estimating observables of noisy circuits in Table I.

**Classical simulation of noiseless circuits via Pauli propagation algorithms.** Moving beyond noisy devices, it has been recently proven that Pauli-path methods allow for the estimation of expectation values in noiseless random quantum circuits in polynomial time for arbitrarily small constant precision, and in quasi-polynomial time for inverse-polynomial precision [39]. Furthermore, this can be improved to almost-polynomial time if the circuit has constant geometric locality and a depth of at most polylogarithmic scale.

Crucially, the algorithm in Ref. [39] truncated all Pauli operators above a certain weight threshold. In contrast, the algorithm considered in the present paper truncates Pauli paths based on their path weight, as sketched in Fig. 1. This distinction explains the differing runtimes of the two methods. Notably, for inversely polynomial precision, the runtime for noiseless circuits in Ref. [39] is significantly higher than that achieved in our work for noisy circuits.

Moreover, Pauli-path methods are also provably efficient for noiseless near-Clifford circuits—those composed of Clifford gates and Pauli rotations by sufficiently small angles [7,48–51]. References [6,52] provide further

numerical evidence of the performance of Pauli Propagation on noiseless circuits of particular relevance in the near-term and early-fault-tolerant era, such as Trotterized circuits for Hamiltonian simulations [6] and quantum convolutional neural networks (QCNNs) [52].

A further application of Pauli-path methods consists in classically spoofing the linear cross-entropy benchmark used in quantum supremacy experiments [2,13,39,53].

## IV. RESULTS

**Theoretical guarantees.** Below we state our main theoretical results. For conciseness, we denote the effective depolarizing rate  $p(\mathcal{N}, \mathcal{D}_{\text{circ}})$  simply by  $p$ . As discussed in Sec. II, for natural circuit distributions and noise models this parameter scales linearly with the underlying physical noise rate. Our results provide *additive* error guarantees for expectation values of observables measured on the output of noisy circuits. More precisely, the error is bounded as  $\varepsilon \|O\|$ , where  $\|O\|$  denotes the operator norm of the observable. For extensive observables,  $\|O\|$  may itself scale with system size; for instance, if  $O = \sum_{P \in \{I, Z\}^{\otimes n}, |P|=1} P$ , then  $\|O\| = n$ , so the guarantee becomes an additive error of order  $\varepsilon n$ . If one instead wishes to achieve a system-size-independent absolute error bound, one may rescale  $\varepsilon$  accordingly. Since our runtime guarantees remain polynomial for any inverse-polynomial accuracy parameter  $\varepsilon$ , such a polynomial rescaling of  $\varepsilon$  preserves the overall polynomial-time scaling.

Our first contribution establishes the efficient classical simulability of typical quantum circuits interspersed by arbitrary *non-unital* noise. For inversely polynomial precision, the computational cost of our Pauli propagation algorithm (described in Sec. V) is polynomial if the measured observable contains polynomially many Pauli operators. The runtime of our algorithm thus aligns with the polynomial scalings found in Refs. [16,47] (see Table I)

but holds for a wider class of noise models. Namely, our result holds for arbitrary local non-unital noise and any random noisy circuit whose layers are sampled independently from a distribution invariant under single-qubit unitary 1-designs. We stress that some form of average-case assumption over circuits is essential due to the existence of the “quantum refrigerator” scheme [35], as discussed in detail in Appendix G.

*Theorem 1 (Non-unital noise, informal).* Let  $O$  be an observable expressed as linear combination of  $M$  Pauli operators, and let  $\rho$  be an initial state. Assume the Pauli coefficients of  $O$  and  $\rho$  can be efficiently computed. Let  $C$  be a random noisy circuit whose layers are sampled independently from a distribution invariant under single-qubit unitary 1-designs. Assume that the noise to be local and non-unital, and the noise rate to be constant. Then the noisy expectation value  $\text{Tr}[OC(\rho)]$  can be estimated classically in time  $Mn(\varepsilon\delta)^{-O(1/p)}$  within error  $\varepsilon \|O\|$  and with probability  $1 - \delta$  over the circuit randomness. In particular, when the effective depolarizing rate is constant, i.e.,  $p \in \Theta(1)$ , the algorithms run in time  $M \text{poly}(n, \varepsilon^{-1}, \delta^{-1})$ .

The proof of Theorem 1 is provided in Supplemental Theorem D2 in Appendix D2. This result also closes a gap in understanding the relationship between barren plateaus and classical simulability [54] in the context of generic noisy circuits. While previous works have shown that non-unital noisy random circuits do not exhibit barren plateaus [37] (although due to only their last few layers), only a quasi-polynomial time classical average-case algorithm was known to achieve inverse-polynomial precision in estimating observable expectation values. Our result improves upon this by reducing the time complexity to polynomial time, demonstrating that in this physically motivated non-unital noisy setting, the absence of barren plateaus is indeed accompanied by efficient average-classical simulation. More specifically, this provides a concrete example where we observe absence of barren plateaus [37], efficient average-classical simulability (our work), and provable worst-case hardness for classical simulability [35].

We also establish the efficient classical simulability of typical quantum circuits interspersed by arbitrary unital noise, possibly within the *dephasing* class. In this case, the computational cost for inversely polynomial precision is polynomial for all observables. Here, we make a slightly stronger assumption on the circuit ensemble, by assuming the circuit layers to be sampled from distributions invariant under single-qubit gates satisfying a property which we call *approximate scrambling* (cf. Definition A4). This natural property is satisfied by a broad class of gate-sets, such as Clifford gates and consecutive Pauli rotations along two orthogonal axes.

*Theorem 2 (Unital noise, informal).* Let  $O$  be an observable and let  $\rho$  be an initial state. Assume that the Pauli coefficients of  $O$  and  $\rho$  can be efficiently computed. Let  $C$  be a random noisy circuit whose layers are sampled independently from a distribution invariant under approximately scrambling single-qubit gates. Assume the noise to be local and unital, possibly dephasing-like, and that the noise rate is constant.

- (1) The noisy expectation value of any observable  $O$  expressed as a linear combination of  $M$  Pauli operators can be estimated classically in time  $Mn(\varepsilon\delta)^{-O(1/p)}$  within additive error  $\varepsilon \|O\|$  with probability at least  $1 - \delta$  over the circuit randomness. In particular, when the effective depolarizing rate is constant, i.e.,  $p \in \Theta(1)$ , the algorithms runs in time  $M \text{poly}(n, \varepsilon^{-1}, \delta^{-1})$ .
- (2) If the circuit depth is at least logarithmic in the number of qubits  $n$  and the effective depolarizing rate is constant, then the noisy expectation value of any observable  $O$  can be estimated classically in time in time  $\text{poly}(n)$  within additive error  $\varepsilon \|O\|$  with probability at least  $1 - \delta$  over the circuit randomness, for arbitrary  $\varepsilon, \delta$  decaying inverse-polynomially in  $n$ .

The proof of Theorem 2 is provided in Supplemental Theorem D2 in Appendix D2.

A distinguishing feature of Theorems 1 and 2 is that they provide runtimes independent of the circuit depth, in contrast to related approaches that achieve only polynomial scaling for very shallow circuits (e.g., circuits with a logarithmic number of non-Clifford gates [28]). This property stems from the fact that, for the simulation of *noisy* circuits, the computational complexity of the Pauli propagation algorithm equipped with path-weight truncation does not grow unboundedly with increasing circuit depth. Crucially, this depth-independence has a direct physical interpretation in the presence of noise. Specifically, we show that any local noise effectively truncates typical circuits to a logarithmic depth for the task of estimating expectation values, generalizing the main result of Ref. [37] to a broader class of circuits. This result implies that for inversely polynomial precision, the full circuit  $C$  can effectively be replaced by a logarithmic-depth circuit, specifically by its last logarithmic-many layers, which can potentially be simulated classically.

*Theorem 3 (Noise-induced shallow depth, informal).* Let  $O$  be an observable and consider an  $L$ -layered circuit  $C$  randomly sampled from  $\mathcal{D}_{\text{circ}}$  where  $\mathcal{D}_{\text{circ}}$  is a distribution over noisy circuits satisfying the assumptions of Theorems 1 or 2. Consider the “truncated” noisy circuit

$$\begin{aligned} C_{[L, L-j]} &:= \mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ U_L \circ \mathcal{N}^{\otimes n} \circ U_{L-1} \circ \dots \\ &\quad \circ \mathcal{N}^{\otimes n} \circ U_{L-j}, \end{aligned} \quad (10)$$

for a suitable  $j \in \mathcal{O}(p^{-1} \log(\varepsilon^{-1} \delta^{-1}))$ . With probability at least  $1 - \delta$  over the choice of  $\mathcal{C}$ , it holds that

$$|\text{Tr}[OC(\rho)] - \text{Tr}[OC_{[L,L-j]}(\sigma)]| \leq \varepsilon \|O\|, \quad (11)$$

where  $\sigma$  is an arbitrary state.

The proof of Theorem 3 is provided in Supplemental Theorem E2 in Appendix E. The above result tells us that, for the task of estimating expectation values, noisy random circuits can be truncated to an effective logarithmic depth: assuming that  $p \in \Theta(1)$  and  $\varepsilon, \delta \in 1/\text{poly}(n)$ , then we can choose  $j \in \mathcal{O}(\log(n))$ . It is noteworthy that combining this result with the recent findings from Ref. [39], already provides a non-trivial simulation algorithm for noisy random circuits. In particular, Ref. [39] demonstrated that observables of logarithmic-depth circuits can be estimated with inversely polynomial precision in almost-polynomial time [i.e.,  $n^{\mathcal{O}(\log \log(n))}$ -time] for circuits with constant geometric dimension [55]. However, the bounds derived in the present paper are significantly stronger, as they establish a polynomial runtime for noisy circuits across all architectures.

The noise-induced effective-depth results reflect the aforementioned fact that the hardness of simulating noisy circuits does not grow indefinitely with circuit depth. In fact, it is well known that in certain regimes simulating *deep* noisy circuits can be easier than simulating *shallow* ones. In the case of depolarizing-like noise, previous results show that sufficiently deep noisy circuits can be well approximated by the maximally mixed state, a phenomenon that can be justified using entropy-accumulation arguments. These arguments, however, do not apply to dephasing channels, for which the maximally mixed state is not the *only* fixed point, or to general non-unital noise, for which the maximally mixed state is not a fixed point at all.

Nevertheless, for random circuits with *arbitrary* local noise we are able to recover analogous conclusions without resorting to entropy-accumulation arguments. For dephasing-like noise, our analysis shows that observables behave essentially as in the depolarizing setting: the expectation value  $\text{Tr}[OC(\rho)]$  of any traceless observable  $O$  concentrates around zero for large depths, so for deep circuits it can be easily estimated by “guessing zero” rather than running a classical simulation algorithm. The situation is more subtle for non-unital channels, where the recent work [37] demonstrates that random noisy circuits do not exhibit barren plateaus, implying that expectation values of traceless observables do not in general converge to zero. Nonetheless, we prove that the deviation from zero is uniformly controlled by a quantity  $\|\mathbf{t}\|_\infty^2 / (p(2-p))$  determined by the parameters of the noise channel. Thus, whenever  $\|\mathbf{t}\|_\infty^2 / (p(2-p)) \leq \varepsilon$ , the expectation value is  $(\varepsilon + o(1))$ -close to zero for sufficiently deep circuits,

and one may effectively replace the simulation by simply guessing 0.

This intuition is formalized by the following Proposition 1, whose proof is provided in Supplemental Proposition F1 in Appendix F.

**Proposition 1 (Noise-induced concentration under arbitrary noise).** For all traceless observables  $O$  and quantum states  $\rho$  it holds that

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[OC(\rho)]^2 \leq \left[ (1-p)^{2L} + \frac{\|\mathbf{t}\|_\infty^2}{p(2-p)} \right] \|O\|^2. \quad (12)$$

For amplitude-damping noise of noise rate  $\gamma$ , this bound becomes particularly transparent. Since  $\|\mathbf{t}\|_\infty = \gamma$  and  $p = \Theta(\gamma)$ , one obtains

$$\frac{\|\mathbf{t}\|_\infty^2}{p(2-p)} = \Theta(\gamma),$$

so the deviation from zero of  $|\text{Tr}[OC(\rho)]|$  scales as  $\mathcal{O}(\gamma^{1/2})$  in expectation. Consequently, if the target accuracy satisfies  $\varepsilon \gg \gamma^{1/2}$  and the circuit is deep enough that the exponential term  $(1-p)^{2L}$  is negligible, then simply outputting 0 allows one to estimate  $\text{Tr}[OC(\rho)]$  to accuracy  $\varepsilon$ .

**Numerical implementation.** We complement our theoretical analysis with precise numerical estimates of the average simulation error for a Hamiltonian variational ansatz [56] and a large-scale simulation of noisy quantum dynamics. All numerical results were collected with the PauliPropagation.jl package [57,58].

As shown in Figs. 3 and 4, the average simulation error encountered in practice is often significantly smaller than predicted by our bounds, demonstrating that Pauli-path methods are very well-suited for simulating typical noisy circuits. Both simulations treat the same system, but with varying degrees of amplitude-damping noise (Fig. 3) or dephasing noise (Fig. 4). The quantum circuits consist of parametrized RX and RZ Pauli rotation gates per qubit followed by parametrized RZZ Pauli rotation gates on a periodic  $6 \times 6$  lattice on 36 qubits. This corresponds to an Hamiltonian variational ansatz (HVA) ansatz of the so-called tilted transverse-field Ising model. The observable here is a Pauli-Z operator in the middle of the lattice, and the initial state is the all-zero state. We emphasize that this circuit ensemble would not satisfy the stronger assumptions of other works, such as Refs. [37,39], which require the presence of local 2-designs. Moreover, we note that simulating deep average-case circuits with amplitude-damping noise is arguably more interesting than with noiseless or depolarizing noise [26,39,59], due to the fact that such circuits can experience “absence of barren plateaus” [37], highlighting the need for non-trivial

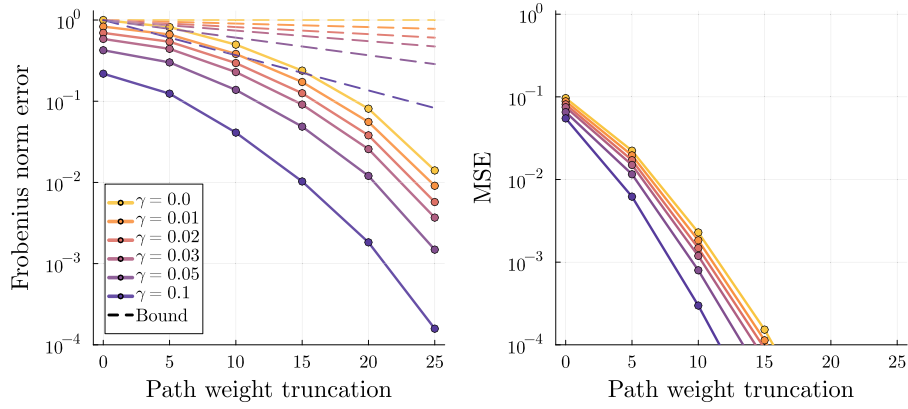


FIG. 3. Error of simulating circuits with amplitude-damping noise. (a) Frobenius norm error and (b) mean square error for different path-weight truncation values and noise strengths in a periodic  $6 \times 6$  lattice of qubits. The observable is a Pauli-Z operator in the middle of the lattice with a circuit consisting of RX, RZ single-qubit rotations and RZZ entangling gates and in (b) the initial state is the all-zero state. The theoretical bound, established in Supplemental Theorem D1 in Appendix D, is calculated via  $(1 - \gamma + \gamma^2)^k$ , where  $k$  is the path-weight truncation order and  $1 - \gamma + \gamma^2$  is the squared contraction coefficient computed in Example B1 in Appendix B.

classical algorithms to tackle such cases. In Appendix H, we further elaborate on this and demonstrate the absence of barren plateaus for our circuit ensemble, which does not meet the stronger assumptions used in Ref. [37].

Our bounds on the Frobenius norm error of the back-propagated observable as a function of the path-weight truncation order are well satisfied [see panel (a)]. This norm also indirectly bounds the error of expectation values with random initial states. We show, however, that the mean square error of expectation values with the ubiquitous all-zero initial state is orders of magnitude lower at the same path-weight truncation order [see panel (b)]. These error estimates were obtained via the Monte Carlo approach provided in Supplemental Theorem A1 in Appendix A4b. Overall, we conclude that the simulation error tends to decay exponentially with the path-weight truncation order, and that circuits affected

by stronger noise are continually easier to simulate on average.

To unquestionably go beyond our theoretical guarantees, and to position Pauli propagation as a key tool for simulating quantum systems, we showcase a large-scale simulation of noisy real-time dynamics. As with most numerical simulations that go beyond brute-force computation, it becomes difficult to judge the reliability of the results. To address this we start by using a system from Ref. [60], which compares the converging Pauli propagation results to iPEPS simulation data. Namely, we consider an  $11 \times 11$  square qubit lattice with a rotated transverse-field Ising Hamiltonian,

$$H = -J \sum_{\langle i,j \rangle} X_i X_j - h \sum_i Z_i, \quad (13)$$

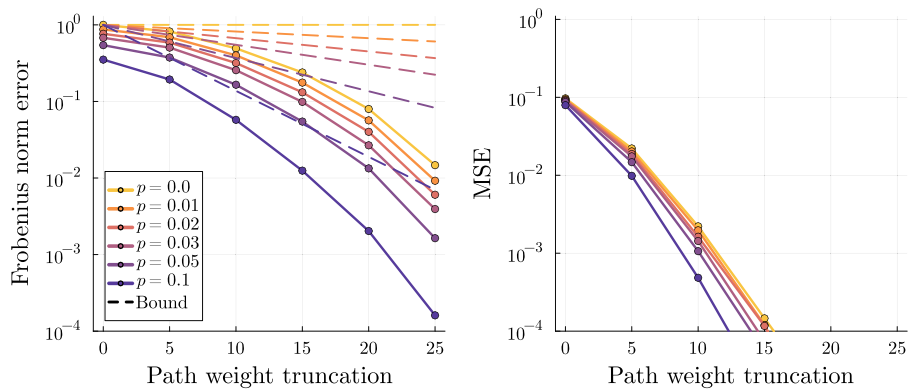


FIG. 4. Error of simulating circuits with dephasing noise. (a) Frobenius norm error and (b) mean square error for the same 36-qubit simulation as in Fig. 3. The theoretical bound, established in Supplemental Theorem D1 in Appendix D, is calculated via  $((1 + (1 - 2p)^2)/2)^k$ , where  $k$  is the path-weight truncation order and  $(1 + (1 - 2p)^2)/2$  is the mean squared contraction coefficient computed in Example B1 in Appendix B.

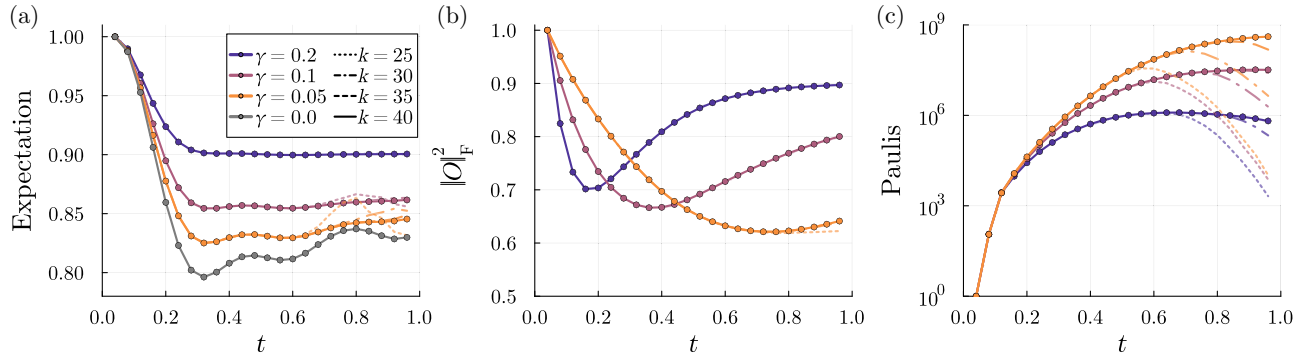


FIG. 5. Simulation of noisy real-time quantum dynamics. We simulate the expectation value of a Pauli-Z observable in the middle of an  $11 \times 11$  arrangement of 121 qubits evolving under a rotated transverse-field Ising model at a quantum critical point. (a) Expectation values simulated for different noise strengths  $\gamma$  and path weight thresholds  $k$ . The curve for  $\gamma = 0$  is taken from Ref. [60]. It is apparent that larger  $\gamma$  allows to converge the simulation with lower  $k$ . (b) Frobenius norm of the evolving observable as a function of time. The nonmonotonic behavior implies the departure from our random circuit assumptions. (c) The number of stored Pauli strings for the employed path-weight truncations, concluding that these noisy dynamics are indeed increasingly more efficient with stronger amplitude-damping noise.

where the XX interaction terms with strength  $J$  act on neighbors on the 2D lattice, and  $h$  is the local field strength. Importantly, we choose  $h = 1$  and  $J = 3.004438$ , which puts the system at a quantum critical point [61]. The quantum circuit is then constructed as the second-order Trotterization of this Hamiltonian with time step  $dt = 0.04$  and up to 23 Trotter steps. The initial state is the all-zero state, and we estimate the expectation value of a Pauli-Z operator in the middle of the 2D lattice as a function of time.

Reference [60] truncates propagating Pauli operators with more than 5 X or Y Paulis (this is not Pauli weight, which also counts Z Paulis) and Pauli operators with small coefficients below a magnitude of  $2^{-18}$  up to  $2^{-23}$ . They then use their comparison with an iPEPS simulation to argue that these truncations are sufficient to ensure convergence. We here employ the most challenging coefficient truncation threshold of  $2^{-23}$  and substitute their XY-Pauli truncation with path weight truncation.

In Fig. 5 we see that, while the noise-free system was reported to converge only with 1TB of memory usage, injecting local amplitude-damping noise into the quantum circuit allows simulating and converging the noisy systems with significantly less computational resources. The initial dependence is exponential in circuit depth, but it plateaus depending on the noise strength. In particular, the operator norm and expectation values for  $\gamma = 0.1$  and  $\gamma = 0.2$  can trivially be converged within seconds to a few minutes on a laptop using open-source code. Only for  $\gamma = 0.05$  do we require moderately larger resources for full convergence. Additionally, we show that the computational scaling of the path weight truncation is mostly independent of the system size (see Fig. 7 in Appendix H), as well as the particular choice of initial state and local observable (see Fig. 8).

We stress that this setup strongly deviates from the theoretical assumptions needed to derive our bounds. Namely, we are considering a highly structured circuit with correlated angles, which is exemplified by the Frobenius norm in Fig. 5(b). Supplemental Lemma B3 in Appendix B and its randomness assumptions would predict the norm to decay monotonically as the amplitude-damping noise effectively acts like depolarizing noise. However, we observe significant nonmonotonic behavior. Despite this departure from our theoretical guarantees, we see that systems with stronger amplitude-damping noise (i) require lower and lower path weight to converge, and (ii) drastically fewer Pauli strings are kept at the same path weight. The eventual drop of Pauli strings in Fig. 5(c) is due to the path weight monotonically increasing over time and eventually surpassing the threshold to be truncated—the drop is more dramatic for lower  $k$  because more terms are truncated. So far, no theoretical accuracy guarantees have been given in this regime. But we show that for systems of interest, even at a quantum critical point, Pauli propagation equipped with suitable truncations can provide accurate results for quantum dynamics subject to non-unital noise.

## V. METHODS

In this section, we outline the primary methods and proof strategies of this work.

*Notation.* In order to illustrate our methods and proof ideas, we first introduce some essential notation. Let  $\mathcal{P}_n := \{I, X, Y, Z\}^{\otimes n}$  be the  $n$ -qubit Pauli basis. Given a Pauli operator  $P = P^{(1)} \otimes P^{(2)} \otimes \dots \otimes P^{(n)} \in \mathcal{P}_n$ , we define its support, denoted as  $\text{supp}(P)$ , as the subset of qubits upon which  $P$  acts nontrivially, and we define its Pauli weight, denoted as  $|P|$  as the cardinality of its support

or, equivalently, as the number of nonidentity single-qubit Pauli operators occurring in the tensor decomposition of  $P$ . Given an observable  $O = \sum_{P \in \mathcal{P}_n} a_P P$ , we define its squared normalized Frobenius norm as

$$\|O\|_F^2 := \sum_{P \in \mathcal{P}_n} a_P^2. \quad (14)$$

**Pauli-path analysis.** Our algorithm and guarantees draw heavily on a Pauli-path based analysis. We start by rewriting the noisy circuit model in Eq. (6) as follows:

$$\begin{aligned} \mathcal{C} &:= \underbrace{\mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_L}_{\mathcal{C}_L} \circ \underbrace{\mathcal{N}^{\otimes n} \circ \mathcal{U}_{L-1}}_{\mathcal{C}_{L-1}} \circ \cdots \circ \underbrace{\mathcal{N}^{\otimes n} \circ \mathcal{U}_1}_{\mathcal{C}_1} \\ &= \mathcal{C}_L \circ \mathcal{C}_{L-1} \circ \cdots \circ \mathcal{C}_1. \end{aligned} \quad (15)$$

In what follows we trade generality for conciseness by assuming that the circuit layer distributions are invariant under random single-qubit Clifford gates. We emphasize that our results extend to more general circuit ensembles, which require a more detailed and nuanced analysis, as elaborated in the Appendixes.

Our technique consists in computing an approximate Heisenberg evolution of the observable via a truncated Pauli-path summation, analogous to similar approaches exploited for simulating noisy and noiseless circuits in the previous literature. In particular, we truncate Pauli paths according to their path weight (also referred as Hamming weight in the prior literature), thus our estimator coincides with that employed in Refs. [13,16,30].

As the Pauli operators forms a basis for the Hermitian operators, we can express the Heisenberg-evolved observable  $\mathcal{C}^\dagger(O)$  as a sum across Pauli paths as follows:

$$\begin{aligned} \mathcal{C}^\dagger(O) &= \sum_{P_0, P_1, \dots, P_L \in \mathcal{P}_n} \langle\langle O | P_L \rangle\rangle \langle\langle P_L | \widehat{\mathcal{C}}_L | P_{L-1} \rangle\rangle \\ &\quad \times \langle\langle P_{L-1} | \widehat{\mathcal{C}}_{L-1} | P_{L-2} \rangle\rangle \cdots \langle\langle P_1 | \widehat{\mathcal{C}}_1 | P_0 \rangle\rangle P_0, \end{aligned} \quad (16)$$

where we employed the normalized vectorized notation  $\langle\langle O | P_L \rangle\rangle := (1/2^n) \text{Tr}[OP_L]$ ,  $\langle\langle P_j | \widehat{\mathcal{C}}_j | P_{j-1} \rangle\rangle := (1/2^n) \text{Tr}[P_j \mathcal{C}_j(P_{j-1})]$ . We can also rearrange Eq. (16) as follows by introducing a more compact notation:

$$\mathcal{C}^\dagger(O) := \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle\langle O | P_L \rangle\rangle \Phi_\gamma(\mathcal{C}) P_0, \quad (17)$$

where  $\gamma = (P_0, P_1, \dots, P_L)$  denotes a Pauli path and  $\Phi_\gamma(\mathcal{C}) := \langle\langle P_L | \widehat{\mathcal{C}}_L | P_{L-1} \rangle\rangle \langle\langle P_{L-1} | \widehat{\mathcal{C}}_{L-1} | P_{L-2} \rangle\rangle \cdots \langle\langle P_1 | \widehat{\mathcal{C}}_1 | P_0 \rangle\rangle$  denotes the associated Fourier coefficient.

As the above sum involves exponentially many Pauli paths, in order to keep a manageable runtime we consider

only those with a small path weight  $|\gamma|$ , which we define as follows:

$$|\gamma| := \sum_{j=1}^L |P_j|. \quad (18)$$

Thus, given a truncation order  $k > 0$ , we consider the following weight-truncated estimator:

$$\mathcal{C}_k^\dagger(O) := \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1} \atop |\gamma| < k} \langle\langle O | P_L \rangle\rangle \Phi_\gamma(\mathcal{C}) P_0. \quad (19)$$

Thus, we approximate the expectation value  $\text{Tr}[\mathcal{C}^\dagger(O)\rho]$  by computing the overlap between  $\mathcal{C}_k^\dagger(O)$  and  $\rho$ , i.e.,  $\text{Tr}[\mathcal{C}_k^\dagger(O)\rho]$ .

**The Pauli propagation algorithm.** The approximate Heisenberg-evolved observable  $\mathcal{C}_k^\dagger(O)$  can be computed with an iterative process. Assume without loss of generality that  $O$  consists of a single Pauli term with weight less than  $k$ . If  $O$  is a linear combination of multiple Pauli terms, we will compute the contribution of each term separately. Moreover, Pauli terms with weight at least  $k$  can be discarded, as they will form Pauli path with weight at least  $k$ . We start by computing  $\mathcal{C}_L^\dagger(O)$  in the Pauli basis:

- (1) If  $\mathcal{C}_L$  contains only Clifford unitaries and Pauli noise channels, then  $\mathcal{C}_L^\dagger(O)$  is also a single Pauli operator.
- (2) In general,  $\mathcal{C}_L^\dagger(O)$  can be expressed as a linear combination of Pauli operators. In this case, we say that the layer  $\mathcal{C}_L$  *splits*  $O$  in different Pauli paths. This is the case if  $\mathcal{C}_L$  contains non-Clifford unitaries or non-unital noise channels.

For each Pauli path obtained as described above, we can then apply the adjoint channel  $\mathcal{C}_{L-1}^\dagger(\cdot)$  and proceed backward until  $\mathcal{C}_1^\dagger(\cdot)$ . Thus, after each iteration the Pauli paths may split again, forming a tree-like structure as shown in Fig. 1(c).

Computing all the Pauli paths would require a runtime exponential in the number of non-Clifford unitaries and non-unital noise channels. Therefore, we combine such iterative method with a carefully chosen *truncation rule*, which consists in discarding all the Pauli paths with path weight larger than a cutoff  $k$ , as in Eq. (19), ensuring that the number of computed Pauli paths remains tractable. In particular, for an observable composed of polynomially many Pauli terms and  $k$  scaling logarithmically in system size, the above procedure requires a polynomial runtime. This truncation rule has also been employed in Refs. [13,16,30] and bears similarities with the strategies adopted in Refs. [6,15,29].

If we further assume that the noise is unital, then the approximate Heisenberg-evolved observable  $\mathcal{C}_k^\dagger(O)$  can also be computed with an alternative strategy proposed in Ref. [13], which yields a polynomial runtime provided that  $k$  scales logarithmically in system size.

**Intuitive explanation of theoretical guarantees.** The widespread success of Pauli-path methods under depolarizing noise can be attributed to a key feature: depolarizing noise has a remarkably simple representation in the Pauli basis. For a Pauli operator  $P \in \mathcal{P}_n$ , its evolution under the local depolarizing channel  $\mathcal{N}_p^{(\text{depo}) \otimes n}$  with noise rate  $p$  is given by

$$\mathcal{N}_p^{(\text{depo}) \otimes n}(P) = \mathcal{N}_p^{\dagger(\text{depo}) \otimes n}(P) = (1-p)^{|P|}P, \quad (20)$$

where the first identity follows from the fact that the depolarizing channel and its adjoint are equal. Therefore, the Frobenius norm of  $P$  is damped by a factor  $(1-p)^{|P|}$ , as we have that  $\|\mathcal{N}_p^{(\text{depo}) \dagger \otimes n}(P)\|_F = (1-p)^{|P|}$ . Moreover, the Frobenius norm of an observable  $O = \sum_{P \in \mathcal{P}_n} a_P P$  takes the following form after the application of the depolarizing channel,

$$\|\mathcal{N}_p^{\dagger(\text{depo}) \otimes n}(O)\|_F^2 = \sum_{P \in \mathcal{P}_n} a_P^2 (1-p)^{2|P|}. \quad (21)$$

Non-unital noise channels, in stark contrast with the behavior of depolarizing noise, can increase the Frobenius norm of observables under Heisenberg evolution. However, as we prove in Supplemental Lemma B2 in Appendix B, when averaged over random gates, any incoherent noise channel contracts the Frobenius norm of Pauli operators under Heisenberg evolution, as we have

$$\mathbb{E}_{V \sim \mathcal{D}} \|\mathcal{N}^{\dagger \otimes n}(V^\dagger P V)\|_F^2 \leq (\chi_{\mathcal{D}}^2(\mathcal{N}))^{|P|}, \quad (22)$$

where  $\chi_{\mathcal{D}}^2(\mathcal{N}) := (\|\mathbf{D}\|_2^2 + \|\mathbf{t}\|_2^2)/3$  is the mean squared contraction coefficient of  $\mathcal{N}$  with respect to the distribution  $\mathcal{D}$ . As this coefficient is strictly smaller than one, the Frobenius norm of Pauli operators is damped exponentially with respect to the Pauli weight  $|P|$ , on average over the random unitaries. Similarly, given an observable  $O = \sum_{P \in \mathcal{P}_n} a_P P$ , we have that

$$\mathbb{E}_{V \sim \mathcal{D}} \|\mathcal{N}^{\dagger \otimes n}(V^\dagger O V)\|_F^2 \leq \sum_{P \in \mathcal{P}_n} a_P^2 (\chi_{\mathcal{D}}^2(\mathcal{N}))^{|P|}, \quad (23)$$

which also matches the effect of local depolarizing noise of Eq. (21).

Given the striking resemblance in Heisenberg picture between the action of depolarizing noise and the *average* action of arbitrary noise, it is convenient to rewrite the channel  $\mathcal{N}$  as follows:

$$\mathcal{N} := \mathcal{N}_p^{(\text{depo})} \circ \tilde{\mathcal{N}}_p, \quad (24)$$

where  $\tilde{\mathcal{N}}_p$  is a suitable (non-physical) linear map and  $p$  is the *effective depolarizing rate* of the channel  $\mathcal{N}$ ,

$$p := 1 - \chi_{\mathcal{D}}(\mathcal{N}). \quad (25)$$

This strategy is inspired by Ref. [16], which proposed a similar decomposition for noisy circuits affected by spontaneous emissions occurring in random directions.

Having established this correspondence between depolarizing noise and arbitrary noise averaged over random gates, we can conduct a Pauli-path analysis akin to that introduced in Ref. [13]. As for the depolarizing case, we find that the contribution of different Pauli paths is exponentially suppressed with respect to their path-weight, which allows us to show that

$$\mathbb{E} \left| \text{Tr}[(\mathcal{C}_k^\dagger(O) - \mathcal{C}^\dagger(O)) \rho] \right| \in \exp(-\Omega(k)). \quad (26)$$

Thus, the average error over the random choice of the circuit can be exponentially suppressed by increasing the truncation order  $k$ . Moreover, we also demonstrate that the truncated observable  $\mathcal{C}_k^\dagger(O)$  can be computed in  $\text{poly}(n)$ -time provided that  $k \in \mathcal{O}(\log(n))$  and the observable  $O$  is a linear combination of at most  $\text{poly}(n)$  Pauli operators.

## VI. OPEN PROBLEMS

In this paper we have conducted a systematic study of random quantum circuits interspersed by arbitrary incoherent noise. Taken together, our results provide a robust evidence that expectation values of typical noisy circuits can be efficiently estimated classically with Pauli-path methods. Nevertheless, several questions remain unaddressed by our results. Here, we outline a series of open problems for future investigation.

- (1) **Estimating “dense” Hamiltonians under non-unital noise.** Our main result, Theorem 1, guarantees that the noisy expectation values of an observable  $H$  can be estimated in polynomial time with precision  $\text{poly}(n^{-1}) \|H\|_F$  on most circuits, provided that  $H$  contains *polynomially many* Pauli terms. However, this constraint excludes an interesting class of observables representing dense Hamiltonians with long-range interactions. For instance,

consider the observable

$$H = \sum_{\substack{P \in \{I, Z\}^{\otimes n} \\ |P| \leq \ell}} P. \quad (27)$$

If  $\ell \in \omega(1)$ , then  $H$  contains  $n^{\omega(1)}$  Pauli terms, which makes the runtime of our algorithm super-polynomial.

- (2) **Classically sampling under arbitrary noise.** In their seminal work, the authors of Ref. [13] provided a polynomial-time classical algorithm for sampling from the distribution of random circuits under depolarizing noise. It is natural to consider whether this result can be extended to circuits subject to arbitrary noise. Notably, non-unital noise introduces several challenges. For instance, the algorithms proposed in Ref. [13] estimate the expectation values of projectors using a truncated path summation, with a truncation order of  $\Theta(\log(n))$ . This approach creates a scenario analogous to that encountered with dense Hamiltonians, as described earlier. Consequently, the algorithms discussed in this paper would require quasi-polynomial time for such computations. On the other hand, the proof technique of Ref. [13] requires the ensemble of noiseless quantum circuit to exhibit anti-concentration. This poses an additional challenge, as random circuits with non-unital noise do not exhibit anti-concentration [36].
- (3) **Circuits with correlated parameters.** While our results apply to a broad class of random quantum circuits, we have assumed that all circuit layers are sampled independently. This assumption, though common in the literature on random quantum circuits, does not account for scenarios where different layers are correlated. Such correlations arise, for example, in Trotterized circuits for simulating Hamiltonian dynamics and in circuits used for variational quantum algorithms [62]. Although variational circuits are often initialized with random parameters, these parameters are updated during training, breaking the independence assumption. Investigating circuits with correlated parameters presents significant theoretical challenges. However, numerical studies such as the one in Fig. 5 could offer valuable insights into their classical simulability.
- (4) **Alternative noise models.** Despite its generality, our noise models also come with its limitations. Namely, we are restricted to *local* incoherent noise models that are independent of the particular gate applied. However, realistic quantum hardware is also subject to coherent errors and errors with long-range correlations (such as those arising from cross talk [63,64]). The classical simulability of such

noisy circuits is an important direction for future research.

## CODE AVAILABILITY

The numerical simulations in this work were performed with the open-source PauliPropagation.jl package [57,58].

## ACKNOWLEDGMENTS

The authors thank Lennart Bittel, Sumeet Khatri, Victor Martinez, Thomas Schuster, Daniel Stilck França, and Chu Zhao for valuable discussions and feedback. A.A. and Z.H. acknowledge support from the Sandoz Family Foundation-Monique de Meuron program for Academic Promotion. A.A.M. acknowledges support by the German Federal Ministry for Education and Research (BMBF) under the project FermiQP. A.A.M. was also supported by the U.S. DOE through a quantum computing program sponsored by the Los Alamos National Laboratory (LANL) Information Science & Technology Institute and by the Laboratory Directed Research and Development (LDRD) program of LANL under Project No. 20230049DR. M.S.R. acknowledges funding from the 2024 Google PhD Fellowship. M.C. acknowledges support by LANL ASC Beyond Moore's Law project and by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research through the Accelerated Research in Quantum Computing Program MACH-Q project. M.S.R. and Z.H. acknowledge support of the NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (Grant No. 205602).

## DATA AVAILABILITY

The data that support the findings of this article are openly available [58].

## APPENDIX A: PRELIMINARIES

### 1. Notation

We briefly introduce the notations employed in this paper. Given a positive integer  $n$ , we denote the set  $[n] := \{1, 2, \dots, n\}$ . Given a distribution  $\mathcal{D}$  over the set  $\mathcal{X}$  and a measurable function  $F : \mathcal{X} \rightarrow \mathbb{R}$ , we denote the expectation value of  $F(X)$  for  $X$  randomly sampled from  $\mathcal{D}$  as  $\mathbb{E}_{X \sim \mathcal{D}}[F(X)]$ . Given two distributions  $\mathcal{D}_1, \mathcal{D}_2$ , we denote by  $\mathcal{D}_1 \otimes \mathcal{D}_2$  the derived distribution obtained by sampling independently  $X_1$  from  $\mathcal{D}_1$  and  $X_2$  from  $\mathcal{D}_2$ , and outputting  $X_1 \otimes X_2$ .

**Linear operators.** Let  $\mathcal{H}_n$  denote the Hilbert space for  $n$ -qubits, and let  $\mathcal{B}(\mathcal{H}_n)$  represent the space of linear operators acting on  $\mathcal{H}_n$ . We introduce the *vectorization* map  $\text{vec} : \mathcal{B}(\mathcal{H}_n) \rightarrow \mathcal{H}_n^{\otimes 2}$ , which is defined as follows: for any  $i, j \in \{0, 1\}^n$ , we set  $\text{vec}(|i\rangle\langle j|) := |i\rangle \otimes |j\rangle$ . By linearity, if  $A \in \mathcal{B}(\mathcal{H}_n)$  is expressed as  $A =$

$\sum_{i,j \in \{0,1\}^n} A_{ij} |i\rangle\langle j|$ , then the vectorization of  $A$  is given by  $\text{vec}(A) = \sum_{i,j \in \{0,1\}^n} A_{ij} |i\rangle \otimes |j\rangle$ . We also adopt notation  $|A\rangle\rangle := \frac{1}{\sqrt{2^n}} \text{vec}(A)$ , which includes a normalization factor. Notably, the canonical inner product between two normalized vectorized operators  $A$  and  $B$  is simply their normalized Hilbert-Schmidt inner product:  $\langle\langle A|B\rangle\rangle = \frac{1}{2^n} \text{Tr}[A^\dagger B]$ , where  $\langle\langle A|B\rangle\rangle := \frac{1}{2^n} \text{vec}(A)^\dagger \text{vec}(B)$ . The following *ABC-rule* is easily verified:

$$|ABC\rangle\rangle = A \otimes C^T |B\rangle\rangle, \quad (\text{A1})$$

for all  $A, B, C \in \mathcal{B}(\mathcal{H}_n)$ .

**Linear maps.** We denote by  $\mathcal{L}_{n \rightarrow n}$  the set of all linear maps of the form  $\Phi : \mathcal{B}(\mathcal{H}_n) \rightarrow \mathcal{B}(\mathcal{H}_n)$ . Quantum channels are linear maps which are also completely positive and trace-preserving. Given  $\Phi \in \mathcal{L}_{n \rightarrow n}$ , it is important to note that the map  $|\Phi(X)\rangle\rangle$  is linear in  $|X\rangle\rangle$  for all  $|X\rangle\rangle \in \mathcal{H}_n^{\otimes 2}$ . This linearity guarantees the existence of a matrix  $\widehat{\Phi} \in \mathcal{B}(\mathcal{H}_n^{\otimes 2})$  that represents this linear transformation. Hence, the action of  $\Phi$  on an operator  $X$  can be written as

$$|\Phi(X)\rangle\rangle = \widehat{\Phi} |X\rangle\rangle \quad \text{for all } X \in \mathcal{B}(\mathcal{H}_n). \quad (\text{A2})$$

Furthermore, for every linear map  $\Phi : \mathcal{B}(\mathcal{H}_n) \rightarrow \mathcal{B}(\mathcal{H}_n)$ , there exists [65] a set of matrices  $\{A_i\}_{i=1}^{4^n}, \{B_i\}_{i=1}^{4^n} \in \mathcal{B}(\mathcal{H}_n)$  such that  $\Phi(X) = \sum_{i=1}^{4^n} A_i X B_i^\dagger$ . Using the *ABC-rule* (A1), we can rewrite  $\Phi(X) = \sum_{i=1}^{4^n} A_i X B_i^\dagger$  as  $|\Phi(X)\rangle\rangle = \sum_{i=1}^{4^n} A_i \otimes B_i^* |X\rangle\rangle$ , which implies  $\widehat{\Phi} = \sum_{i=1}^{4^n} A_i \otimes B_i^*$ , that we denote as the matrix form of the linear map.

We will always use the *hat* symbol over a linear map ( $\widehat{\Phi}$ ) to indicate the matrix form of the linear map. It is also easy to observe that the matrix form of composition of two linear maps is equal to the product of the associated matrix form: Let  $\Phi_1, \Phi_2$  be two linear maps, then we have  $\widehat{\Phi_2 \circ \Phi_1} = \widehat{\Phi_2} \cdot \widehat{\Phi_1}$ .

For every linear map  $\Phi : \mathcal{B}(\mathcal{H}_n) \rightarrow \mathcal{B}(\mathcal{H}_n)$ , written as  $\Phi(X) = \sum_{i=1}^{4^n} A_i X B_i^\dagger$ , its adjoint linear map (or Heisenberg representation) can be defined as  $\Phi^\dagger(X) := \sum_{i=1}^{4^n} A_i^\dagger X B_i$ . The matrix form of the adjoint linear map  $\Phi^\dagger$  is equal to the adjoint of the matrix form of the linear map, i.e.,  $\widehat{\Phi^\dagger} = \widehat{\Phi}^\dagger$ .

**Pauli basis.** We define the Pauli basis as  $\mathcal{P}_n := \{I, X, Y, Z\}^{\otimes n}$ . Since the Pauli basis forms an orthonormal Hermitian basis with respect to the normalized Hilbert-Schmidt inner product, we have for all  $P, Q \in \mathcal{P}_n$ ,

$$\langle\langle P|Q\rangle\rangle = \frac{1}{2^n} \text{Tr}[PQ] = \delta_{PQ}. \quad (\text{A3})$$

Given a linear map  $\mathcal{C} : \mathcal{B}(\mathcal{H}_n) \rightarrow \mathcal{B}(\mathcal{H}_n)$  we denote the transition amplitude associated to two Pauli operators

$P, Q \in \mathcal{P}_n$  as follows:

$$\langle\langle P|\widehat{\mathcal{C}}|Q\rangle\rangle = \langle\langle P|\mathcal{C}(Q)\rangle\rangle = \langle\langle \mathcal{C}^\dagger(P)|Q\rangle\rangle, \quad (\text{A4})$$

then, note that  $\mathcal{C}$  can be completely described by the associated transition amplitudes. For any Pauli operator  $P \in \mathcal{P}_n$  we can write

$$\mathcal{C}(P) = \sum_{Q \in \mathcal{P}_n} \langle\langle Q|\widehat{\mathcal{C}}|P\rangle\rangle Q. \quad (\text{A5})$$

By linearity, for any operator  $O \in \mathcal{B}(\mathcal{H}_n)$  we have

$$\mathcal{C}(O) = \sum_{P, Q \in \mathcal{P}_n} \langle\langle O|P\rangle\rangle \langle\langle Q|\widehat{\mathcal{C}}|P\rangle\rangle Q. \quad (\text{A6})$$

Equivalently, the transition amplitudes  $\langle\langle Q|\widehat{\mathcal{C}}|P\rangle\rangle$  can be expressed as a  $4^n \times 4^n$  matrix, which is usually referred as the *Pauli transfer matrix* of the channel.

**Support and weight of Pauli operators.** Given a Pauli operator  $P = P^{(1)} \otimes P^{(2)} \otimes \dots \otimes P^{(n)} \in \mathcal{P}_n$ , we define its support, denoted as  $\text{supp}(P)$ , as the subset of qubits upon which  $P$  acts nontrivially, i.e.,

$$\text{supp}(P) := \{i \mid P^{(i)} \neq I\}, \quad (\text{A7})$$

and we use  $|P|$  to denote the Pauli weight of  $P$ , which is defined as the number of nonidentity single-qubit Pauli operators in the Pauli tensor decomposition of  $P$ , i.e.,  $|P| = |\text{supp}(P)|$ .

**Matrix norms.** For a matrix  $A \in \mathcal{B}(\mathcal{H}_n)$ , the *Schatten  $p$ -norm* is defined as  $\|A\|_p := (\text{Tr}[(\sqrt{A^\dagger A})^p])^{1/p}$ , representing the  $p$ -norm of the matrix's singular values. Closely related to the Schatten  $p$ -norm are the trace norm ( $\|\cdot\|_1$ ) and the Hilbert-Schmidt norm ( $\|\cdot\|_2$ ). Then, we recall that the infinity norm ( $\|\cdot\|_\infty$ ) of a matrix corresponds to its largest singular value and can also be obtained as the limit of the Schatten  $p$ -norm when  $p \rightarrow \infty$ .

We remind that an operator  $O \in \mathcal{B}(\mathcal{H}_n)$  is Hermitian if and only if it can be expressed as a linear combination of Pauli operators with real coefficients, i.e.,

$$O = \sum_{P \in \mathcal{P}_n} a_P P, \quad \text{where } a_P \in \mathbb{R} \text{ for all } P \in \mathcal{P}_n. \quad (\text{A8})$$

Given a Hermitian operator  $O$ , we define its squared *normalized Frobenius norm*  $\|O\|_F^2$  as

$$\|O\|_F^2 := \frac{\|O\|_2^2}{2^n} = \frac{1}{4^n} \sum_{P \in \mathcal{P}_n} \text{Tr}[OP]^2 = \sum_{P \in \mathcal{P}_n} a_P^2, \quad (\text{A9})$$

which is also referred as normalized Hilbert-Schmidt norm or squared Pauli 2-norm. We will also use the following

useful relation:

$$\|O\|_2^2 = \text{Tr}[\mathbb{F}O^{\otimes 2}], \quad (\text{A10})$$

where  $\mathbb{F}$  is the *flip operator*, also known as SWAP operator, defined as

$$\mathbb{F} := \sum_{i,j \in \{0,1\}^n} |i,j\rangle\langle j,i|. \quad (\text{A11})$$

## 2. Ensembles of linear maps and states

In this section, we define several ensembles of linear maps that will serve as essential tools throughout this work. These ensembles enable us to analyze the limitations of noisy random circuits without relying on the global [45] or local [37] unitary 2-design assumption commonly adopted in previous works on noisy random circuits.

We start by recalling the definition of the Haar measure, which formalizes the notion of uniform distribution over unitaries. For a more detailed explanation we refer to Ref. [66].

*Definition A1 (Haar measure and  $t$ -designs).* The Haar measure  $\mu_H$  is the (unique) probability distribution over the unitary group  $\mathbb{U}(2^n)$  which is left and right-invariant, which means that for any integrable function  $f$ , we have

$$\mathbb{E}_{U \sim \mu_H} [f(U)] = \mathbb{E}_{U \sim \mu_H} [f(UV)] = \mathbb{E}_{U \sim \mu_H} [f(VU)], \quad (\text{A12})$$

for any  $U, V \in \mathbb{U}(2^n)$ . Furthermore, a probability distribution  $\mathcal{D}$  over the unitary group  $\mathbb{U}(d)$  is a unitary  $t$ -design if it matches the moments of the Haar measure up to the  $t$ th order, that is

$$\mathbb{E}_{U \sim \mathcal{D}} [U^{\otimes t} O U^{\dagger \otimes t}] = \mathbb{E}_{U \sim \mu_H} [U^{\otimes t} O U^{\dagger \otimes t}], \quad (\text{A13})$$

for all linear operators  $O \in \mathcal{B}(\mathcal{H}_n)$ .

Equivalently, in the vectorized notation, the previous definition reads as follows:

$$\mathbb{E}_{U \sim \mathcal{D}} [U^{\otimes t} \otimes U^{*\otimes t}] = \mathbb{E}_{U \sim \mu_H} [U^{\otimes t} \otimes U^{*\otimes t}]. \quad (\text{A14})$$

We further provide some examples of unitary 1- and 2-designs.

*Example A1 (Unitary 1- and 2-designs).* A distribution  $\mathcal{D}$  over  $\mathbb{U}(2^n)$  is a unitary 1-design if, for all linear operators  $O \in \mathcal{B}(\mathcal{H}_n)$ , it satisfies

$$\mathbb{E}_{U \sim \mathcal{D}} [U O U^\dagger] = \mathbb{E}_{U \sim \mu_H} [U O U^\dagger] = \frac{I^{\otimes n}}{2^n} \text{Tr}[O]. \quad (\text{A15})$$

An example of unitary 1-design is the uniform distribution over the Pauli basis  $\mathcal{P}_n$ , as we have

$$\frac{1}{4^n} \sum_{P \in \mathcal{P}_n} [POP] = \frac{I^{\otimes n}}{2^n} \text{Tr}[O]. \quad (\text{A16})$$

The above identity is also commonly referred to as *Pauli twirling*. It is also noteworthy that the Haar measure over the orthogonal group  $\mathbb{O}(2^n)$ , which we denote as  $\mu_H(\mathbb{O}(2^n))$ , also forms a unitary 1-design.

A distribution  $\mathcal{D}$  over  $\mathbb{U}(2^n)$  is a unitary 2-design if, for all linear operators  $O \in \mathcal{B}(\mathcal{H}_n^{\otimes 2})$ , it satisfies

$$\begin{aligned} \mathbb{E}_{U \sim \mathcal{D}} [U^{\otimes 2} O U^{\dagger \otimes 2}] &= \mathbb{E}_{U \sim \mu_H} [U^{\otimes 2} O U^{\dagger \otimes 2}] \\ &= c_{\mathbb{I},O} \mathbb{I} + c_{\mathbb{F},O} \mathbb{F}, \end{aligned} \quad (\text{A17})$$

where

$$\begin{aligned} c_{\mathbb{I},O} &= \frac{\text{Tr}(O) - 2^{-n} \text{Tr}(\mathbb{F}O)}{4^n - 1} \quad \text{and} \\ c_{\mathbb{F},O} &= \frac{\text{Tr}(\mathbb{F}O) - 2^{-n} \text{Tr}(O)}{4^n - 1}. \end{aligned} \quad (\text{A18})$$

An example of unitary 2-design is the uniform distribution over the Clifford group  $\text{Cl}(2^n)$ .

In the following, we define the class of *locally unbiased* distributions over linear maps. In loose terms, a distribution belongs to this class if it remains invariant under single-qubit unitary 1-designs. We refer to these distributions as “locally unbiased” because the expectation value of a quantum state evolved under a unitary 1-design is the maximally mixed state. This state lies at the center of the Bloch sphere, thus it is not biased toward any specific direction.

*Definition A2 (Locally unbiased distribution).* A distribution  $\mathcal{D}$  over linear maps  $\mathcal{L}_{n \rightarrow n}$  is locally unbiased if it is invariant under right-multiplication of single-qubit unitary 1-designs up to the second moment, i.e., if there exist some distribution  $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$  over  $\mathbb{U}(2)$  such that (i) all  $\mathcal{D}_i$  are unitary 1-designs, and (ii) the following identity holds:

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\widehat{\mathcal{C}^{\otimes 2}}] = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \mathbb{E}_{V \sim \bigotimes_{i=1}^n \mathcal{D}_i} [\widehat{\mathcal{C}^{\otimes 2}} (V^{\otimes 2} \otimes V^{*\otimes 2})]. \quad (\text{A19})$$

Locally unbiased distributions generalize the notion of *Pauli invariance* previously proposed in Ref. [13], in the sense that Pauli invariance implies local unbiasedness. Pauli invariance is also a natural structural assumption in

the context of classical shadow tomography. In particular, Ref. [67] studied classical shadows generated by Pauli-invariant unitary ensembles and showed that, under this assumption, the shadow channel is diagonal in the Pauli basis and the corresponding reconstruction map admits an explicit formula, thereby clarifying how Pauli invariance facilitates inversion of the measurement channel.

*Definition A3 (Pauli-invariant distribution).* A distribution  $\mathcal{D}$  over linear maps  $\mathcal{L}_{n \rightarrow n}$  is Pauli-invariant if it is invariant under right-multiplication of random Pauli operators up to the second moment, i.e., if

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\widehat{\mathcal{C}^{\otimes 2}}] = \frac{1}{4^n} \sum_{P \in \mathcal{P}_n} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\widehat{\mathcal{C}^{\otimes 2}} (P^{\otimes 2} \otimes P^{*\otimes 2})]. \quad (\text{A20})$$

The following technical lemma plays a central role in the analysis of noisy random circuits.

*Supplemental Lemma A1 (Orthogonality).* Let  $\mathcal{D}$  be a distribution over linear maps  $\mathcal{L}_{n \rightarrow n}$ .

- (1) If  $\mathcal{D}$  is locally unbiased, then for all  $P, Q \in \mathcal{P}_n$  such that  $\text{supp}(P) \neq \text{supp}(Q)$ , we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (P \otimes Q)] = 0. \quad (\text{A21})$$

for some unitary 1-designs  $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$ . Let  $P, Q \in \mathcal{P}_n$  be two Pauli operators which do not share the same support. Consider their tensor decompositions  $P = P_1 \otimes P_2 \otimes \dots \otimes P_n$  and  $Q = Q_1 \otimes Q_2 \otimes \dots \otimes Q_n$ . As  $\text{supp}(P) \neq \text{supp}(Q)$ , there exists an index  $j \in [n]$  such that  $P_j \neq I$  and  $Q_j = I$ . From the previous, we have that

$$\mathbb{E}_{V_j \sim \mathcal{D}_j} [V_j^{\dagger \otimes 2} (P_j \otimes Q_j) V_j^{\otimes 2}] = \mathbb{E}_{V_j \sim \mathcal{D}_j} [V_j^\dagger P_j V_j] \otimes I = 0, \quad (\text{A26})$$

where we used the fact that  $\mathcal{D}_j$  is a unitary 1-design and unitaries are trace-preserving. Then, the desired result easily follows:

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (P \otimes Q)] \quad (\text{A27})$$

$$= \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (\mathbb{E}_{V \sim \bigotimes_{i=1}^n \mathcal{D}_i} V^{\dagger \otimes 2} (P \otimes Q) V^{\otimes 2})] \quad (\text{A28})$$

$$= \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left[ \mathcal{C}^{\otimes 2} \left( \left( \mathbb{E}_{V_j \sim \mathcal{D}_j} V_j^{\otimes 2} (P_j \otimes Q_j) V_j^{\dagger \otimes 2} \right) \otimes \left( \mathbb{E}_{V_1 \sim \mathcal{D}_1, \dots, V_{j-1} \sim \mathcal{D}_{j-1}, V_{j+1} \sim \mathcal{D}_{j+1}, \dots, V_n \sim \mathcal{D}_n} \bigotimes_{i \neq j} (V_i^{\otimes 2} (P_i \otimes Q_i) V_i^{\dagger \otimes 2}) \right) \right) \right] = 0, \quad (\text{A29})$$

where in the last step we plugged Eq. (A26). Then Eq. (A22) readily follows:

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (O^{\otimes 2})] = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left( \mathcal{C} \left( \sum_{P \in \mathcal{P}_n} \langle O | P \rangle P \right) \right)^{\otimes 2} = \sum_{P, Q \in \mathcal{P}_n} \langle O | P \rangle \langle O | Q \rangle \underbrace{\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (P \otimes Q)]}_{=0 \text{ if } \text{supp}(P) \neq \text{supp}(Q)} \quad (\text{A30})$$

$$= \sum_{A \subseteq [n]} \left( \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P) = A}} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \langle O | P \rangle \mathcal{C}(P) \right)^{\otimes 2}. \quad (\text{A31})$$

In particular, for all observables  $O$ , we have

$$\begin{aligned} & \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \mathcal{C}^{\otimes 2} (O^{\otimes 2}) \\ &= \sum_{A \subseteq [n]} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left( \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P) = A}} \langle O | P \rangle \mathcal{C}(P) \right)^{\otimes 2}. \end{aligned} \quad (\text{A22})$$

- (2)  $\mathcal{D}$  is Pauli-invariant if and only if, for all  $P, Q \in \mathcal{P}_n$  such that  $P \neq Q$ , we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (P \otimes Q)] = 0. \quad (\text{A23})$$

In particular, for all observables  $O$ , we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \mathcal{C}^{\otimes 2} (O^{\otimes 2}) = \sum_{P \in \mathcal{P}_n} \langle O | P \rangle^2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \mathcal{C}^{\otimes 2} (P^{\otimes 2}). \quad (\text{A24})$$

*Proof.* We start by proving the first statement. Recall that, by Definition A2, we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\widehat{\mathcal{C}^{\otimes 2}}] = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \mathbb{E}_{V \sim \bigotimes_{i=1}^n \mathcal{D}_i} [\widehat{\mathcal{C}^{\otimes 2}} V^{\otimes 2} \otimes V^{*\otimes 2}], \quad (\text{A25})$$

Regarding the second statement, Lemma 2 in Ref. [13] established the result for the case where  $\mathcal{D}$  is a distribution over  $\mathbb{U}(4)$ . Here, we demonstrate that this argument can be naturally extended to linear maps  $\mathcal{L}_{n \rightarrow n}$ . First, let us use the invariance under right-multiplication of random Pauli,

$$\begin{aligned} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (P \otimes Q)] \\ = \frac{1}{4^n} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left[ \mathcal{C}^{\otimes 2} \left( \sum_{R \in \mathcal{P}_n} RPR \otimes RQR \right) \right]. \end{aligned} \quad (\text{A32})$$

From here we can see that it suffices to show that

$$\sum_{R \in \mathcal{P}_n} RPR \otimes RQR = 0. \quad (\text{A33})$$

Let  $\langle P, Q \rangle := 1[P \text{ and } Q \text{ anticommute}]$ , i.e.,  $\langle P, Q \rangle$  is the indicator function which equals 1 if  $\{P, Q\} = 0$  and 0 if  $[P, Q] = 0$ . We have

$$\begin{aligned} \sum_{R \in \mathcal{P}_n} RPR \otimes RQR &= \sum_{R \in \mathcal{P}_n} (-1)^{\langle P, R \rangle + \langle Q, R \rangle} P \otimes Q \\ &= \sum_{R \in \mathcal{P}_n} (-1)^{\langle PQ, R \rangle} P \otimes Q = 0, \end{aligned} \quad (\text{A34})$$

where the last line follows from the fact that  $PQ$  is not identity, and therefore commutes with half Paulis and anticommutes with the other half.

We now prove the other direction, i.e., that  $\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (P \otimes Q)] = 0$  implies the Pauli invariance property. Let  $O$  be an arbitrary Hermitian operator. We have

$$\frac{1}{4^n} \sum_{R \in \mathcal{P}_n} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \mathcal{C}^{\otimes 2} (R^{\otimes 2} O R^{\otimes 2}) = \frac{1}{4^n} \sum_{R \in \mathcal{P}_n} \sum_{P \in \mathcal{P}_{2n}} \langle O | P \rangle \underbrace{\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (R^{\otimes 2} P R^{\otimes 2})]}_{=0 \text{ unless } P = Q^{\otimes 2} \text{ with } Q \in \mathcal{P}_n} \quad (\text{A35})$$

$$= \frac{1}{4^n} \sum_{Q, R \in \mathcal{P}_n} \langle O | Q^{\otimes 2} \rangle \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (Q^{\otimes 2})] (-1)^{2\langle Q, R \rangle} \quad (\text{A36})$$

$$= \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \mathcal{C}^{\otimes 2} (O), \quad (\text{A37})$$

where in the last step we used that  $(-1)^{2\langle Q, R \rangle} = 1$ . Equation (A24) can also be proven analogously

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (O^{\otimes 2})] = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left( \mathcal{C} \left( \sum_{P \in \mathcal{P}_n} \langle O | P \rangle P \right) \right)^{\otimes 2} \quad (\text{A38})$$

$$= \sum_{P, Q \in \mathcal{P}_n} \langle O | P \rangle \langle O | Q \rangle \underbrace{\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} [\mathcal{C}^{\otimes 2} (P \otimes Q)]}_{=0 \text{ if } P \neq Q} \quad (\text{A39})$$

$$= \sum_{P \in \mathcal{P}_n} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \langle O | P \rangle^2 \mathcal{C}^{\otimes 2} (P^{\otimes 2}). \quad (\text{A40})$$

■

In the following, we provide an example of a distribution over  $\mathbb{U}(2)$  that constitutes a unitary 1-design but does not exhibit Pauli invariance.

*Example A2 (A class of unitary 1-designs which are not Pauli-invariant).* Given a parameter  $\theta \in [0, 2\pi)$ , let  $\mathcal{E}$  be the uniform distribution over the following ensemble:

$$\left\{ X, Y, \exp\left(-i\frac{\theta}{2}Z\right), \exp\left(-i\left(\frac{\theta+\pi}{2}\right)Z\right) \right\}. \quad (\text{A41})$$

For  $\theta = 0, \pi$ , this ensemble coincides with the single-qubit Pauli basis. We can easily verify that  $\mathcal{E}$  forms a 1-design. However,  $\mathcal{E}$  is not Pauli invariant, as we have

$$\mathbb{E}_{U \sim \mathcal{E}} [U^{\otimes 2} (X \otimes Y) U^{\dagger \otimes 2}] \quad (\text{A42})$$

$$\begin{aligned} &= -\frac{1}{2} (X \otimes Y) + \frac{1}{2} (\cos^2(\theta) X \otimes Y + \sin^2(\theta) Y \otimes X \\ &\quad + \cos(\theta) \sin(\theta) (X \otimes X - Y \otimes Y)), \end{aligned} \quad (\text{A43})$$

which equals 0 if and only if  $\theta = 0$  or  $\pi$ .

We also introduce the class of approximately locally scrambling unitary distributions, which will be particularly useful for analyzing noise within the dephasing class.

*Definition A4 (Approximate local scrambling).* Let  $\eta \in [0, 1)$ . A distribution  $\mathcal{D}$  over  $\mathbb{U}(2)$  is an  $\eta$ -approximate scrambler if it satisfies the following properties:

$$\forall P, Q \in \{I, X, Y, Z\}, P \neq Q: \quad \mathbb{E}_{U \sim \mathcal{D}} [U^{\dagger \otimes 2} (P \otimes Q) U^{\otimes 2}] = 0 \quad (\text{orthogonality}), \quad (\text{A44})$$

$$\max_{P, Q \in \{X, Y, Z\}} \mathbb{E}_{U \sim \mathcal{D}} \langle \langle P | U \otimes U^* | Q \rangle \rangle^2 \leq \frac{1}{3} (1 + 2\eta) \quad (\text{approximate Pauli-mixing}). \quad (\text{A45})$$

A distribution  $\mathcal{D}$  over  $\mathbb{U}(2^n)$  is an  $\eta$ -approximate local scrambler if there exists a distribution  $\mathcal{D}'$  over  $\mathbb{U}(2^n)$  and some distribution  $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n$  over  $\mathbb{U}(2)$  such that (i) all  $\mathcal{D}_i$  are  $\eta$ -approximate scrambler, and (ii) the following identity holds:

$$\mathbb{E}_{U \sim \mathcal{D}} [U^{\otimes 2} \otimes U^{*\otimes 2}] = \mathbb{E}_{U \sim \mathcal{D}'} \mathbb{E}_{V_1 \sim \mathcal{D}_1, V_2 \sim \mathcal{D}_2, \dots, V_n \sim \mathcal{D}_n} \left[ U^{\otimes 2} \bigotimes_{i=1}^n V_i^{\otimes 2} \otimes U^{*\otimes 2} \bigotimes_{i=1}^n V_i^{*\otimes 2} \right], \quad (\text{A46})$$

where  $V_1, V_2, \dots, V_n$  are sampled independently.

This notion generalizes that of locally scrambling distributions [42,43], which have found fruitful application in previous literature on the classical simulation of random noiseless circuits [39]. In particular, setting  $\eta = 0$  and  $\mathcal{D} = \mathcal{D}'$ , we recover the definition of locally scrambling distribution (up to the second moment) proposed in the previous literature [42,43,68,69].

We further note that the notion of approximate local scrambling is closely related to that of max-relative entropy, a connection we will leverage in our subsequent technical analysis.

*Observation A1 (Entropic interpretation of approximate scrambling).* Let  $\mathcal{D}$  be a single-qubit  $\eta$ -approximate scrambler and let  $P \in \{X, Y, Z\}$ . By Eq. (A44), we have

$$\mathbb{E}_{U \sim \mathcal{D}} [U^{\dagger \otimes 2} P U^{\otimes 2}] = \sum_{Q \in \{X, Y, Z\}} a_P^2 P^{\otimes 2}. \quad (\text{A47})$$

Moreover, the vector  $\mathbf{a} \in \mathbb{R}^3$  satisfies

$$\|\mathbf{a}\|_2^2 = 1 \quad \text{and} \quad \|\mathbf{a}\|_\infty^2 \leq \frac{1 + 2\eta}{3}, \quad (\text{A48})$$

where the identity follows from the unitarily invariance of the Frobenius norm and the inequality follows from Eq. (A45). Therefore, the vector  $p_{\mathbf{a}} = (a_X^2, a_Y^2, a_Z^2)$  can be interpreted as a probability distribution with min-entropy bounded from below as

$$H_{\min}(p_{\mathbf{a}}) := -\log \left( \max_{P \in \{X, Y, Z\}} a_P^2 \right) \geq \log(3) - \log(1 + 2\eta). \quad (\text{A49})$$

Equivalently, the min-entropy can be expressed in terms of the max-relative entropy between  $p_{\mathbf{a}}$  and the uniform distribution  $p_{\mathbf{u}} := (1/3, 1/3, 1/3)$ ,

$$D_{\max}(p_{\mathbf{a}} \| p_{\mathbf{u}}) := -H_{\min}(p_{\mathbf{a}}) + \log(3) \leq \log(1 + 2\eta). \quad (\text{A50})$$

Hinging on connection between approximate scrambling and the max-relative entropy, we can demonstrate that the approximate scrambling property is preserved under unitary evolution. Specifically, if  $\mathcal{D}$  is a single-qubit  $\eta$ -approximate scrambler, then the derived distribution obtained by sampling a unitary operator  $U$  from  $\mathcal{D}$  and applying the transformation  $UV$ , is also a single-qubit  $\eta$ -approximate scrambler, as formalized in the following Lemma.

*Supplemental Lemma A2 (Monotonicity of approximate scrambling).* Let  $\mathcal{D}$  be a single-qubit  $\eta$ -approximate scrambler. Then for all  $V \in \mathbb{U}(2)$  we have

$$\forall P, Q \in \mathcal{P}_1, P \neq Q: \quad \mathbb{E}_{U \sim \mathcal{D}} [(VU)^{\dagger \otimes 2} (P \otimes Q) (UV)^{\otimes 2}] = 0, \quad (\text{A51})$$

$$\max_{P, Q \in \{X, Y, Z\}} \mathbb{E}_{U \sim \mathcal{D}} \langle \langle P | UV \otimes U^* V^* | Q \rangle \rangle^2 \leq \frac{1}{3} (1 + 2\eta). \quad (\text{A52})$$

*Proof.* Equation (A51) follows from the linearity of the expectation:

$$\begin{aligned} & \mathbb{E}_{U \sim \mathcal{D}} [(UV)^{\dagger \otimes 2} (P \otimes Q) (UV)^{\otimes 2}] \\ &= V^{\dagger \otimes 2} \mathbb{E}_{U \sim \mathcal{D}} [U^{\dagger \otimes 2} (P \otimes Q) U^{\otimes 2}] V^{\otimes 2}, \end{aligned} \quad (\text{A53})$$

which equals to zero if  $P \neq Q$  by Eq. (A44). The second part of the Lemma can be proven exploiting the connection between approximate scrambling and the max-relative entropy described in Observation A1. We have

$$\mathbb{E}_{U \sim \mathcal{D}} [U^{\dagger \otimes 2} P^{\otimes 2} U^{\otimes 2}] = \sum_{Q \in \{X, Y, Z\}} a_Q^2 Q^{\otimes 2} \text{ where } \|a\|_2^2 = 1, \quad (\text{A54})$$

$$\begin{aligned} \mathbb{E}_{U \sim \mathcal{D}} [V^{\dagger \otimes 2} U^{\dagger \otimes 2} P^{\otimes 2} U^{\otimes 2} V^{\otimes 2}] \\ = \sum_{Q, R \in \{X, Y, Z\}} a_Q^2 \langle Q | V \otimes V^* | R \rangle^2 R^{\otimes 2}. \end{aligned} \quad (\text{A55})$$

We further observe that, for all  $Q \in \{X, Y, Z\}$  we have

$$\sum_{R \in \{X, Y, Z\}} \langle Q | V \otimes V^* | R \rangle^2 = \|V^{\dagger} Q V\|_F^2 = \|Q\|_F^2 = 1, \quad (\text{A56})$$

where in the first step we used the fact that  $\langle Q | V \otimes V^* | I \rangle = 0$  as  $V$  is trace-preserving, and in the second step we used the unitarily invariance of the Frobenius norm. Thus, the squared amplitudes  $\langle Q | V | R \rangle^2$  sum up to 1 and can be interpreted as transition probabilities. Specifically, we introduce the stochastic channel  $\mathcal{K} : \{X, Y, Z\} \rightarrow \{X, Y, Z\}$  defined as follows:

$$\Pr[\mathcal{K}(Q) = R] = \langle Q | V | R \rangle^2. \quad (\text{A57})$$

We can represent the action of  $\mathcal{K}$  on the distribution  $p_a = (a_X^2, a_Y^2, a_Z^2)$  as

$$\mathcal{K}(p_a) = p_b := (b_X^2, b_Y^2, b_Z^2), \quad (\text{A58})$$

$$\text{where } b_R^2 := \sum_{Q \in \{X, Y, Z\}} a_Q^2 \langle Q | V | R \rangle^2. \quad (\text{A59})$$

We can see that the uniform distribution  $p_u = (1/3, 1/3, 1/3)$  is invariant under  $\mathcal{K}$ , i.e.,

$$\mathcal{K}(p_u) = p_u. \quad (\text{A60})$$

This is a consequence of the fact that the operator  $\frac{1}{3}(X^{\otimes 2} + Y^{\otimes 2} + Z^{\otimes 2})$  is invariant under the unitary evolution  $V^{\dagger \otimes 2}(\cdot)V^{\otimes 2}$ . Then, by the data-processing inequality, we have

$$D_{\max}(p_b \| p_u) = D_{\max}(\mathcal{K}(p_a) \| \mathcal{K}(p_u)) \quad (\text{A61})$$

$$\leq D_{\max}(p_a \| p_u) \leq \log(1 + 2\eta), \quad (\text{A62})$$

which implies Eq. (A52).  $\blacksquare$

It is easy to see that any 2-designs over  $\mathbb{U}(2)$  is also a single-qubit 0-approximate scrambler. Notably, this

include the uniform distribution over the single-qubit Clifford group discussed in Example A1. However, we will demonstrate below that there are other practically relevant distributions over  $\mathbb{U}(2)$  that are  $\eta$ -approximate scramblers with  $\eta < 1$ , highlighting the broad applicability of our analysis.

*Example A3 (Pauli rotations along 2 orthogonal axes).* For simplicity, we consider two rotations along the  $X$  and  $Y$  axes, but our conclusions hold for any choice of the orthogonal axes. Assume that  $\phi$  and  $\theta$  are sampled independently from the uniform distribution over  $[0, 2\pi)$ . We can easily verify that this distribution is Pauli invariant, as we have for all  $P, Q \in \mathcal{P}_1$  such that  $P \neq Q$ ,

$$\mathbb{E}_{\theta, \phi} (R_X(\phi) R_Z(\theta))^{\otimes 2} (P \otimes Q) (R_Z(\theta)^{\dagger} R_X(\phi)^{\dagger})^{\otimes 2} = 0. \quad (\text{A63})$$

Moreover, this distribution is also an  $\eta$ -approximate scrambler with  $\eta = 1/4$ , as we have

$$\mathbb{E}_{\theta, \phi} (R_Z(\theta)^{\dagger} R_X(\phi)^{\dagger} X R_X(\phi) R_Z(\theta))^{\otimes 2} = \frac{1}{2} (X^{\otimes 2} + Y^{\otimes 2}), \quad (\text{A64})$$

$$\begin{aligned} \mathbb{E}_{\theta, \phi} (R_Z(\theta)^{\dagger} R_X(\phi)^{\dagger} Y R_X(\phi) R_Z(\theta))^{\otimes 2} \\ = \frac{1}{4} (X^{\otimes 2} + Y^{\otimes 2}) + \frac{1}{2} Z^{\otimes 2}, \end{aligned} \quad (\text{A65})$$

$$\begin{aligned} \mathbb{E}_{\theta, \phi} (R_Z(\theta)^{\dagger} R_X(\phi)^{\dagger} Z R_X(\phi) R_Z(\theta))^{\otimes 2} \\ = \frac{1}{4} (X^{\otimes 2} + Y^{\otimes 2}) + \frac{1}{2} Z^{\otimes 2}. \end{aligned} \quad (\text{A66})$$

In all the three cases above, the largest coefficient is  $1/2$ , which implies that the considered distribution is a  $(1/4)$ -approximate scrambler.

On the other hand, we also note that uniform Pauli rotations along a single axis are neither  $\eta$ -approximate scramblers for  $\eta < 1$ , nor do they form a unitary 1-design.

*Example A4 (Rotations along a single axis).* We can further consider the case of a uniformly random Pauli rotation along a single axis [e.g.,  $R_Z(\theta)$ ]. This distribution fails to be a 1-design or approximately locally scrambling, as we have  $R_Z(\theta) Z R_Z(\theta)^{\dagger} = Z$  for all  $\theta$ . In particular, this implies

$$\mathbb{E}_{\theta} R_Z(\theta) Z R_Z(\theta)^{\dagger} = Z, \quad (\text{A67})$$

$$\mathbb{E}_{\theta} (R_Z(\theta)^{\dagger} Z R_Z(\theta))^{\otimes 2} = Z^{\otimes 2}. \quad (\text{A68})$$

We conclude this section by defining a class of distributions over quantum states, previously introduced in Ref. [16]. This definition generalizes that of state 1-design.

*Definition A5 (Low-average ensemble, Ref. [16]).* A distribution  $\mathcal{D}$  over quantum states is a low-average distribution with purity  $c$  if it satisfies

$$\|\mathbb{E}_{\rho \sim \mathcal{D}}[\rho]\|_{\infty} \leq \frac{c}{2^n}. \quad (\text{A69})$$

In particular, if  $\mathcal{D}$  is a distribution over pure states and  $c = 1$ , then we recover the definition of state 1-design. More generally, if  $\mathcal{D}$  is unitary 1-design, then we have for any state  $\rho$ ,

$$\|\mathbb{E}_{U \sim \mathcal{D}}[U\rho U^{\dagger}]\|_{\infty} = \left\| \frac{I^{\otimes n}}{2^n} \right\|_{\infty} = \frac{1}{2^n}, \quad (\text{A70})$$

i.e.,  $U\rho U^{\dagger}$  is sampled from a low-average ensemble with purity 1.

The second moment of the expectation value of an observable  $O$ , when measured on a state sampled from a low-average ensemble, can be upper bounded as follows.

*Supplemental Lemma A3 (Lemma 3 in Ref. [16]).* Let  $O$  be an observable and  $\mathcal{D}$  be a low-average ensemble over quantum states with purity  $c$ , i.e.  $\|\mathbb{E}_{\rho \sim \mathcal{D}}[\rho]\|_{\infty} \leq c/2^n$ . We have

$$\mathbb{E}_{\rho \sim \mathcal{D}} \text{Tr}[O\rho]^2 \leq c \|O\|_F^2. \quad (\text{A71})$$

*Proof.* We provide an elementary proof of this fact. We have

$$\begin{aligned} \mathbb{E}_{\rho \sim \mathcal{D}} \text{Tr}[O\rho]^2 &\leq \mathbb{E}_{\rho \sim \mathcal{D}} \text{Tr}[O^2 \rho] \leq \|O^2\|_1 \|\mathbb{E}_{\rho \sim \mathcal{D}} \rho\|_{\infty} \\ &= \|O\|_2^2 \|\mathbb{E}_{\rho \sim \mathcal{D}} \rho\|_{\infty} \leq \frac{c}{2^n} \|O\|_2^2. \end{aligned} \quad (\text{A72})$$

In the first step, we used the well-known fact that the variance of a quantum observable is non-negative. In the second step, we applied Hölder's inequality. In the third step, we used the property that the Schatten  $p$ -norm of an observable corresponds to the  $p$ -norm of its eigenvalue vector. Finally, in the last step, we used the assumption on the ensemble  $\mathcal{D}$ . ■

### 3. Noise channels

In this work, we study quantum circuits interspersed by local noise. To this end, we model noise with channels of form  $\mathcal{N}^{\otimes n}$ , where  $\mathcal{N}$  is an arbitrary single-qubit channel. In general, a single-qubit channel  $\mathcal{N}$  can be fully characterized by its action on the Pauli matrices  $\mathcal{P}_1 = \{I, X, Y, Z\}$ . Thus,  $\mathcal{N}$  can be identified by 16 transition amplitudes of the form  $\langle\langle P | \hat{\mathcal{N}} | Q \rangle\rangle$ , where  $P, Q \in \mathcal{P}_1$ . Since  $\mathcal{N}$  is trace-preserving, it satisfies  $\langle\langle I | \hat{\mathcal{N}} | Q \rangle\rangle = 0$  for all  $Q \in \{X, Y, Z\}$  and  $\langle\langle I | \hat{\mathcal{N}} | I \rangle\rangle = 1$ . Consequently, there are 12 remaining transition amplitudes whose values must be determined.

Moreover, as shown in Refs. [18,37], any single-qubit channel can be decomposed into unitary and non-unitary components. For the purposes of the present analysis, we focus exclusively on the non-unitary component, which can be fully described using only 6 parameters. In this context, we state a useful lemma from Ref. [37].

*Supplemental Lemma A4 (Normal form of a quantum channel [18,70]).* Any single-qubit quantum channel  $\mathcal{N}$  can be written in the so-called “normal” form

$$\mathcal{N}(\cdot) = U\mathcal{N}'(V^{\dagger}(\cdot)V)U^{\dagger}, \quad (\text{A73})$$

where  $U, V$  are unitaries and  $\mathcal{N}'(\cdot)$  is a quantum channel defined by the following transition amplitudes:

$$\langle\langle P | \hat{\mathcal{N}}' | Q \rangle\rangle := \begin{cases} 1 & \text{if } P = Q = I, \\ D_P & \text{if } P = Q \neq I, \\ t_P & \text{if } P \neq I \text{ and } Q = I, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A74})$$

where  $\mathbf{t} := (t_X, t_Y, t_Z)$  and  $\mathbf{D} := (D_X, D_Y, D_Z) \in \mathbb{R}^3$ , such that the entries of  $\mathbf{D}$  have all the same sign.

In the following, we present an exhaustive classification of single-qubit (noise) channels.

- (1) **Unitary noise.** A noise channel  $\mathcal{N}$  is unitary, or coherent, if  $\mathcal{N} = U(\cdot)U^{\dagger}$  for some unitary  $U \in \mathbb{U}(2)$ . As our analysis hinges on the contraction properties of non-unitary channels, unitary noise is beyond the scope of the present work.
- (2) **Non-unitary unital noise.** A quantum channel  $\mathcal{N}$  is *unital* if it preserves the identity, i.e., if it satisfies  $\mathcal{N}(I) = I$ . Let the normal form of the channel be  $\mathcal{N}(\cdot) = U\mathcal{N}'(V^{\dagger}(\cdot)V)U^{\dagger}$ . As  $U(\cdot)U^{\dagger}$  and  $V(\cdot)V^{\dagger}$  are unital, then  $\mathcal{N}'$  must be unital as well. By unitality, we have that  $\mathbf{t} = (0, 0, 0)$ . Moreover, we observe that  $\mathbf{D} \neq (1, 1, 1)$ , otherwise the  $\mathcal{N}'$  would be the identity channel and  $\mathcal{N}$  would be unitary. We can further classify non-unitary unital noise in two classes:
  - (a) **Depolarizing class.** A channel  $\mathcal{N}$  belongs to the depolarizing class if all the entries  $D_X, D_Y$  and  $D_Z$  are strictly smaller than 1. As a consequence, the limit of repeatedly applying  $\mathcal{N}$  on any input state is the maximally mixed state, which is also the only fixed point of the noise. Notably, the depolarizing class includes the depolarizing channel, which corresponds to the case  $D_X = D_Y = D_Z = 1 - p$ , where  $p \in (0, 1]$  is the noise rate of the channel.
  - (b) **Dephasing class.** A channel  $\mathcal{N}$  belongs to the dephasing class if exactly one entry of the vector  $\mathbf{D}$  is one and the other two are strictly smaller

than one. This implies that the limit of repeatedly applying  $\mathcal{N}$  on any input state lies on a diameter of the Bloch sphere. The dephasing class includes the dephasing noise, which satisfy  $D_Z = 1$  and  $D_X = D_Y = (1 - 2p)$ , where  $p \in (0, 1]$  is the noise rate of the channel.

All state  $\rho$  of the form  $\rho = (I + aZ)/2$  for  $a \in [-1, 1]$  is a fixed point of the dephasing noise.

Furthermore, since there is no single-qubit channel with exactly two entries in the vector  $\mathbf{D}$  strictly less than 0, then any non-unitary, unital noise must belong to either the depolarizing class or the dephasing class.

- (3) **Non-unital noise.** A channel  $\mathcal{N}$  is non-unital if  $\mathcal{N}(I) \neq I$ . Consequently, we have that also  $\mathcal{N}'(I) \neq I$  and  $\mathbf{t} \neq (0, 0, 0)$ . This implies that channel  $\mathcal{N}$  is also non-unitary. The limit of repeatedly applying  $\mathcal{N}$  on any input state leads to a state  $\rho$  which is not the maximally mixed state. Crucially, unlike noise channels in the other two classes, non-unital noise can increase the purity of an operator.

Our analysis extensively relies on the normal form of the channels. As a preliminary step, we derive the following constraint on the parameters of this normal form.

*Supplemental Lemma A5 (Constraint on normal form parameters).* Let  $\mathcal{N}(\cdot) = U\mathcal{N}'(V(\cdot)V^\dagger)U^\dagger$  be an arbitrary single-qubit channel with normal form parameters  $\mathbf{D}, \mathbf{t}$ . Let  $\Upsilon(\cdot, \cdot)$  be the function defined as follows:

$$\Upsilon(\mathbf{D}, \mathbf{t}) := \max_{\|\mathbf{a}\|_2=1} \left\{ \sum_{Q \in \{X, Y, Z\}} a_Q^2 D_Q^2 + \left( \sum_{Q \in \{X, Y, Z\}} a_Q t_Q \right)^2 \right\}. \quad (\text{A75})$$

We have that

$$\Upsilon(\mathbf{D}, \mathbf{t}) \leq 1. \quad (\text{A76})$$

Moreover the inequality is strict provided that  $\|\mathbf{D}\|_\infty^2 \in (0, 1)$  or  $\|\mathbf{t}\|_2^2 \in (0, 1)$ .

*Proof.* Let  $(b_X, b_Y, b_Z)$  be the coefficients maximizing Eq. (A75), i.e.,

$$(b_X, b_Y, b_Z) := \arg \max_{\substack{\mathbf{a} \in \mathbb{R}^3 \\ \|\mathbf{a}\|_2=1}} \sum_{Q \in \{X, Y, Z\}} a_Q^2 D_Q^2 + \left( \sum_{Q \in \{X, Y, Z\}} a_Q t_Q \right)^2. \quad (\text{A77})$$

We define the observable  $O$  and the state  $\rho$  as follows:

$$O = \sum_{P \in \{X, Y, Z\}} |b_P| \cdot \text{sgn}(t_P \cdot D_P) P \quad \text{and} \quad \rho = \frac{I + O}{2}, \quad (\text{A78})$$

where  $\text{sgn}(x)$  is the sign of  $x$ . We observe that the operator norm of  $O$  equals 1. First, we can upper bound it as follows:

$$\begin{aligned} \|O\|_\infty &= \max_{\sigma} |\text{Tr}[O\sigma]| \\ &= \max_{\substack{\mathbf{r} \in \mathbb{R}^3 \\ \|\mathbf{r}\|_2=1}} \left| \sum_{P \in \{X, Y, Z\}} |b_P| r_P \cdot \text{sgn}(t_P \cdot D_P) \right| \\ &\leq \max_{\substack{\mathbf{r} \in \mathbb{R}^3 \\ \|\mathbf{r}\|_2=1}} \|\mathbf{b}\|_2 \|\mathbf{r}\|_2 = 1, \end{aligned} \quad (\text{A79})$$

where we used the Cauchy-Schwarz inequality. Moreover, the inequality is saturated choosing  $\sigma = \rho$ , so  $\|O\|_\infty = 1$ . We can exploit this fact to upper bound Eq. (A77). Recall that the channel  $\mathcal{N}'$  acts as

$$\begin{aligned} \mathcal{N}' \left( \frac{1}{2} \left( I + \sum_{P \in \{X, Y, Z\}} r_P P \right) \right) \\ = \frac{I}{2} + \frac{1}{2} \sum_{P \in \{X, Y, Z\}} (r_P D_P + t_P) P. \end{aligned} \quad (\text{A80})$$

Replacing the coefficients  $r_P$  with the appropriate values for the state  $\rho$ , we obtain that

$$\text{Tr}[O\mathcal{N}'(\rho)] = \frac{1}{2} \text{Tr} \left[ \left( \sum_{P \in \{X, Y, Z\}} |b_P| \cdot \text{sgn}(t_P \cdot D_P) P \right) \mathcal{N} \left( I + \sum_{P \in \{X, Y, Z\}} |b_P| \cdot \text{sgn}(t_P \cdot D_P) P \right) \right] \quad (\text{A81})$$

$$= \sum_{P \in \{X, Y, Z\}} b_P^2 D_P + |b_P t_P| \text{sgn}(D_P). \quad (\text{A82})$$

By Hölder's inequality, we have that

$$1 = \|O\|_\infty \|\rho\|_1 \geq |\text{Tr}[ON'(\rho)]|$$

$$= \left| \sum_{P \in \{X,Y,Z\}} b_P^2 D_P + |b_P t_P| \text{sgn}(D_P) \right| \quad (\text{A83})$$

$$= \sum_{P \in \{X,Y,Z\}} b_P^2 |D_P| + \sum_{P \in \{X,Y,Z\}} |b_P t_P| \quad (\text{A84})$$

$$\geq \sum_{P \in \{X,Y,Z\}} b_P^2 D_P^2 + \left( \sum_{P \in \{X,Y,Z\}} b_P t_P \right)^2, \quad (\text{A85})$$

where the second-to-last step follows from the fact that the coefficients  $D_P$  have all the same sign (cf. Supplemental Lemma A4) and in the last step we used the fact that  $x^2 \leq x$  if  $|x| \leq 1$ .

It remains to determine under which circumstances the last inequality is strict. We observe that

$$\|\mathbf{D}\|_\infty^2 \in (0, 1) \implies \forall P \in \{X, Y, Z\} : D_P^2 < |D_P| \quad (\text{A86})$$

$$\implies \sum_{P \in \{X,Y,Z\}} b_P^2 D_P^2 < \sum_{P \in \{X,Y,Z\}} b_P^2 |D_P|, \quad (\text{A87})$$

where in the first step we used the fact that  $x^2 < x$  if  $|x| < 1$  and in the second step we used the fact that  $\|\mathbf{b}\|_2^2 = 1$ . Moreover, by Cauchy-Schwarz inequality, we have

$$\left( \sum_{P \in \{X,Y,Z\}} b_P t_P \right)^2 \leq \|\mathbf{t}\|_2^2 \|\mathbf{b}\|_2^2 = \|\mathbf{t}\|_2^2, \quad (\text{A88})$$

and therefore

$$\|\mathbf{t}\|_2^2 \in (0, 1) \implies \left( \sum_{P \in \{X,Y,Z\}} b_P t_P \right)^2 < \sum_{P \in \{X,Y,Z\}} |b_P t_P|, \quad (\text{A89})$$

where we used again the fact that  $x^2 < x$  if  $|x| < 1$ . Putting all together, we obtain that

$$\sum_{P \in \{X,Y,Z\}} b_P^2 D_P^2 + \left( \sum_{P \in \{X,Y,Z\}} b_P t_P \right)^2 < 1, \quad (\text{A90})$$

provided that  $\|\mathbf{D}\|_\infty^2 \in (0, 1)$  or  $\|\mathbf{t}\|_2^2 \in (0, 1)$ . ■

#### 4. The Pauli propagation method

We provide a brief introduction to the Pauli Propagation framework for classically simulating quantum circuits. Let a circuit be represented by an  $L$ -layered quantum channel  $\mathcal{C}$ ,

$$\mathcal{C} = \mathcal{C}_L \circ \mathcal{C}_{L-1} \cdots \circ \mathcal{C}_1. \quad (\text{A91})$$

Then, its matrix (vectorized) form is

$$\widehat{\mathcal{C}} = \widehat{\mathcal{C}}_L \widehat{\mathcal{C}}_{L-1} \cdots \widehat{\mathcal{C}}_1. \quad (\text{A92})$$

Given a Pauli path  $\gamma = (P_0, P_1, P_2, \dots, P_L) \in \mathcal{P}_n^{L+1}$ , we denote the associated Fourier coefficients as

$$\Phi_\gamma(\mathcal{C}) := \langle P_L | \widehat{\mathcal{C}}_L | P_{L-1} \rangle \times \langle P_{L-1} | \widehat{\mathcal{C}}_{L-1} | P_{L-2} \rangle$$

$$\times \cdots \times \langle P_1 | \widehat{\mathcal{C}}_1 | P_0 \rangle. \quad (\text{A93})$$

Thus, we can express the Schrödinger-evolved state  $\mathcal{C}(\rho)$  and the Heisenberg-evolved observable  $\mathcal{C}^\dagger(O)$  as sums over Pauli paths,

$$\mathcal{C}(\rho) = \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle \rho | P_0 \rangle \Phi_\gamma(\mathcal{C}) P_L, \quad (\text{A94})$$

$$\mathcal{C}^\dagger(O) = \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle O | P_L \rangle \Phi_\gamma(\mathcal{C}) P_0. \quad (\text{A95})$$

In order to compute such evolutions approximately, we consider a suitable subset of Pauli paths  $\mathcal{A} \subseteq \mathcal{P}_n^{L+1}$  and the associated “truncated linear map” defined as

$$\mathcal{C}_\mathcal{A}(\cdot) := \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{A}} \langle \cdot | P_0 \rangle \Phi_\gamma(\mathcal{C}) P_L. \quad (\text{A96})$$

Note that the adjoint of  $\mathcal{C}_\mathcal{A}$  is

$$\mathcal{C}_\mathcal{A}^\dagger(\cdot) = \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{A}} \langle \cdot | P_L \rangle \Phi_\gamma(\mathcal{C}) P_0. \quad (\text{A97})$$

Then we have

$$\mathcal{C}_\mathcal{A}(\rho) = \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{A}} \langle \rho | P_0 \rangle \Phi_\gamma(\mathcal{C}) P_L, \quad (\text{A98})$$

$$\mathcal{C}_\mathcal{A}^\dagger(O) = \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{A}} \langle O | P_L \rangle \Phi_\gamma(\mathcal{C}) P_0. \quad (\text{A99})$$

##### a. Orthogonality of Pauli paths

In this work, we categorize Pauli paths based on their support and weight, defined as follows.

**Definition A6 (Path support and path weight).** Let  $\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}$  be a Pauli path.

- (1) The support of  $\gamma$  is the vector of the supports of its components,

$$\text{supp}(\gamma) = (\text{supp}(P_0), \text{supp}(P_1), \dots, \text{supp}(P_L)). \quad (\text{A100})$$

- (2) The weight of  $\gamma$  is the sum of the Pauli weight of its components,

$$|\gamma| = \sum_{i=1}^L |P_i|. \quad (\text{A101})$$

We will also make extensive use of the following orthogonality relationships. The key idea is that locally unbiased distributions eliminate all cross-terms between Pauli paths with distinct supports when taking second moments, while Pauli-invariant distributions satisfy an even stronger property, eliminating cross-terms between all distinct Pauli paths.

*Supplemental Lemma A6 (Orthogonality of Pauli paths).* Let  $\mathcal{C} = \mathcal{C}_L \circ \mathcal{C}_{L-1} \circ \dots \circ \mathcal{C}_1$  be an  $L$ -layered quantum channel sampled from a distribution  $\mathcal{D}_{\text{circ}}$ .

- (1) Assume that all layers  $\mathcal{C}_j$  are sampled independently from locally unbiased distributions  $\mathcal{D}_j$ , and moreover  $\mathcal{C}_L^\dagger$  is also sampled from a locally unbiased distribution. Then the Fourier coefficients of paths with different supports are uncorrelated,

$$\begin{aligned} \text{supp}(\gamma) \neq \text{supp}(\gamma') \\ \implies \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C})] = 0. \end{aligned} \quad (\text{A102})$$

- (2) Assume that all layers  $\mathcal{C}_j$  are sampled independently from Pauli-invariant distributions  $\mathcal{D}_j$ , and moreover  $\mathcal{C}_L^\dagger$  is also sampled from a Pauli invariant distribution. Then the Fourier coefficients of different paths are uncorrelated,

$$\gamma \neq \gamma' \implies \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C})] = 0. \quad (\text{A103})$$

*Proof.* We can express the product of the two coefficients  $\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C})$  as follows:

$$\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C}) \quad (\text{A104})$$

$$= \langle P_L | \widehat{\mathcal{C}}_L | P_{L-1} \rangle \langle P'_L | \widehat{\mathcal{C}}_L | P'_{L-1} \rangle \langle P_{L-1} | \widehat{\mathcal{C}}_{L-1} | P_{L-2} \rangle \langle P'_{L-1} | \widehat{\mathcal{C}}_{L-1} | P'_{L-2} \rangle \dots \langle P_1 | \widehat{\mathcal{C}}_1 | P_0 \rangle \langle P'_1 | \widehat{\mathcal{C}}_1 | P'_0 \rangle \quad (\text{A105})$$

$$= \langle P_L \otimes P'_L | \widehat{\mathcal{C}}_L^{\otimes 2} | P_{L-1} \otimes P'_{L-1} \rangle \langle P_{L-1} \otimes P'_{L-1} | \widehat{\mathcal{C}}_{L-1}^{\otimes 2} | P_{L-2} \otimes P'_{L-2} \rangle \dots \langle P_1 \otimes P'_1 | \widehat{\mathcal{C}}_1^{\otimes 2} | P_0 \otimes P'_0 \rangle. \quad (\text{A106})$$

We will prove the two statements separately. Consider two paths  $\gamma, \gamma'$  with different supports. Then there exists an index  $j$  such that  $\text{supp}(P_j) \neq \text{supp}(P'_j)$ . If  $j \neq L$ , we can use the fact that  $\mathcal{C}_{j+1}$  is sampled from a locally unbiased distribution and therefore by Supplemental Lemma A1,

$$\mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} [\mathcal{C}_{j+1}^{\otimes 2} (P_j \otimes P'_j)] = 0. \quad (\text{A107})$$

Therefore, by linearity of the expectation and the trace we have

$$\begin{aligned} \mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} \langle P_{j+1} \otimes P'_{j+1} | \widehat{\mathcal{C}}_{j+1}^{\otimes 2} | P_j \otimes P'_j \rangle \\ = \mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} \frac{1}{4^n} \text{Tr} \left\{ \mathcal{C}_{j+1}^{\otimes 2} (P_j \otimes P'_j) (P_{j+1} \otimes P'_{j+1}) \right\} \end{aligned} \quad (\text{A108})$$

$$\begin{aligned} &= \frac{1}{4^n} \text{Tr} \left\{ \mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} [\mathcal{C}_{j+1}^{\otimes 2} (P_j \otimes P'_j)] (P_{j+1} \otimes P'_{j+1}) \right\} \\ &= 0, \end{aligned} \quad (\text{A109})$$

which implies that  $\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C})] = 0$ . Similarly, if  $j = L$ , we can instead use the fact that  $\mathcal{C}_L^\dagger$  is sampled from a locally unbiased distribution and therefore

$$\mathbb{E}_{\mathcal{C}_L \sim \mathcal{D}_L} [\mathcal{C}_L^{\dagger \otimes 2} (P_L \otimes P'_L)] = 0. \quad (\text{A110})$$

Hence, we can conclude that  $\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C})] = 0$ .

The proof of the second statement is analogous. Consider two different paths  $\gamma, \gamma'$ . Then there exists an index  $j$  such that  $P_j \neq P'_j$ . If  $j \neq L$ , we can use the fact that  $\mathcal{C}_{j+1}$  is sampled from a Pauli-invariant distribution and therefore by Supplemental Lemma A1,

$$\mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} [\mathcal{C}_{j+1}^{\otimes 2} (P_j \otimes P'_j)] = 0. \quad (\text{A111})$$

Thus, employing the linearity of the expectation and the trace as above we can prove that  $\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C})] = 0$ . Similarly, if  $j = L$ , we can instead use the fact

that  $\mathcal{C}_L^\dagger$  is sampled from a Pauli-invariant distribution and therefore

$$\mathbb{E}_{\mathcal{C}_L \sim \mathcal{D}_L} [\mathcal{C}_L^{\dagger \otimes 2} (P_L \otimes P'_L)] = 0, \quad (\text{A112})$$

which also allows us to conclude that  $\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C})] = 0$ . ■

### b. Estimating second moments by sampling Pauli paths

Let  $\mathcal{D}_{\text{circ}}$  be a distribution over  $L$ -layered quantum channels  $\mathcal{C}_L \circ \mathcal{C}_{L-1} \circ \dots \circ \mathcal{C}_1$  such that all the channels  $\mathcal{C}_j$  are sampled from Pauli-invariant distributions, which by Supplemental Lemma A6 implies that the Fourier coefficients of any two distinct Pauli paths are uncorrelated, i.e.,

$$\gamma \neq \gamma' \implies \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C})] = 0. \quad (\text{A113})$$

In this section we will demonstrate that several second moment quantities associated to the distribution  $\mathcal{D}_{\text{circ}}$  can be efficiently estimated with a simple classical randomized algorithm.

Specifically, given an observable  $O$ , we will consider the following general expression:

$$F(\mathcal{D}_{\text{circ}}) := \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle O | P_L \rangle^2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(\mathcal{C})] f(\gamma), \quad (\text{A114})$$

where  $f : \mathcal{P}_n^{L+1} \rightarrow [0, 1]$  is a function of the Pauli paths. It is easy to see that many relevant quantities associated to  $\mathcal{D}_{\text{circ}}$  can be written in this form:

- (1) **Variances of expectation values.** Given an observable  $O$  and a state  $\rho$ , the associated variance can be written as

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\text{Tr}[O\mathcal{C}(\rho)]^2] = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left( \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \Phi_\gamma(\mathcal{C}) \langle O | P_L \rangle \text{Tr}[P_0 \rho] \right)^2 \quad (\text{A115})$$

$$= \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(\mathcal{C})] \langle O | P_L \rangle^2 \text{Tr}[P_0 \rho]^2, \quad (\text{A116})$$

where the second identity follows from the orthogonality relation in Eq. (A113). Setting  $f_{\text{var}}(\gamma) = \text{Tr}[P_0 \rho]^2$ , we find that the variance can be rewritten as in Eq. (A114).

- (2) **Mean squared errors.** Given an observable  $O$ , a state  $\rho$ , a subset of Pauli paths  $\mathcal{A} \subseteq \mathcal{P}_n^{L+1}$  and its complement  $\bar{\mathcal{A}} = \mathcal{P}_n^{L+1} \setminus \mathcal{A}$ , we consider the following mean squared error:

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\text{Tr}[(\mathcal{C}^\dagger(O) - \mathcal{C}_{\mathcal{A}}^\dagger(O)) \rho]^2] = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left( \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \bar{\mathcal{A}}} \Phi_\gamma(\mathcal{C}) \langle O | P_L \rangle \text{Tr}[P_0 \rho] \right)^2 \quad (\text{A117})$$

$$= \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \bar{\mathcal{A}}} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(\mathcal{C})] \langle O | P_L \rangle^2 \text{Tr}[P_0 \rho]^2 \quad (\text{A118})$$

$$= \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(\mathcal{C})] f_{\mathcal{A}}(\gamma), \quad (\text{A119})$$

where we used again orthogonality relation of Pauli paths in Eq. (A113) and we defined the function

$$f_{\mathcal{A}}(\gamma) = \begin{cases} \text{Tr}[P_0 \rho]^2 & \text{if } \gamma \notin \mathcal{A}, \\ 0 & \text{if } \gamma \in \mathcal{A}. \end{cases} \quad (\text{A120})$$

Alternatively, one can evaluate the mean squared error with respect to the normalized Frobenius norm  $\|\mathcal{C}^\dagger(O) - \mathcal{C}_{\mathcal{A}}^\dagger(O)\|_F$ ,

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \|\mathcal{C}^\dagger(O) - \mathcal{C}_{\mathcal{A}}^\dagger(O)\|_F^2 = \frac{1}{4^n} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \left( \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \bar{\mathcal{A}}} \Phi_\gamma(\mathcal{C}) \langle\langle O | P_L \rangle\rangle P_0 \right)^2 \right] \quad (\text{A121})$$

$$= \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \bar{\mathcal{A}}} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(\mathcal{C})] \langle\langle O | P_L \rangle\rangle^2 \quad (\text{A122})$$

$$= \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(\mathcal{C})] f_{\mathcal{A}, \text{Frob}}(\gamma), \quad (\text{A123})$$

where we used again orthogonality relation of Pauli paths in Eq. (A113) and we defined the function

$$f_{\mathcal{A}, \text{Frob}}(\gamma) = \begin{cases} 1 & \text{if } \gamma \notin \mathcal{A}, \\ 0 & \text{if } \gamma \in \mathcal{A}. \end{cases} \quad (\text{A124})$$

We are now ready to state the main result of this section.

*Supplemental Theorem A1 (Monte Carlo estimates).* Let  $\mathcal{D}_{\text{circ}}$  be a distribution over  $L$ -layered quantum channels  $\mathcal{C}_L \circ \mathcal{C}_{L-1} \circ \dots \circ \mathcal{C}_1$  such that all the channels  $\mathcal{C}_j$  are sampled from Pauli-invariant distributions. Let  $f : \mathcal{P}_n^{L+1} \rightarrow [0, 1]$  be a function of the Pauli paths computable in time  $T$ . Given an observable  $O$ , let  $F(\mathcal{D}_{\text{circ}})$  be defined as follows:

$$F(\mathcal{D}_{\text{circ}}) := \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle\langle O | P_L \rangle\rangle^2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(\mathcal{C})] f(\gamma). \quad (\text{A125})$$

Assume that we can sample  $P \in \mathcal{P}_n$  with probability proportional to  $\langle\langle O | P_L \rangle\rangle^2$  in time  $T$ . Moreover, assume that for all  $j \in [L]$  and for all  $P \in \mathcal{P}_n$ , we can also sample  $Q \in \mathcal{P}_n$  with probability proportional to  $\mathbb{E}_{\mathcal{C}_j \sim \mathcal{D}_j} \langle\langle P | \hat{\mathcal{C}}_j | Q \rangle\rangle^2$  in time  $T$ . Then for any  $\varepsilon, \delta \in (0, 1]$ , there exists a classical randomized algorithm that runs in time

$$\mathcal{O}(TL\varepsilon^{-2} \log(1/\delta)) \quad (\text{A126})$$

and outputs a value  $\Lambda$  such that

$$|\Lambda - F(\mathcal{D}_{\text{circ}})| \leq \varepsilon \|O\|_F^2, \quad (\text{A127})$$

with probability at least  $1 - \delta$ .

*Proof.* We can estimate  $F(\mathcal{D}_{\text{circ}})$  using the following randomized algorithm.

Let  $M := \lceil 2 \log(2/\delta)/\varepsilon^2 \rceil$ . For  $j = 1, 2, \dots, M$ , repeat the following steps:

(1) Sample  $P_L$  with probability

$$\frac{\langle\langle O | P_L \rangle\rangle^2}{\sum_{P \in \mathcal{P}_n} \langle\langle O | P \rangle\rangle^2} = \frac{\langle\langle O | P_L \rangle\rangle^2}{\|O\|_F^2}. \quad (\text{A128})$$

(2) For  $i = L, \dots, 1$ , given  $P_i$ , sample  $P_{i-1}$  with probability

$$\begin{aligned} & \frac{\mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \langle\langle P_i | \hat{\mathcal{C}}_i | P_{i-1} \rangle\rangle^2}{\sum_{P \in \mathcal{P}_n} \mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \langle\langle P_i | \hat{\mathcal{C}}_i | P \rangle\rangle^2} \\ &= \frac{\mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \langle\langle P_i | \hat{\mathcal{C}}_i | P_{i-1} \rangle\rangle^2}{\mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \|\mathcal{C}_i^\dagger(P_i)\|_F^2}. \end{aligned} \quad (\text{A129})$$

(3) Given  $\gamma = (P_1, P_2, \dots, P_L)$ , output  $\lambda_j = K(\gamma) \cdot f(\gamma)$ , where the normalization factor  $K(\gamma)$  is defined as

$$K(\gamma) = \|O\|_F^2 \prod_{i=1}^L \left( \mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \|\mathcal{C}_i^\dagger(P_i)\|_F^2 \right). \quad (\text{A130})$$

Finally, we compute the empirical mean  $\Lambda = \frac{1}{M} \sum_{j=1}^M \lambda_j$ .

We notice that, at each iteration, a path  $\gamma$  is sampled with probability

$\text{Pr}[\gamma = (P_0, P_1, \dots, P_L) \text{ is sampled}]$

$$= \frac{\langle\langle O | P_L \rangle\rangle^2}{\|O\|_F^2} \prod_{i=1}^L \left( \frac{\mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \langle\langle P_i | \hat{\mathcal{C}}_i | P_{i-1} \rangle\rangle^2}{\mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \|\mathcal{C}_i^\dagger(P_i)\|_F^2} \right). \quad (\text{A131})$$

Therefore procedure described above yields an unbiased estimator for the target value  $F(\mathcal{D}_{\text{circ}})$ , as the expected value of  $\lambda_j$  is

$$\mathbb{E}[\lambda_j] = \sum_{\gamma=(P_0, P_1, \dots, P_L)} \Pr[\gamma \text{ is sampled}] K(\gamma) f(\gamma) = \langle O | P_L \rangle^2 \prod_{i=1}^L (\mathbb{E}_{C_i \sim \mathcal{D}_i} \langle P_i | \hat{C}_i | P_{i-1} \rangle^2) f(\gamma) \quad (\text{A132})$$

$$= \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle O | P_L \rangle^2 \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(C)] f(\gamma) = F(\mathcal{D}_{\text{circ}}). \quad (\text{A133})$$

Moreover, we observe that

$$\|C_i^\dagger(P_i)\|_F^2 \leq \|C_i^\dagger(P_i)\|_\infty^2 \leq \|P_i\|_\infty^2 = 1, \quad (\text{A134})$$

where the first inequality follows from the relation  $2^n \|A\|_F^2 = \|A\|_2^2 \leq 2^n \|A\|_\infty^2$  and the second inequality follows from the Russo-Dye theorem [71] (which states  $\|C^\dagger(A)\|_\infty \leq \|A\|_\infty$  for any quantum channel  $C$ ), and in the final step we noted that the operator norm of any Pauli matrix is 1. Thus, the random variable  $\lambda_j$  can be bounded as follows:

$$\begin{aligned} \lambda_j = K(\gamma) f(\gamma) &\leq \|O\|_F^2 \prod_{i=1}^L \left( \mathbb{E}_{C_i \sim \mathcal{D}_i} \|C_i^\dagger(P_i)\|_F^2 \right) \\ &\leq \|O\|_F^2, \end{aligned} \quad (\text{A135})$$

where we used the fact that  $f(\gamma) \leq 1$  and  $\|C_i^\dagger(P_i)\|_F^2 \leq 1$ . Moreover,  $\lambda_j$  is always positive since both  $K(\gamma) f(\gamma)$  are positive for all  $\gamma$ . Therefore,  $\lambda_j$  lies in the interval  $[0, 1]$ .

We can conclude the proof by invoking the Chernoff-Hoeffding bound, which guarantees the final outcome approximates the mean squared error with precision  $\varepsilon \|O\|_F^2$  with probability at least  $1 - \delta$ . ■

## APPENDIX B: TECHNICAL LEMMAS FOR NOISY RANDOM CHANNELS

In this section we introduce several technical tools for studying the effect of noise on random circuits.

### 1. Motivation: Bounding the Frobenius norm under Heisenberg evolution

While our work encompasses arbitrary local noise channels, most of the previous research focuses on the “depolarizing class.” A widely employed feature of the depolarizing class is that it contracts the Frobenius norm of any traceless observable. In particular, let  $\mathcal{N}_p^{(\text{depo})}$  be the single-qubit depolarizing channel with noise rate  $p \in [0, 1]$  and  $O = \sum_{P \in \mathcal{P}_n} a_P P$  be an observable with Frobenius

norm  $\|O\|_F^2 = \sum_{P \in \mathcal{P}_n} a_P^2$ . We have

$$\mathcal{N}_p^{(\text{depo})\dagger \otimes n}(O) = \mathcal{N}_p^{(\text{depo}) \otimes n}(O) = \sum_{P \in \mathcal{P}_n} a_P (1-p)^{|P|} P, \quad (\text{B1})$$

and therefore

$$\|\mathcal{N}_p^{(\text{depo})\dagger \otimes n}(O)\|_F^2 = \sum_{P \in \mathcal{P}_n} a_P^2 (1-p)^{2|P|}. \quad (\text{B2})$$

Thus  $\mathcal{N}_p^{(\text{depo})\dagger \otimes n}$  strictly contracts the Frobenius norm if  $O$  is traceless, as we have

$$\|\mathcal{N}_p^{(\text{depo})\dagger \otimes n}(O)\|_F^2 \leq (1-p)^2 \|O\|_F^2. \quad (\text{B3})$$

However, this is not true in general for other kinds of noise. For instance, observables supported on  $\{I, Z\}^{\otimes n}$  are invariant under the dephasing channel. Moreover, non-unital noise can *increase* the Frobenius norm under Heisenberg evolution. Let  $\mathcal{A}_q$  be the amplitude-damping channel with noise rate  $q$ . We have

$$\begin{aligned} \mathcal{A}_q^{\dagger \otimes 2}(ZZ + IZ + ZI) \\ = (1-q)^2 ZZ + (1-q^2)(IZ + ZI) + (2q + q^2)II. \end{aligned} \quad (\text{B4})$$

Therefore, the corresponding Frobenius norm is

$$\begin{aligned} \|\mathcal{A}_q^{\dagger \otimes 2}(ZZ + IZ + ZI)\|_F^2 \\ = q^2(2+q)^2 + (1-q)^4 + 2(1-q^2)^2. \end{aligned} \quad (\text{B5})$$

In particular, setting  $q = 1$ , we find that the squared Frobenius norm strictly increases:

$$\frac{\|\mathcal{A}_q^{\dagger \otimes 2}(ZZ + IZ + ZI)\|_F^2}{\|ZZ + IZ + ZI\|_F^2} = 3. \quad (\text{B6})$$

This preliminary observation suggests that techniques relying on norm truncation may only be applicable when

the noise channel belongs to the depolarizing class. Despite this intuition, we will demonstrate that the adjoint of any local noise channel—potentially non-unital or of a dephasing nature—contracts the Frobenius norm of any traceless observable under averaging over random unitaries. Our results generalize the bounds presented in previous works such as Refs. [37,45]. While these previous results require the presence of local and global unitary 2-designs respectively, our approach encompasses a wider class of circuit ensembles.

## 2. Average-case norm contraction

We provide the following definition of contraction coefficients for single-qubit channels, which will serve as a useful technical tool for studying the average-case contraction of the Frobenius norm under local noise channels.

*Definition B1 (Contraction coefficients).* Given an observable  $O = \sum_{P \in \mathcal{P}_n} a_P P$  and a subset of qubits  $A \subseteq [n]$ , denote by  $O_A$  the restriction to its Pauli terms being nonidentity on  $A$  and identity elsewhere,

$$O_A := \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P P. \quad (\text{B7})$$

Given a single-qubit quantum channel  $\mathcal{N}$ , we define the contraction coefficient of  $\mathcal{N}$  as

$$\chi(\mathcal{N}) := \max_{A \subseteq [n]} \max_{\substack{O \\ \|O_A\|_F \neq 0}} \left( \frac{\|\mathcal{N}^{\dagger \otimes n}(O_A)\|_F}{\|O_A\|_F} \right)^{1/|A|}. \quad (\text{B8})$$

Moreover, given a distribution  $\mathcal{D}$  over  $\mathbb{U}(2)$ , we define the *mean squared* contraction coefficient of  $\mathcal{N}$  with respect to the  $\mathcal{D}$  as

$$\chi_{\mathcal{D}}^2(\mathcal{N}) := \max_{A \subseteq [n]} \max_{\substack{O \\ \|O_A\|_F \neq 0}} \left( \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \frac{\|\mathcal{N}^{\dagger \otimes n}(W^\dagger O_A W)\|_F^2}{\|O_A\|_F^2} \right)^{1/|A|}. \quad (\text{B9})$$

We will also refer to the square root of  $\chi_{\mathcal{D}}^2(\mathcal{N})$  as the *average* contraction coefficient, and denote it by  $\chi_{\mathcal{D}}(\mathcal{N})$ .

We also remark that the contraction coefficient of a channel  $\mathcal{N}(\cdot) = U\mathcal{N}'V(\cdot)V^\dagger U^\dagger$  can be expressed in terms of  $\mathcal{N}'$ .

*Observation B1.* Given a noise channel expressed in normal form as  $\mathcal{N}(\cdot) = U\mathcal{N}'V(\cdot)V^\dagger U^\dagger$ , then  $\mathcal{N}$  and  $\mathcal{N}'$  have the same contraction coefficient  $\chi(\mathcal{N})$ . This follows from the following two facts: first, for any observable

$O$  and subset of qubits  $A$ , we have that  $(U^{\dagger \otimes n} O U^{\otimes n})_A = U^{\dagger \otimes n} O_A U^{\otimes n}$ , and second, the Frobenius norm is unitarily invariant. Thus, we obtain that

$$\begin{aligned} \chi(\mathcal{N}) &= \max_{A \subseteq [n]} \max_{\substack{O \\ \|O_A\|_F \neq 0}} \\ &\quad \times \left( \frac{\|V^{\dagger \otimes n} \mathcal{N}'^{\dagger \otimes n} (U^{\dagger \otimes n} O_A U^{\otimes n}) V^{\otimes n}\|_F}{\|O_A\|_F} \right)^{1/|A|} \end{aligned} \quad (\text{B10})$$

$$= \max_{A \subseteq [n]} \max_{\substack{O \\ \|O_A\|_F \neq 0}} \left( \frac{\|\mathcal{N}'^{\dagger \otimes n}(O_A)\|_F}{\|O_A\|_F} \right)^{1/|A|} = \chi(\mathcal{N}'). \quad (\text{B11})$$

Applying iteratively Supplemental Lemma A5, we can prove that the contraction coefficient  $\chi(\mathcal{N})$  is strictly smaller than 1 for noise channels within the depolarizing class or the non-unital class.

*Supplemental Lemma B1 (General upper bound on the contraction coefficients).* Let  $\mathcal{N}(\cdot) = U\mathcal{N}'(V(\cdot)V^\dagger)U^\dagger$  be a single-qubit channel with normal form parameters  $\mathbf{D}$  and  $\mathbf{t}$ . We have

$$\begin{aligned} \chi^2(\mathcal{N}) &\leq \Upsilon(\mathbf{D}, \mathbf{t}) \\ &:= \max_{\|\mathbf{a}\|_2^2=1} \sum_{Q \in \{X,Y,Z\}} a_Q^2 D_Q^2 + \left( \sum_{Q \in \{X,Y,Z\}} a_Q t_Q \right)^2. \end{aligned} \quad (\text{B12})$$

In particular, by Supplemental Lemma A5, this implies that the contraction coefficient of any single-qubit channel  $\mathcal{N}$  is at most 1, and it is strictly smaller than 1 if  $\mathcal{N}$  satisfies  $\|\mathbf{D}\|_\infty^2 \in (0, 1)$  or  $\|\mathbf{t}\|_2^2 \in (0, 1)$ .

*Proof.* As a preliminary step, we establish the following claim. Intuitively, we upper-bound the contraction by isolating the contribution of each single-qubit Pauli component and analyzing how the normal-form parameters  $(\mathbf{D}, \mathbf{t})$  scale these components under Heisenberg evolution.

*Claim B1.* Let  $H = \sum_{P \in \{X,Y,Z\}} \sum_{Q \in \{I,X,Y,Z\}^{\otimes(k-1)}} b_P \otimes Q$  be an observable. We have

$$\|\mathcal{N}^{\dagger} \otimes \mathcal{I}_{k-1}(H)\|_F^2 \leq c \|H\|_F^2, \quad (\text{B13})$$

where  $\mathcal{I}_{k-1}(\cdot)$  is the identity channel acting on  $k-1$  qubits and  $c := \Upsilon(\mathbf{D}, \mathbf{t})$ .

*Proof of the Claim.* We start by writing the expression of  $\mathcal{N}^{\dagger} \otimes \mathcal{I}_{k-1}(H)$

$$\mathcal{N}^{\dagger} \otimes \mathcal{I}_{k-1}(H) = \sum_{Q \in \{I, X, Y, Z\}^{\otimes(k-1)}} \left( \sum_{P \in \{X, Y, Z\}} D_P b_{P \otimes Q} P \otimes Q + \left( \sum_{P \in \{X, Y, Z\}} b_{P \otimes Q} t_P \right) I \otimes Q \right). \quad (\text{B14})$$

Therefore the normalized Frobenius norm of  $\mathcal{N}^{\dagger} \otimes \mathcal{I}_{k-1}(H)$  satisfies

$$\|\mathcal{N}^{\dagger} \otimes \mathcal{I}_{k-1}(H)\|_F^2 = \sum_{Q \in \{I, X, Y, Z\}^{\otimes(k-1)}} \left( \sum_{P \in \{X, Y, Z\}} D_P^2 b_{P \otimes Q}^2 + \left( \sum_{P \in \{X, Y, Z\}} b_{P \otimes Q} t_P \right)^2 \right) \quad (\text{B15})$$

$$= \sum_{Q \in \{I, X, Y, Z\}^{\otimes(k-1)}} \left( \sum_{P \in \{X, Y, Z\}} D_P^2 \frac{b_{P \otimes Q}^2}{\|\mathbf{b}_Q\|_2^2} + \left( \sum_{P \in \{X, Y, Z\}} \frac{b_{P \otimes Q}}{\|\mathbf{b}_Q\|_2} t_P \right)^2 \right) \|\mathbf{b}_Q\|_2^2 \quad (\text{B16})$$

$$\leq \Upsilon(\mathbf{D}, \mathbf{t}) \sum_{Q \in \{I, X, Y, Z\}^{\otimes(k-1)}} \|\mathbf{b}_Q\|_2^2 := c \|H\|_F^2, \quad (\text{B17})$$

where we denoted  $\|\mathbf{b}_Q\|_2^2 := \sum_{P \in \{X, Y, Z\}} b_{P \otimes Q}^2$  and the only inequality follows from Supplemental Lemma A5. This concludes the proof. ■

Thanks to Claim B1, we have reduced the  $n$ -qubit statement to a single step on one qubit, after which we can iterate the argument over all qubits in the support. Given a subset of qubits  $A \subseteq [n]$  with size  $|A| = k$ , let  $O_A = \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P) = A}} a_P P$  be a linear combination of Pauli operators supported on  $A$ . Given  $j \in A$ , there exist some suitable coefficients  $b_{P \otimes Q}$  such that

$$\bigotimes_{i \in A \setminus \{j\}} \mathcal{N}_{\{i\}}^{\dagger}(O_A) = \sum_{P \in \{X, Y, Z\}} \sum_{Q \in \{I, X, Y, Z\}^{\otimes(k-1)}} b_{P \otimes Q} P \otimes Q \otimes I^{\otimes(n-k)}, \quad (\text{B18})$$

where the channel  $\mathcal{N}_{\{i\}}^{\dagger}$  act on the  $i$ th qubit and we have omitted the identity channel acting on the remaining  $n + 1 - k$  qubits in order to simplify the notation. Hence the above observable satisfies the assumptions of Claim B1. We can thus prove the desired result by applying iteratively Claim B1,

$$\|\mathcal{N}^{\dagger \otimes n}(O_A)\|_F^2 = \left\| \bigotimes_{i \in A} \mathcal{N}_{\{i\}}^{\dagger}(O_A) \right\|_F^2 = \left\| \mathcal{N}_{\{j_1\}}^{\dagger} \left( \bigotimes_{i \in A \setminus \{j_1\}} \mathcal{N}_{\{i\}}^{\dagger}(O_A) \right) \right\|_F^2 \quad (\text{B19})$$

$$\leq c \left\| \bigotimes_{i \in A \setminus \{j_1\}} \mathcal{N}_{\{i\}}^{\dagger}(O_A) \right\|_F^2 = c \left\| \mathcal{N}_{\{j_2\}}^{\dagger} \left( \bigotimes_{i \in A \setminus \{j_1, j_2\}} \mathcal{N}_{\{i\}}^{\dagger}(O_A) \right) \right\|_F^2 \quad (\text{B20})$$

$$\leq c^2 \left\| \bigotimes_{i \in A \setminus \{j_1, j_2\}} \mathcal{N}_{\{i\}}^{\dagger}(O_A) \right\|_F^2 \leq \dots \leq c^k \|O_A\|_F^2, \quad (\text{B21})$$

where we have again omitted the identity channels to simplify the notation. The final inequality follows by applying the single-qubit bound sequentially to each site in  $A$ . Since each step contributes at most a factor of  $c$ , the total effect is upper bounded by  $c^{|A|} = c^k$ . Since  $c = \Upsilon(\mathbf{D}, \mathbf{t})$ , the statement of the Lemma follows. ■

Since  $\Upsilon(\mathbf{D}, \mathbf{t}) < 1$  only when the channel  $\mathcal{N}$  is either depolarizing-like or non-unital, a different approach is required to analyze the contraction properties of dephasing-like noise. In particular, we will demonstrate that such noise still induces norm contraction when the initial Pauli observable is scrambled—i.e., it spreads into at least two distinct Pauli components—under the action of a unitary transformation. This idea is captured by the notion of *approximate scrambling*, as defined in Definition A4. We formalize this insight in the following lemma.

*Supplemental Lemma B2 (Contraction coefficients under approximately locally scrambling evolution).* Let  $\mathcal{N}$  be a single-qubit channel with normal form parameters  $\mathbf{D}$  and  $\mathbf{t}$  and let  $\mathcal{D}$  be an  $\eta$ -approximate scrambler over  $\mathbb{U}(2)$  for some  $\eta \in [0, 1]$ .

- (1) If  $\eta = 0$  (i.e., if  $\mathcal{D}$  is a single-qubit unitary 2-design), we have

$$\chi_{\mathcal{D}}^2(\mathcal{N}) = \frac{\|\mathbf{D}\|_2^2 + \|\mathbf{t}\|_2^2}{3}. \quad (\text{B22})$$

- (2) If  $\mathcal{N}$  belongs to the dephasing class, we have

$$\chi_{\mathcal{D}}^2(\mathcal{N}) \leq \eta + (1 - \eta) \frac{\|\mathbf{D}\|_2^2}{3}. \quad (\text{B23})$$

*Proof.* Let  $\mathcal{N}(\cdot) = U\mathcal{N}'(V(\cdot)V^\dagger)U^\dagger$  be an arbitrary single-qubit channel. Given  $P \in \{X, Y, Z\}$ , let  $V^\dagger W^\dagger P W V = \sum_{Q \in \{X, Y, Z\}} a_Q Q$ . By the definition of  $\eta$ -approximate scrambler and Supplemental Lemma A2, we have that

$$\begin{aligned} & \max_{Q \in \{X, Y, Z\}} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} a_Q^2 \\ &= \max_{Q \in \{X, Y, Z\}} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \langle \langle P | W U \otimes W^* U^* | Q \rangle \rangle^2 \\ &\leq \frac{1 + 2\eta}{3}, \end{aligned} \quad (\text{B24})$$

where for  $\eta = 0$  the inequality is in fact an equality. We obtain that

$$\begin{aligned} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \|\mathcal{N}^\dagger(W^\dagger P W)\|_F^2 &= \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \sum_{Q \in \{X, Y, Z\}} a_Q^2 D_Q^2 \\ &\quad + \left( \sum_{Q \in \{X, Y, Z\}} a_Q t_Q \right)^2. \end{aligned} \quad (\text{B25})$$

*Case 1:  $\eta = 0$ ,  $\mathcal{N}$  arbitrary.* We have

$$\mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \|\mathcal{N}^\dagger(W^\dagger P W)\|_F^2 = \frac{1}{3} (\|\mathbf{D}\|_2^2 + \|\mathbf{t}\|_2^2). \quad (\text{B26})$$

Moreover, given  $A \subseteq [n]$  a  $O = \sum_{P \in \mathcal{P}_n} b_P P$ , we have that

$$\begin{aligned} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \|\mathcal{N}^\dagger(W^\dagger O W)\|_F^2 &= \sum_{P \in \mathcal{P}_n} b_P^2 \left\{ \frac{\|\mathbf{D}\|_2^2 + \|\mathbf{t}\|_2^2}{3} \right\}^{|A|} \\ &= \left\{ \frac{\|\mathbf{D}\|_2^2 + \|\mathbf{t}\|_2^2}{3} \right\}^{|A|} \|O\|_F^2. \end{aligned} \quad (\text{B27})$$

*Case 2:  $\eta \neq 0$ ,  $\mathcal{N}$  within the dephasing class.* Assume without loss of generality that  $D_X = D_Y = D < 1$  and

$D_Z = 1$ . We need to upper bound the following quantity:

$$\begin{aligned} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \|\mathcal{N}^\dagger(W^\dagger P W)\|_F^2 &= \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \sum_{Q \in \{X, Y, Z\}} a_Q^2 D_Q^2 \\ &= \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} (a_X^2 + a_Y^2) D + a_Z^2, \end{aligned} \quad (\text{B28})$$

under the constraint that  $\max_{Q \in \{X, Y, Z\}} \mathbb{E}_{W \sim \mathcal{D}} a_Q^2 \leq \frac{1}{3}(1 + 2\eta)$ . Since only the  $Z$  component is preserved under dephasing, the worst-case contraction arises when the scrambling distribution concentrates as much weight as possible on the  $Z$  basis element. Thus, when  $\mathbb{E}_{W \sim \mathcal{D}} a_Z^2 = \frac{1}{3}(1 + 2\eta)$  the expression is maximized and takes value

$$\left(1 - \frac{1}{3}(1 + 2\eta)\right) D + \frac{1}{3}(1 + 2\eta) = \eta + (1 - \eta) \frac{\|\mathbf{D}\|_2^2}{3}. \quad (\text{B29})$$

Moreover, given  $A \subseteq [n]$  a  $O = \sum_{P \in \mathcal{P}_n} b_P P$ , we have that

$$\begin{aligned} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \|\mathcal{N}^\dagger(W^\dagger O W)\|_F^2 &= \sum_{P \in \mathcal{P}_n} b_P^2 \left\{ \eta + (1 - \eta) \frac{\|\mathbf{D}\|_2^2}{3} \right\}^{|A|} \\ &= \left\{ \eta + (1 - \eta) \frac{\|\mathbf{D}\|_2^2}{3} \right\}^{|A|} \|O\|_F^2. \end{aligned} \quad (\text{B30})$$

■

The contraction coefficients proposed in Definition B1 can be related to the average-case contraction of the Frobenius norm.

*Supplemental Lemma B3 (Average norm contraction).* Let  $\mathcal{N}$  be a single-qubit channel and let  $\mathcal{D}$  be a 1-design over  $\mathbb{U}(2)$ . For all observables  $O = \sum_{P \in \mathcal{P}_n} a_P P$ , we have

$$\mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \|\mathcal{N}^{\dagger \otimes n}(W^\dagger O W)\|_F^2 \leq \sum_{P \in \mathcal{P}_n} \{\chi_{\mathcal{D}}^2(\mathcal{N})\}^{|P|} a_P^2. \quad (\text{B31})$$

In particular, we also have

$$\begin{aligned} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \|\mathcal{N}^{\dagger \otimes n}(W^\dagger O W)\|_F^2 &- 4^{-n} \text{Tr}[O]^2 \\ &\leq \{\chi_{\mathcal{D}}^2(\mathcal{N})\} (\|O\|_F^2 - 4^{-n} \text{Tr}[O]^2). \end{aligned} \quad (\text{B32})$$

*Proof.* In general, we can write

$$O = \sum_{w=1}^n \sum_{\substack{A \subseteq [n] \\ |A|=w}} \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P P. \quad (\text{B33})$$

We have

$$\mathbb{E}_{U \sim \mathcal{D}} (U^\dagger O U)^{\otimes 2} = \mathbb{E}_{U \sim \mathcal{D}} \sum_{w=1}^n \sum_{\substack{A \subseteq [n] \\ |A|=w}} \left( \sum_{P \in \mathcal{P}_n} a_P U^\dagger P U \right)^{\otimes 2}. \quad (\text{B34})$$

Fix a subset  $A$  of size  $k$ ,

$$\left\| \mathcal{N}^{\otimes n \dagger} \left( \sum_{P \in \mathcal{P}_n} a_P U^\dagger P U \right) \right\|_F^2 = \left\| \mathcal{N}^{\otimes k \dagger} \left( \sum_{P \in \{X, Y, Z\}^{\otimes k}} b_P P \right) \right\|_F^2 \leq \{\chi_{\mathcal{D}}^2(\mathcal{N})\}^k \|O\|_F^2, \quad (\text{B35})$$

for some suitable coefficients  $b_P$ . Putting all together

$$\mathbb{E}_{U \sim \mathcal{D}} \left\| \mathcal{N}^{\otimes n} (U^\dagger O U) \right\|_F^2 = \sum_{w=1}^n \sum_{\substack{A \subseteq [n] \\ |A|=w}} \sum_{P \in \mathcal{P}_n} \mathbb{E}_{U \sim \mathcal{D}} \left\| \mathcal{N}^{\otimes n \dagger} \left( \sum_{P \in \mathcal{P}_n} a_P U^\dagger P U \right) \right\|_F^2 \quad (\text{B36})$$

$$\leq \sum_{w=1}^n \sum_{\substack{A \subseteq [n] \\ |A|=w}} \sum_{P \in \mathcal{P}_n} \{\chi_{\mathcal{D}}^2(\mathcal{N})\}^w a_P^2 = \sum_{P \in \mathcal{P}_n} \{\chi_{\mathcal{D}}^2(\mathcal{N})\}^{|P|} a_P^2. \quad (\text{B37})$$

■

We conclude this section by explicitly providing the values of  $\chi(\mathcal{N})$  and  $\chi_{\mathcal{D}}(\mathcal{N})$  for amplitude-damping and dephasing noise.

*Example B1 (Contraction coefficients).* Let  $\mathcal{N}_\gamma^{(\text{amp})}$  be the amplitude-damping channel with noise strength  $\gamma$ , characterized by the normal form parameters  $\mathbf{D} = (\sqrt{1-\gamma}, \sqrt{1-\gamma}, 1-\gamma)$  and  $\mathbf{t} = (0, 0, \gamma)$ .

Then we obtain

$$\chi^2(\mathcal{N}_\gamma^{(\text{amp})}) \leq \max_{\|\mathbf{a}\|_2^2=1} \sum_{Q \in \{X, Y, Z\}} a_Q^2 D_Q^2 + \left( \sum_{Q \in \{X, Y, Z\}} a_Q t_Q \right)^2 \quad (\text{B38})$$

$$\leq \|\mathbf{D}\|_\infty^2 + \|\mathbf{t}\|_2^2 = 1 - \gamma + \gamma^2, \quad (\text{B39})$$

where the first step follows from Supplemental Lemma B1 and the second step follows from Hölder's inequality.

Furthermore, we also consider the dephasing channel  $\mathcal{N}_p^{(\text{deph})}$  with noise strength  $p$ , characterized by the normal

form parameters  $\mathbf{D} = (1-2p, 1-2p, 1)$  and  $\mathbf{t} = (0, 0, 0)$ . Let  $\mathcal{D}$  be the distribution over  $\mathbb{U}(2)$  obtained by performing two Pauli rotations along orthogonal directions with angles sampled independently and uniformly from  $[0, 2\pi)$ .

By Example A3, we know that the  $\mathcal{D}$  is an  $\eta$ -approximate scrambler with  $\eta = 1/4$ . Therefore, Supplemental Lemma B2 yields the following mean squared contraction coefficient:

$$\chi_{\mathcal{D}}^2(\mathcal{N}_p^{(\text{deph})}) = \frac{1 + (1-2p)^2}{2}. \quad (\text{B40})$$

### 3. Effective depolarizing rate of noisy random channels

As previously observed in Ref. [16], every quantum channel can be re-expressed as a sequential composition of the depolarizing channel and a suitable linear map, which is in general non-physical. In particular, given a value  $p \in [0, 1]$ , we decompose the noisy channel  $\mathcal{N}$  as

$$\mathcal{N} = \mathcal{N}_p^{(\text{depo})} \circ \tilde{\mathcal{N}}_p, \quad (\text{B41})$$

where  $\mathcal{N}_p^{(\text{depo})}$  is the depolarizing channel with noise rate  $p$ , i.e.,  $\mathcal{N}_p^{(\text{depo})}(\cdot) = (1-p)(\cdot) + p \frac{I}{2} \text{Tr}[\cdot]$ , and  $\tilde{\mathcal{N}}_p$  is the linear map given by

$$\tilde{\mathcal{N}}_p(I) = \mathcal{N}(I)/(1-p) - Ip/(1-p), \quad (\text{B42})$$

$$\tilde{\mathcal{N}}_p(X) = \mathcal{N}(X)/(1-p), \quad (\text{B43})$$

$$\tilde{\mathcal{N}}_p(Y) = \mathcal{N}(Y)/(1-p), \quad (\text{B44})$$

$$\tilde{\mathcal{N}}_p(Z) = \mathcal{N}(Z)/(1-p). \quad (\text{B45})$$

Moreover, the adjoint channel  $\tilde{\mathcal{N}}_p^\dagger$  is given by

$$\tilde{\mathcal{N}}_p^\dagger(I) = I, \quad (\text{B46})$$

$$\tilde{\mathcal{N}}_p^\dagger(X) = \mathcal{N}^\dagger(X)/(1-p), \quad (\text{B47})$$

$$\tilde{\mathcal{N}}_p^\dagger(Y) = \mathcal{N}^\dagger(Y)/(1-p), \quad (\text{B48})$$

$$\tilde{\mathcal{N}}_p^\dagger(Z) = \mathcal{N}^\dagger(Z)/(1-p). \quad (\text{B49})$$

In particular, in Ref. [16] the above decomposition is applied to a specific class of non-unital noise, which models a spontaneous emission in a random direction with rate  $\gamma_s$ . Intuitively, such randomization is performed implicitly

in random circuits, via the random gates before and after each noise channel.

Crucially, the authors of Ref. [16] observed that the associated nonphysical map,  $\tilde{\mathcal{N}}_p^\dagger$ , does not increase the Frobenius norm of any operator on average over random noise channel directions, provided that  $p \leq \gamma_s/2$ .

We extend this result by demonstrating that a similar phenomenon holds for *any* noise channel  $\mathcal{N}$  when averaged over a tensor product of single-qubit gates sampled from unitary 1-designs. Specifically, we demonstrate that the Frobenius norm does not increase on average over  $\mathcal{D}$ , as long as value of  $p$  in Eq. (B41) is appropriately chosen according to the contraction coefficient  $\chi_{\mathcal{D}}(\mathcal{N})$ .

*Supplemental Lemma B4.* Let  $\mathcal{D}$  be a 1-design over  $\mathbb{U}(2)$  and let  $p := 1 - \chi_{\mathcal{D}}(\mathcal{N})$ . For all observables  $O$  we have

$$\mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \left\| \tilde{\mathcal{N}}_p^{\dagger \otimes n}(V^\dagger O V) \right\|_F^2 \leq \|O\|_F^2, \quad (\text{B50})$$

that is, the linear map  $\tilde{\mathcal{N}}_p^{\dagger \otimes n}$  does not increase the Frobenius norm in expectation over a randomly sampled  $V$ .

*Proof.* We denote  $O = \sum_{P \in \mathcal{P}_n} a_P P$ . We have

$$\mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \left\| \tilde{\mathcal{N}}_p^{\dagger \otimes n}(V^\dagger O V) \right\|_F^2 = \mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \sum_{k=1}^n \sum_{\substack{A \subseteq [n] \\ |A|=k}} \left\| \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P \tilde{\mathcal{N}}_p^{\dagger \otimes n}(V^\dagger P V) \right\|_F^2 \quad (\text{B51})$$

$$= \mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \sum_{k=1}^n \sum_{\substack{A \subseteq [n] \\ |A|=k}} \frac{1}{(1-p)^{2k}} \left\| \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P \mathcal{N}_p^{\dagger \otimes n}(V^\dagger P V) \right\|_F^2 \quad (\text{B52})$$

$$= \mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \sum_{k=1}^n \sum_{\substack{A \subseteq [n] \\ |A|=k}} \frac{1}{(1-p)^{2k}} \left\| (1-p)^k \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P V^\dagger P V \right\|_F^2 \quad (\text{B53})$$

$$= \mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \sum_{k=1}^n \sum_{\substack{A \subseteq [n] \\ |A|=k}} \left\| \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P V^\dagger P V \right\|_F^2 \quad (\text{B54})$$

$$= \|V^\dagger O V\|_F^2 = \|O\|_F^2, \quad (\text{B55})$$

where the first line follows from the orthogonality property of Pauli terms with different supports (Supplemental Lemma A1), the second line follows from the definition of  $\tilde{\mathcal{N}}_p$ , and the third line follows from the fact that the observable

$$\sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P V^\dagger P V \quad (\text{B56})$$

is a linear combination of Pauli terms with weight  $|A| = k$ , and therefore

$$\sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P \mathcal{N}_p^{\otimes n} (V^\dagger P V) = (1-p)^k \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} a_P V^\dagger P V. \quad (\text{B57})$$

Finally, the fourth line follows from the absolute homogeneity of Schatten norms (i.e.,  $\|\alpha X\|_p = |\alpha| \|X\|_p$  for all scalar  $\alpha$ ), and the fifth line from the unitarily invariance of Schatten norms. ■

### APPENDIX C: NOISY CIRCUIT MODEL

We introduce our noisy circuit model.

*Definition C1 (Noisy circuit model).* A noisy  $L$ -layered quantum circuit  $\mathcal{C}$  consists of  $n$ -qubit unitary layers interleaved with local (single-qubit) noise, concluding with a final layer of single-qubit gates. Formally, we express the circuit as

$$\mathcal{C} := \mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_L \circ \dots \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_1. \quad (\text{C1})$$

Let  $\mathcal{D}_{\text{single}}$  a 1-design over  $\mathbb{U}(2)$ . We make the following assumptions:

- (1)  $\mathcal{N}$  is a single-qubit quantum channel.
- (2)  $\mathcal{V}^{\text{single}} := V(\cdot)V^\dagger$ , with  $V := \bigotimes_{i=1}^n V_i$ , represents a layer of single-qubit gates. We assume that  $V$  is sampled from a distribution  $\mathcal{D}_{L+1}$ .
- (3) For  $i = 1, 2, \dots, L$ ,  $\mathcal{U}_i = U_i(\cdot)U_i^\dagger$  represents the quantum channel associated with the  $i$ th unitary circuit layer. We assume that  $U_i$  is sampled from a distribution  $\mathcal{D}_i$  and that it is composed of non-overlapping gates, each acting on  $\mathcal{O}(1)$ -qubits.
- (4) For all  $i = 1, 2, \dots, L+1$ , the distribution  $\mathcal{D}_i$  satisfies the following:

$$\mathbb{E}_{U \sim \mathcal{D}_i} [U \otimes U^*] = \mathbb{E}_{U \sim \mathcal{D}_i'} \mathbb{E}_{W \sim \mathcal{D}_{\text{single}}^{\otimes n}} [UW \otimes U^* W^*]. \quad (\text{C2})$$

We denote by  $\mathcal{D}_{\text{circ}}$  the overall distribution of the noisy circuits described above.

It is worth noting that the final unitary layer in the circuit is not essential. Moreover, if the circuit were to end with an additional layer of local noise  $\mathcal{N}^{\otimes n}$ , this noise could be absorbed into the measured observable, effectively replacing  $O$  with  $\mathcal{N}^{\dagger \otimes n}(O)$ .

Moreover, our results remain valid if we replace Eq. (C2) with the following more general expression:

$$\mathbb{E}_{U \sim \mathcal{D}_i} [U \otimes U^*] = \mathbb{E}_{U \sim \mathcal{D}_i'} \mathbb{E}_{W \sim \bigotimes_{j=1}^n \mathcal{D}_{\text{single},i}^{(j)}} [UW \otimes U^* W^*], \quad (\text{C3})$$

where  $\mathcal{D}_{\text{single},i}^{(j)}$  represents 1-designs over  $\mathbb{U}(2)$  for all  $i \in [L+1]$  and  $j \in [n]$ . For simplicity and to avoid overly complex notation, we assume a single-qubit distribution  $\mathcal{D}_{\text{single}}$  across all qubits and circuit layers.

Similarly, our results can be easily extended to circuits influenced by noise channels that vary across qubits. Specifically, this involves replacing  $\mathcal{N}^{\otimes n}$  with the more general expression  $\bigotimes_{j=1}^n \mathcal{N}^{(j)}$ .

Supplemental Lemma B4 hints to the fact that all local noise channels behave similarly to the depolarizing noise on typical instances of random circuits. Thus, given a distribution over noisy circuits such as the one in Definition C1, it is natural to assign to it an *effective depolarizing rate*.

*Definition C2 (Effective depolarizing rate of noisy random circuits).* Let  $\mathcal{D}_{\text{circ}}$  be the circuit distribution introduced in Definition C1. We define the associated effective depolarizing rate of  $\mathcal{D}_{\text{circ}}$  as

$$p(\mathcal{N}, \mathcal{D}_{\text{circ}}) := 1 - \chi_{\mathcal{D}_{\text{single}}}(\mathcal{N}), \quad (\text{C4})$$

where  $\chi_{\mathcal{D}_{\text{single}}}(\mathcal{N})$  is the average contraction coefficient introduced in Definition B1.

### APPENDIX D: CLASSICAL SIMULATION OF NOISY RANDOM CIRCUITS

In this section, we simplify notation by writing the effective depolarizing rate  $p(\mathcal{N}, \mathcal{D}_{\text{circ}})$  simply as  $p$ . We will provide some additional notation to facilitate the proof of our main result. For convenience, we also rewrite the noisy circuit as

$$\mathcal{C} := \mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_L \circ \dots \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_1 \quad (\text{D1})$$

$$= \mathcal{V}^{\text{single}} \circ \mathcal{N}_p^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_L \circ \dots \circ \mathcal{N}_p^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_1 \quad (\text{D2})$$

$$= \mathcal{C}_L \circ \mathcal{C}_{L-1} \circ \dots \circ \mathcal{C}_1, \quad (\text{D3})$$

where the layers  $\mathcal{C}_j$  are given by

$$\mathcal{C}_j := \begin{cases} \mathcal{V}^{\text{single}} \circ \mathcal{N}_p^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_L & \text{if } j = L, \\ \mathcal{N}_p^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_j & \text{if } j \neq L. \end{cases} \quad (\text{D4})$$

We also consider the (non-physical) linear map obtained by “taking out” the depolarizing channels,

$$\tilde{\mathcal{C}} := \mathcal{V}^{\text{single}} \circ \tilde{\mathcal{N}}_p^{\otimes n} \circ \mathcal{U}_L \circ \dots \circ \tilde{\mathcal{N}}_p^{\otimes n} \circ \mathcal{U}_1 \quad (\text{D5})$$

$$= \tilde{\mathcal{C}}_L \circ \tilde{\mathcal{C}}_{L-1} \circ \dots \circ \tilde{\mathcal{C}}_1, \quad (\text{D6})$$

where we defined the layers  $\tilde{\mathcal{C}}_j$  as

$$\tilde{\mathcal{C}}_j := \begin{cases} \mathcal{V}^{\text{single}} \circ \tilde{\mathcal{N}}_p^{\otimes n} \circ \mathcal{U}_L & \text{if } j = L, \\ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_j & \text{if } j \neq L. \end{cases} \quad (\text{D7})$$

Given an observable  $O$  and an initial state  $\rho$ , we aim to approximate the expectation value

$$\text{Tr}[O\mathcal{C}(\rho)] = \text{Tr}[\mathcal{C}^\dagger(O)\rho]. \quad (\text{D8})$$

To this end, we fix a positive value  $k$  and we approximate the Heisenberg-evolved observable  $\mathcal{C}^\dagger(O)$  by computing only the Pauli paths with total weight smaller than  $k$ ,

$$O_C^{(k)} := \sum_{\substack{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}: \\ |\gamma| < k}} \langle\langle O | P_L \rangle\rangle \Phi_\gamma(\mathcal{C}) P_0, \quad (\text{D9})$$

where we recall that the Fourier coefficients  $\Phi_\gamma(\mathcal{C})$  are defined as

$$\Phi_\gamma(\mathcal{C}) = \langle\langle P_L | \hat{\mathcal{C}}_L | P_{L-1} \rangle\rangle \langle\langle P_{L-1} | \hat{\mathcal{C}}_{L-1} | P_{L-2} \rangle\rangle \dots \langle\langle P_1 | \hat{\mathcal{C}}_1 | P_0 \rangle\rangle. \quad (\text{D10})$$

We also consider the observable  $\tilde{\mathcal{C}}^\dagger(O)$  obtained by Heisenberg-evolving  $O$  with the linear map  $\tilde{\mathcal{C}}$ ,

$$\tilde{\mathcal{C}}^\dagger(O) = \sum_{\gamma=(P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle\langle O | P_L \rangle\rangle \Phi_\gamma(\tilde{\mathcal{C}}) P_0, \quad (\text{D11})$$

where the Fourier coefficients  $\Phi_\gamma(\tilde{\mathcal{C}})$  are defined as

$$\Phi_\gamma(\tilde{\mathcal{C}}) = \langle\langle P_L | \hat{\tilde{\mathcal{C}}}_L | P_{L-1} \rangle\rangle \langle\langle P_{L-1} | \hat{\tilde{\mathcal{C}}}_{L-1} | P_{L-2} \rangle\rangle \dots \langle\langle P_1 | \hat{\tilde{\mathcal{C}}}_1 | P_0 \rangle\rangle. \quad (\text{D12})$$

### 1. Average contraction implies efficient classical simulation

We start by proving that the Fourier coefficients of  $\mathcal{C}$  and  $\tilde{\mathcal{C}}$  are connected by a proportionality factor, which is exponential in the path weight.

*Supplemental Lemma D1 (Proportionality of Fourier coefficients).* Let  $\gamma \in \mathcal{P}_n^{L+1}$  be a Pauli path. Then, the associated Fourier coefficients  $\Phi_\gamma(\mathcal{C})$  and  $\Phi_\gamma(\tilde{\mathcal{C}})$  satisfy the identity

$$\Phi_\gamma(\mathcal{C}) = (1-p)^{|\gamma|} \Phi_\gamma(\tilde{\mathcal{C}}). \quad (\text{D13})$$

---

*Proof.* The lemma follows from the fact that  $\mathcal{C}_j = \mathcal{N}_p^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{C}}_j$  for all  $j \in [L]$ . Therefore,

$$\langle\langle P_j | \hat{\mathcal{C}}_j | P_{j-1} \rangle\rangle = \langle\langle \mathcal{N}_p^{(\text{depo}) \otimes n}(P_j) | \hat{\tilde{\mathcal{C}}}_j | P_{j-1} \rangle\rangle = (1-p)^{|P_j|} \langle\langle P_j | \hat{\tilde{\mathcal{C}}}_j | P_{j-1} \rangle\rangle. \quad (\text{D14})$$

Iterating over all  $j \in [L]$  we obtain the desired result,

$$\Phi_\gamma(\mathcal{C}) = \langle\langle P_L | \hat{\mathcal{C}}_L | P_{L-1} \rangle\rangle \langle\langle P_{L-1} | \hat{\mathcal{C}}_{L-1} | P_{L-2} \rangle\rangle \dots \langle\langle P_1 | \hat{\mathcal{C}}_1 | P_0 \rangle\rangle \quad (\text{D15})$$

$$= (1-p)^{|P_L|} \langle\langle P_L | \hat{\tilde{\mathcal{C}}}_L | P_{L-1} \rangle\rangle (1-p)^{|P_{L-1}|} \langle\langle P_{L-1} | \hat{\tilde{\mathcal{C}}}_{L-1} | P_{L-2} \rangle\rangle \dots (1-p)^{|P_1|} \langle\langle P_1 | \hat{\tilde{\mathcal{C}}}_1 | P_0 \rangle\rangle \quad (\text{D16})$$

$$= (1-p)^{|P_L|+|P_{L-1}|+\dots+|P_1|} \Phi_\gamma(\tilde{\mathcal{C}}) := (1-p)^{|\gamma|} \Phi_\gamma(\tilde{\mathcal{C}}). \quad (\text{D17})$$

■

Leveraging the definition of effective depolarizing rate of random noisy circuits, we show that, on average over the choice of random unitaries, the linear map  $\tilde{\mathcal{C}}$  does not increase the Frobenius norm under Heisenberg evolution.

*Supplemental Corollary D1 (Nonincreasing Frobenius norm).* It holds that

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \tilde{\mathcal{C}}^\dagger(O) \right\|_F^2 \leq \|O\|_F^2. \quad (\text{D18})$$

*Proof.* We start by making the following preliminary observation:

$$\mathbb{E}_{V \sim \mathcal{D}_{\text{single}}^{\otimes n}} \left\| \mathcal{U}_j \circ \tilde{\mathcal{N}}_p^{\otimes n} (V^\dagger O V) \right\|_F^2 = \mathbb{E}_{V \sim \mathcal{D}_{\text{single}}^{\otimes n}} \left\| \tilde{\mathcal{N}}_p^{\otimes n} (V^\dagger O V) \right\|_F^2 \leq \|O\|_F^2, \quad (\text{D19})$$

where the first step follows from the unitarily invariance of Schatten norms, and the second step follows applying Supplemental Lemma B4 and noting that the effective depolarizing rate is defined as  $p := 1 - \chi_{\mathcal{D}_{\text{single}}}(\mathcal{N})$ . Employing Eq. (D5), we can write

$$\mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \left\| \tilde{C}^\dagger(O) \right\|_F^2 = \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \left\| \tilde{C}_1^\dagger \circ \dots \circ \tilde{C}_{L-1}^\dagger \circ \tilde{C}_L^\dagger(O) \right\|_F^2 \quad (\text{D20})$$

$$= \mathbb{E}_{U_1 \sim \mathcal{D}_1, \dots, U_L \sim \mathcal{D}_L, V \sim \mathcal{D}_{\text{circ}}^{\otimes n}} \left\| \mathcal{U}_1^\dagger \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \dots \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \mathcal{U}_L^\dagger \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \mathcal{V}^{\text{single}^\dagger}(O) \right\|_F^2, \quad (\text{D21})$$

where we denoted  $\mathcal{U}_j := U_j(\cdot)U_j^\dagger$  and  $\mathcal{V}^{\text{single}} := V(\cdot)V^\dagger$ . Using Eq. (C2), we can insert independently sampled layers of random single-qubit gates within the above expression

$$\mathbb{E}_{U_1 \sim \mathcal{D}_1, \dots, U_L \sim \mathcal{D}_L, V \sim \mathcal{D}_{\text{circ}}^{\otimes n}} \left\| \mathcal{U}_1^\dagger \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \dots \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \mathcal{U}_L^\dagger \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \mathcal{V}^{\text{single}^\dagger}(O) \right\|_F^2 \quad (\text{D22})$$

$$= \mathbb{E}_{U_1 \sim \mathcal{D}'_1, \dots, U_L \sim \mathcal{D}'_L} \mathbb{E}_{V_1, V_2, \dots, V_L \sim \mathcal{D}_{\text{circ}}^{\otimes n}} \left\| \mathcal{U}_1^\dagger \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \mathcal{V}_1^{\text{single}^\dagger} \circ \dots \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \mathcal{V}_{L-1}^{\text{single}^\dagger} \circ \mathcal{U}_L^\dagger \circ \tilde{\mathcal{N}}_p^{\dagger \otimes n} \circ \mathcal{V}_L^{\text{single}^\dagger}(O) \right\|_F^2 \quad (\text{D23})$$

$$\leq \|O\|_F^2, \quad (\text{D24})$$

where the final inequality follows by applying iteratively Eq. (D19). ■

The core result follows by combining Supplemental Lemma D1 and Supplemental Corollary D1 with the orthogonality properties of different Pauli paths.

*Supplemental Theorem D1 (Mean squared error).* Let  $\mathcal{D}_{\text{circ}}$  be an  $L$ -layered locally unbiased distribution over noisy circuits, and let  $p$  be the effective depolarizing rate of  $\mathcal{D}_{\text{circ}}$ . Let  $O_C^{(k)} := \sum_{\gamma: |\gamma| < k} \langle O | P_L \rangle \Phi_\gamma(C) P_0$  be the observable obtained by computing the approximate evolution of  $O$  under  $C^\dagger$  restricted to the Pauli paths of weight less than  $k$ . We have

$$\mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \left\| O_C^{(k)} - C^\dagger(O) \right\|_F^2 \leq (1-p)^{2k} \|O\|_F^2. \quad (\text{D25})$$

Moreover, for any initial state  $\rho$  we have

$$\mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \left( O_C^{(k)} - C^\dagger(O) \right) \rho \right]^2 \leq (1-p)^{2k} \|O\|_F^2. \quad (\text{D26})$$

*Proof.* We start by proving the following technical claim, which implies that the contribution of paths with different weights to the mean squared error can be evaluated separately.

*Claim D1.* For all subsets of Pauli paths  $\mathcal{S} \subseteq \mathcal{P}_n^{L+1}$ , we have

$$\begin{aligned} & \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \left( \sum_{\gamma \in \mathcal{S}} \langle O | P_{L+1} \rangle \Phi_\gamma(\tilde{C}) P_0 \right)^{\otimes 2} \\ &= \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \sum_{w=1}^n \left( \sum_{\substack{\gamma \in \mathcal{S}: \\ |\gamma|=w}} \langle O | P_{L+1} \rangle \Phi_\gamma(\tilde{C}) P_0 \right)^{\otimes 2}. \end{aligned} \quad (\text{D27})$$

*Proof of the Claim.* We first explain the intuition behind the claim. Under the locally unbiased assumption, Pauli paths with different weights satisfy an orthogonality relation: their coefficients are uncorrelated, as formalized in Supplemental Lemma A6. As a consequence, cross-terms between paths of different weights vanish after averaging over  $\mathcal{D}_{\text{circ}}$ , so that the second moment decomposes into a sum of independent contributions labeled by the path weight. More formally, we compute

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left( \sum_{\gamma \in \mathcal{S}} \langle O | P_{L+1} \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_0 \right)^{\otimes 2} = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left( \sum_{w=1}^n \sum_{\substack{\gamma \in \mathcal{S}: \\ |\gamma|=w}} \langle O | P_{L+1} \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_0 \right)^{\otimes 2} \quad (\text{D28})$$

$$= \sum_{w,w'=1}^n \sum_{\gamma: |\gamma|=w} \sum_{\gamma': |\gamma'|=w'} \langle O | P_{L+1} \rangle \langle O | P'_{L+1} \rangle \underbrace{\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \Phi_{\gamma}(\tilde{\mathcal{C}}) \Phi_{\gamma'}(\tilde{\mathcal{C}}) P_0 \otimes P'_0}_{=0 \text{ if } w \neq w'} \quad (\text{D29})$$

$$= \sum_{w=1}^n \sum_{\gamma: |\gamma|=w} \sum_{\gamma': |\gamma'|=w} \langle O | P_{L+1} \rangle \langle O | P'_{L+1} \rangle \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \Phi_{\gamma}(\tilde{\mathcal{C}}) \Phi_{\gamma'}(\tilde{\mathcal{C}}) P_0 \otimes P'_0 \quad (\text{D30})$$

$$= \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \sum_{w=1}^n \left( \sum_{\substack{\gamma \in \mathcal{S}: \\ |\gamma|=w}} \langle O | P_{L+1} \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_0 \right)^{\otimes 2}, \quad (\text{D31})$$

where in the third step we invoked Supplemental Lemma A6. ■

This claim shows that, for the purpose of bounding the mean squared error, it is enough to understand the total contribution of all paths of a given weight  $w$  and then sum over  $w$ . We now exploit this decomposition together with the relation between the Fourier coefficients of the noisy circuit  $\mathcal{C}$  and of its locally depolarized version  $\tilde{\mathcal{C}}$ . In particular, we prove the desired results using the fact that the Fourier coefficients of  $\mathcal{C}$  and those of  $\tilde{\mathcal{C}}$  coincide, up to a proportionality factor which is exponential in the path weight. As for the first statement, we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| O_C^{(k)} - \mathcal{C}^\dagger(O) \right\|_F^2 = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \sum_{\gamma: |\gamma| \geq k} \langle O | P_{L+1} \rangle \Phi_{\gamma}(\mathcal{C}) P_0 \right\|_F^2 \quad (\text{D32})$$

$$= \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \sum_{\gamma: |\gamma| \geq k} (1-p)^{|\gamma|} \langle O | P_{L+1} \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_0 \right\|_F^2 \quad (\text{D33})$$

$$= \frac{1}{2^n} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \mathbb{F} \left( \sum_{\gamma: |\gamma| \geq k} (1-p)^{|\gamma|} \langle O | P_{L+1} \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{\gamma} \right)^{\otimes 2} \right] \quad (\text{D34})$$

$$= \frac{1}{2^n} \sum_{w \geq k} (1-p)^{2w} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \mathbb{F} \left( \sum_{\gamma: |\gamma|=w} \langle O | P_{L+1} \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{\gamma} \right)^{\otimes 2} \right] = (\star). \quad (\text{D35})$$

In the second step, we applied Supplemental Lemma D1, while in the fourth step, we invoked Claim 2. In words, we have isolated the contribution of all Pauli paths of a fixed weight; orthogonality ensures that cross-terms between different weights are zero in expectation. Supplemental Lemma D1 lets us replace each Fourier coefficient of the full noisy circuit with that of the non-physical map  $\tilde{\mathcal{C}}$ , weighted by a factor  $(1-p)^{|\gamma|}$ , which drives the exponential decay. We now bound the contribution of the discarded paths  $\{\gamma : |\gamma| \geq k\}$  to the mean squared Frobenius error,

$$(\star) \leq (1-p)^{2k} \frac{1}{2^n} \sum_{w \geq k} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \mathbb{F} \left( \sum_{\gamma: |\gamma|=w} \langle O | P_{L+1} \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{\gamma} \right)^{\otimes 2} \right] \quad (\text{D36})$$

$$\leq (1-p)^{2k} \frac{1}{2^n} \sum_{w \geq 0} \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \mathbb{F} \left( \sum_{\gamma: |\gamma|=w} \langle\langle O | P_{L+1} \rangle\rangle \Phi_\gamma(\tilde{C}) P_\gamma \right)^{\otimes 2} \right] \quad (\text{D37})$$

$$= (1-p)^{2k} \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \left\| \sum_{\gamma: |\gamma| \in \mathcal{P}_n^{L+2}} \langle\langle O | P_{L+1} \rangle\rangle \Phi_\gamma(\tilde{C}) P_\gamma \right\|_F^2 = (1-p)^{2k} \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \left\| \tilde{C}^\dagger(O) \right\|_F^2 \leq (1-p)^{2k} \|O\|_F^2. \quad (\text{D38})$$

The first inequality follows from the observation that the sum is taken over indices  $w \geq k$ , implying  $(1-p)^w \leq (1-p)^k$ . The second inequality holds because all the additional terms included in the sum are positive. Finally, the last inequality is a direct consequence of Supplemental Corollary D1.

The second statement can be shown as follows:

$$\mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \left( O_C^{(k)} - C^\dagger(O) \right) \rho \right]^2 = \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}, V \sim \mathcal{D}_{\text{single}}^{\otimes n}} \text{Tr} \left[ \left( O_C^{(k)} - C^\dagger(O) \right) (V \rho V^\dagger) \right]^2 \quad (\text{D39})$$

$$= \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} \left\| O_C^{(k)} - C^\dagger(O) \right\|_F^2 \leq (1-p)^{2k} \|O\|_F^2, \quad (\text{D40})$$

where the inequality follows from Supplemental Lemma A3 and from the fact that  $\mathcal{D}_{\text{single}}^{\otimes n}$  is a 1-design, which implies that  $V \rho V^\dagger$  is sampled from a low-average ensemble with purity 1. ■

The theorem states that, when we truncate the Pauli-path expansion of the Heisenberg-evolved observable at total Pauli weight  $k$ , the mean squared Frobenius error incurred by discarding all heavier paths decays exponentially in  $k$  with rate  $(1-p)^2$ . In particular, each additional unit of allowed Pauli weight suppresses the contribution of the discarded sector by a factor  $(1-p)^2$ , independently of the circuit depth  $L$  and of the particular observable  $O$  (up to its Frobenius norm). The second statement shows that the same bound controls the mean squared error of expectation values, so that truncation at weight  $k$  is sufficient to guarantee accurate prediction of measurement statistics in the noisy circuit.

## 2. Time complexity—Proof of Theorems 1 and 2

It remains to show that the weight-truncated Pauli-path summation can be evaluated efficiently. This has been already proven in Ref. [13] for circuits interspersed by depolarizing noise. While for unital noise the same proof can be adapted straightforwardly, the upper bound found in Ref. [13] does not hold anymore for the case of non-unital noise, due to the additional Pauli paths created by the noisy layers. Here, we provide a modified argument which holds under any local noise, provided that the observable  $O$  is a linear combination of polynomially many Pauli operators.

We start by providing the definition of *legal* Pauli path, which are the Fourier coefficients that bear nonzero contributions.

*Definition D1.* We say that a Pauli path  $\gamma = (P_0, P_1, \dots, P_{L+1})$  is *legal* if its Fourier coefficient gives a nonzero contribution, i.e., if it satisfies  $\langle\langle O | P_{L+1} \rangle\rangle \mathbb{E}_{C \sim \mathcal{D}_{\text{circ}}} [\Phi_\gamma^2(C)] \neq 0$ .

If the noise channel  $\mathcal{N}$  is unital, then the Fourier coefficients of all the legal Pauli paths can be efficiently enumerated. In particular, the proof of the unital case coincides with that of Lemma 8 in Ref. [13] for circuits with 2-qubit gates interleaved by local depolarizing noise. To see this, it suffices to notice that any unital single-qubit noise channel can be rewritten as  $\mathcal{N} = U \mathcal{N}' (V(\cdot) V^\dagger) U^\dagger$ , where

$$P \neq Q \implies \langle\langle P | \hat{\mathcal{N}}' | Q \rangle\rangle = 0, \quad (\text{D41})$$

for  $P, Q \in \mathcal{P}_1$ , i.e., the Pauli transfer matrix of  $\mathcal{N}'$  is diagonal. Therefore,  $\mathcal{N}'$  does not create additional Pauli paths, and thus it is covered by same proof of Ref. [13].

*Supplemental Lemma D2 (Runtime for the unital case, adapted from Lemma 8 in Ref. [13]).* If the noise is unital, then the number of legal Pauli paths with weight at most  $k$  is at most  $n^{k/L} \cdot \exp(\mathcal{O}(k))$ . Furthermore, there is an efficient algorithm to enumerate the legal paths in time  $n^{k/L} \cdot \exp(\mathcal{O}(k))$ .

However, the result discussed above does not hold if the channel  $\mathcal{N}$  is non-unital. The following observation provides an easy counter-example.

*Observation D1.* Let  $\mathcal{C}$  be a quantum circuit interspersed by amplitude-damping noise and let  $O$  be a  $k$ -local

Hamiltonian such as

$$O = \sum_{P \in \{I, Z\}^{\otimes n} : |P| < k} P. \quad (\text{D42})$$

Assume for the sake of simplicity that the circuit ends with a layer of Pauli gates. Then there are at least  $\Omega(n^{k-1})$  legal Pauli paths.

Crucially, projectors over computational basis states, e.g.,  $|0^n\rangle\langle 0^n| = 2^{-n} \sum_{P \in \{I, Z\}^{\otimes n}} P$  constitute another possible counter-example. These observations motivate the adoption of a different strategy for circuits with non-unital noise, which we formalize in the following Lemma.

*Supplemental Lemma D3 (Runtime for arbitrary noise).* Assume that each unitary layer  $U_j$  consists of non-overlapping 1- and 2-qubit gates. For any local noise, if the observable  $O$  contains  $M$  different Pauli terms, then the number of legal Pauli paths with weight at most  $k$  is at most  $M \exp(\mathcal{O}(k))$ . Furthermore, there is an efficient algorithm to enumerate the legal paths in time  $Mn \exp(\mathcal{O}(k))$ .

*Proof.* We will enumerate all the legal Pauli paths of the form  $\gamma = (P_0, P_1, \dots, P_L)$  with weight smaller than  $k$  with an iterative algorithm. First, we notice that these paths contain at most  $k$  nonidentity terms otherwise their path weights would exceed  $k$  (we recall that  $P_0$  does not contribute to the Pauli weight). Moreover, such nonidentity terms need to be consecutive, since the adjoint of any quantum channel is unital. Thus, if  $L > k$ , then the legal paths  $\gamma$  with weight smaller than  $k$  it holds that

$$P_0 = P_1 = \dots = P_{L-k+1} = I^{\otimes n}. \quad (\text{D43})$$

Therefore, we can assume without loss of generality that  $L \leq k$ .

The enumeration algorithm corresponds to a breadth-first search (BFS) of the subtree constituted by the legal paths with weight less than  $k$ . Specifically, it works by iterative forming some *partial* Pauli paths, which are vectors of the form

$$(\star, \star, \dots, \star, P_j, P_{j+1}, \dots, P_L), \quad (\text{D44})$$

where the symbol  $\star$  denotes an empty entry.

- (1) Let  $\mathbb{S}_t$  be a set of partial Pauli paths at time  $t$ . We set  $t = L$  at the beginning of the algorithm, and we decrease the counter by one at each iteration until 0.
- (2) We initialize the set  $\mathbb{S}_L$  as follows:

$$\mathbb{S}_L \leftarrow \{(\star, \star, \dots, \star, P_L) : \langle P_L | O \rangle \neq 0 \text{ and } |P_L| < k\}. \quad (\text{D45})$$

- (3) At iteration  $0 < j < L$ , for all partial paths  $(\star, \star, \dots, \star, P_{j+1}, \dots, P_L) \in \mathbb{S}_{j+1}$  we compute the Heisenberg-evolved observable  $\mathcal{C}_j^\dagger(P_{j+1})$  and we add the following paths to the set  $\mathbb{S}_j$ :

$$\mathbb{S}_j \leftarrow \{(\star, \dots, \star, P_j, P_{j+1}, \dots, P_L) : \langle P_{j+1} | \widehat{\mathcal{C}}_j | P_j \rangle \neq 0 \text{ and } |P_j| + |P_{j+1}| + \dots + |P_L| < k\}. \quad (\text{D46})$$

- (4) At iteration  $j = 0$ , for all partial paths  $(\star, P_2, \dots, P_L) \in \mathbb{S}_1$  we compute the Heisenberg-evolved observable  $\mathcal{C}_1^\dagger(P_1)$  and we add the following paths to the set  $\mathbb{S}_0$ :

$$\mathbb{S}_0 \leftarrow \{(P_0, P_1, \dots, P_L) : \langle P_1 | \widehat{\mathcal{C}}_1 | P_0 \rangle \neq 0\}. \quad (\text{D47})$$

For any fixed index  $j$ , we can upper bound the number of partial Pauli paths  $(\star, \dots, \star, P_j, \dots, P_L)$  contained in  $\mathbb{S}_j$  with the following counting argument.

- (1) We assign to the Pauli operators  $P_j, P_{j+1}, \dots, P_L$  their Pauli weights  $\ell_j, \ell_{j+1}, \dots, \ell_L$ . The number of possible choices is at most

$$\sum_{w=0}^k \left\{ \text{number of solutions to the equation } \sum_{i=j}^L |\ell_i| = w \right\} \quad (\text{D48})$$

$$= \sum_{w=0}^k \binom{w+L-j}{L-j} \leq (k+1)2^{k+L} \in \exp(\mathcal{O}(k)), \quad (\text{D49})$$

where in the first step we used the fact that the number of solutions to the equation  $\sum_{i=1}^t x_i = m$  under the constraint  $x_i \leq 0$  equals  $\binom{m+t-1}{t-1} \leq 2^{m+t-1}$ , and the last step follows because we can assume  $L \leq k$  without loss of generality.

- (2) We choose the support (positions of identities and nonidentities) for each Pauli operators  $P_j, P_{j+1}, \dots, P_L$ . We recall that  $P_L$  satisfies  $\langle\langle P_L | O \rangle\rangle \neq 0$  and  $|P_L| < k$ , i.e.,  $P_L$  is in the Pauli decomposition  $O$  and has Pauli weight less than  $k$ . Thus,  $P_L$  can take  $\mathcal{O}(\min\{n^k, M\})$  different values. For any given value of  $P_{i+1}$ , we can upper bound the number of possible choices of  $P_i$  by light-cone argument. As  $U_{i+1}$  contains only non-overlapping gates, each acting on  $c \in \mathcal{O}(1)$  qubits, and  $P_{i+1}$  has weight  $\ell_{i+1}$ , then  $P_i$  is supported on a set of  $c^{\ell_{i+1}}$  qubits. Those qubits can take values  $\{I, X, Y, Z\}$ , which yields  $(c^{\ell_{i+1}})^4 = c^{4\ell_{i+1}}$  different choices. Overall, the number of possible choices for fixed values of  $\ell_j, \ell_{j+1}, \dots, \ell_L$  is at most

$$\prod_{i=j}^L c^{4\ell_{i+1}} = c^{4\sum_{i=j}^L \ell_{i+1}} \leq c^{4k} \in \exp(\mathcal{O}(k)). \quad (\text{D50})$$

We emphasize that this argument relies solely on the fact that a layer is comprised of nonoverlapping  $\mathcal{O}(1)$ -qubit gates and does not require the circuit to be geometrically local.

- (3) For each fixed choice of the supports, we assign a value among  $X, Y, Z$  to each nonidentity single-qubit operator. The number of possible choices is at most

$$\prod_{i=j}^L 3^{\ell_{i+1}} = 3^{\sum_{i=j}^L \ell_{i+1}} \leq 3^k \in \exp(\mathcal{O}(k)). \quad (\text{D51})$$

Thus  $\mathbb{S}_j$  contains at most  $M \exp(\mathcal{O}(k))$  elements. As  $\mathcal{C}_j$  contains only non-overlapping  $\mathcal{O}(1)$ -qubit channels, all the required transition amplitudes  $\langle\langle P_{j+1} | \hat{\mathcal{C}}_j | P_j \rangle\rangle$  for  $|P_{j+1}| \leq k$  can be computed in time  $\exp(\mathcal{O}(k))$ . Moreover, the algorithm needs to store in memory each  $n$ -qubit Pauli operator, leading to an additional factor  $\mathcal{O}(n)$ . Putting all together, we find that the runtime is at most

$$Mn \exp(\mathcal{O}(k)). \quad (\text{D52})$$

■

Combining the above results, we obtain the following upper bound on the time complexity.

*Supplemental Theorem D2 (Time complexity).* Let  $\mathcal{C}$  be a noisy quantum circuit sampled from the distribution

$\mathcal{D}_{\text{circ}}$  with effective depolarizing rate  $p = p(\mathcal{N}, \mathcal{D}_{\text{circ}}) \in \Theta(1)$ .

- (1) If the measured observable  $O$  is expressed as a linear combination of  $M$  Pauli operators, then there exists a classical algorithm running in time  $M \cdot \text{poly}(n, 1/\varepsilon, 1/\delta)$  that approximates  $\text{Tr}[OC(\rho)]$  within additive error  $\varepsilon \|O\|_F$  with probability at least  $1 - \delta$  over the choice of  $\mathcal{C}$ .
- (2) If the noise channel  $\mathcal{N}$  is unital and the depth  $L$  is at least logarithmic in the number of qubits  $n$ , then there exists a classical algorithm running in time  $\text{poly}(n)$  that approximates  $\text{Tr}[OC(\rho)]$  within additive error  $\varepsilon \|O\|_F$  with probability at least  $1 - \delta$  over the choice of  $\mathcal{C}$ , for arbitrary  $\varepsilon, \delta$  decaying inverse-polynomially in  $n$ .

*Proof.* By Supplemental Theorem D1, the approximate Heisenberg-evolved observable  $O_C^{(k)}$  satisfies the following:

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[(O_C^{(k)} - C^\dagger(O))\rho]^2 \leq (1-p)^{2k} \|O\|_F^2. \quad (\text{D53})$$

Thus, Markov's inequality yields

$$\begin{aligned} \Pr_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\{ \left| \text{Tr}[(O_C^{(k)} - C^\dagger(O))\rho] \right| \geq \varepsilon \|O\|_F \right\} \\ = \Pr_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\{ \text{Tr}[(O_C^{(k)} - C^\dagger(O))\rho]^2 \geq \varepsilon^2 \|O\|_F^2 \right\} \end{aligned} \quad (\text{D54})$$

$$\leq \frac{\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[(O_C^{(k)} - C^\dagger(O))\rho]^2}{\|O\|_F^2 \varepsilon^2} \leq \frac{(1-p)^{2k}}{\varepsilon^2}. \quad (\text{D55})$$

Choosing  $k = \lceil \log(\delta^{1/2}\varepsilon) / \log(1-p) \rceil \leq \lceil p^{-1} \log(\delta^{-1/2}\varepsilon^{-1}) \rceil$ , we obtain the desired failure probability  $\delta$ , i.e.,

$$\Pr_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\{ \left| \text{Tr}[(O_C^{(k)} - C^\dagger(O))\rho] \right| \geq \varepsilon \right\} \leq \delta. \quad (\text{D56})$$

By Supplemental Lemma D3, the runtime of the Pauli propagation algorithm with path-weight cutoff  $k \in \mathcal{O}(p^{-1} \log(\delta^{-1}\varepsilon^{-1}))$  is

$$Mn \exp(\mathcal{O}(k)) = Mn \left( \frac{1}{\varepsilon\delta} \right)^{\mathcal{O}(1/p)}. \quad (\text{D57})$$

Moreover, assuming that  $L \in \Omega(\log(n))$  and the noise channel  $\mathcal{N}$  is unital, then by Supplemental Lemma D2 we can achieve  $\varepsilon \|O\|_F$  precision and with probability  $\delta$  in polynomial time, for arbitrary  $\varepsilon, \delta$  decaying inverse-polynomially in  $n$ . ■

### APPENDIX E: EFFECTIVE DEPTH BEYOND LOCAL 2-DESIGNS—PROOF OF THEOREM 3

We conclude our theoretical investigation by extending the main result of Ref. [37] to the broad classes of circuits studied in this work. Specifically, we demonstrate that, on average over the choice of random unitaries, the Frobenius norm of the traceless components of an observable undergoes exponential suppression with circuit depth under noisy Heisenberg evolution.

*Supplemental Theorem E1.* For all observables  $O$  it holds that

$$\begin{aligned} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left( \|\mathcal{C}^\dagger(O)\|_F^2 - 4^{-n} \text{Tr}[\mathcal{C}^\dagger(O)]^2 \right) \\ \leq (1-p)^{2L} \|O\|_F^2. \end{aligned} \quad (\text{E1})$$

Moreover, given two quantum states  $\rho, \sigma$ , we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[OC(\rho - \sigma)]^2 \leq 4(1-p)^{2L} \|O\|_F^2. \quad (\text{E2})$$

*Proof.* We will exploit the fact that  $\mathcal{C} = \bigcirc_{j=1}^L \mathcal{C}_j = \bigcirc_{j=1}^L \mathcal{N}_p^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{C}}_j$ . Therefore, for all observables  $O = \sum_{P \in \mathcal{P}_n} a_P P$  we have

$$\begin{aligned} \mathbb{E}_{\mathcal{C}_j} \left( \|\mathcal{C}_j^\dagger(O)\|_F^2 - 4^{-n} \text{Tr}[\mathcal{C}_j^\dagger(O)]^2 \right) \\ = \mathbb{E}_{\mathcal{C}_j} \sum_{w=1}^n \left\| \tilde{\mathcal{C}}_j^\dagger \circ \mathcal{N}_p^{(\text{depo}) \otimes n} \left( \sum_{P: |P|=w} a_P P \right) \right\|_F^2 \end{aligned} \quad (\text{E3})$$

$$= \mathbb{E}_{\mathcal{C}_j} \sum_{w=1}^n (1-p)^{2w} \left\| \tilde{\mathcal{C}}_j^\dagger \left( \sum_{P: |P|=w} a_P P \right) \right\|_F^2 \quad (\text{E4})$$

$$\leq (1-p)^2 (\|O\|_F^2 - 4^{-n} \text{Tr}[O]^2), \quad (\text{E5})$$

where the inequality follows from Supplemental Lemma B4. Iterating over all  $j$  we obtain the desired result.

The we can prove the second part of the Theorem as follows:

$$\begin{aligned} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[OC(\rho - \sigma)]^2 \\ = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \left( O - \text{Tr}[O] \frac{I^{\otimes n}}{2^n} \right) \mathcal{C}(\rho - \sigma) \right]^2 \end{aligned} \quad (\text{E6})$$

$$\begin{aligned} \leq 2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \left( O - \text{Tr}[O] \frac{I^{\otimes n}}{2^n} \right) \mathcal{C}(\rho) \right]^2 \\ + 2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} \left[ \left( O - \text{Tr}[O] \frac{I^{\otimes n}}{2^n} \right) \mathcal{C}(\sigma) \right]^2 \end{aligned} \quad (\text{E7})$$

$$\leq 4 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left( \|\mathcal{C}^\dagger(O)\|_F^2 - 4^{-n} \text{Tr}[\mathcal{C}^\dagger(O)]^2 \right) \quad (\text{E8})$$

$$\leq 4(1-p)^2 (\|O\|_F^2 - 4^{-n} \text{Tr}[O]^2), \quad (\text{E9})$$

where the second inequality follows by Supplemental Lemma A3. ■

We also obtain the following corollary for the trace distance.

*Supplemental Corollary E1.* Given two quantum states  $\rho, \sigma$ , we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \|\mathcal{C}(\rho - \sigma)\|_1^2 \leq 4^{n+1} (1-p)^{2L}. \quad (\text{E10})$$

*Proof.* We will exploit the inequality  $\|\cdots\|_1^2 \leq 2^n \|\cdots\|_2^2$  and the identity  $\|H\|_2^2 = \frac{1}{2^n} \sum_{P \in \mathcal{P}_n} \text{Tr}[HP]^2$ , which holds for all Hermitian operators  $H$ . We obtain

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \|\mathcal{C}(\rho - \sigma)\|_1^2 \leq 2^n \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \|\mathcal{C}(\rho - \sigma)\|_2^2 \quad (\text{E11})$$

$$\begin{aligned} &= \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \sum_{P \in \mathcal{P}_n} \text{Tr}[PC(\rho - \sigma)]^2 \\ &\leq 4^{n+1} (1-p)^{2L}, \end{aligned} \quad (\text{E12})$$

where we invoked Supplemental Theorem E1 in the last step. ■

A consequence of the above results is that any random noisy circuit can be truncated to an effective logarithmic depth for the task of estimating expectation values. We formalize this intuition with the following Theorem.

*Supplemental Theorem E2 (Noise-induced shallow depth).* Let  $O$  be an observable and  $\rho$  be an initial state. Consider the “truncated” noisy circuit

$$\begin{aligned} \mathcal{C}_{[L, L-j]} &:= \mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_L \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_{L-1} \circ \cdots \\ &\quad \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_{L-j}, \end{aligned} \quad (\text{E13})$$

for a suitable  $j \in \mathcal{O}(p^{-1} \log(\varepsilon^{-1} \delta^{-1}))$ . With probability at least  $1 - \delta$  over the choice of  $\mathcal{C}$ , it holds that

$$|\text{Tr}[OC(\rho)] - \text{Tr}[OC_{[L, L-j]}(\sigma)]| \leq \varepsilon \|O\|, \quad (\text{E14})$$

where  $\sigma$  is an arbitrary state.

*Proof.* The result follows as an immediate consequence of Supplemental Theorem E1. We simply replace the channel  $\mathcal{C}$  in the statement of Supplemental Theorem E1 with  $\mathcal{C}_{[L, L-j]}$ . Then we obtain that, for any two quantum

states  $\sigma, \sigma'$ ,

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} [\mathcal{OC}_{[L,L-j]}(\sigma - \sigma')]^2 \leq 4(1-p)^{2j} \|O\|^2. \quad (\text{E15})$$

Replacing  $\sigma'$  with  $\mathcal{N}^{\otimes n} \circ \mathcal{U}_{L-j-1} \circ \dots \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_1(\rho)$ , we obtain

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} [O(\mathcal{C}_{[L,L-j]}(\sigma) - \mathcal{C}(\rho))]^2 \leq 4(1-p)^{2j} \|O\|^2. \quad (\text{E16})$$

Choosing  $j = \left\lceil \frac{\log(\varepsilon)}{2 \log(1-p)} \right\rceil \leq \left\lceil \frac{\log(1/\varepsilon)}{2p} \right\rceil \in \mathcal{O}(p^{-1} \log(1/\varepsilon))$ , we obtain

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} [O(\mathcal{C}_{[L,L-j]}(\sigma) - \mathcal{C}(\rho))]^2 \leq \varepsilon \|O\|^2. \quad (\text{E17})$$

Applying Markov's inequality leads to

$$\begin{aligned} \Pr_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} (|\text{Tr} [O(\mathcal{C}_{[L,L-j]}(\sigma) - \mathcal{C}(\rho))]| \leq \varepsilon \|O\|) \\ \geq 1 - \frac{4(1-p)^{2j} \|O\|^2}{\varepsilon^2}. \end{aligned} \quad (\text{E18})$$

Therefore, choosing  $j = \left\lceil \frac{\log(\varepsilon^2 \delta)}{2 \log(1-p)} \right\rceil \leq \left\lceil \frac{\log(\varepsilon^{-2} \delta^{-1})}{2p} \right\rceil \in \mathcal{O}(p^{-1} \log(\varepsilon^{-1} \delta^{-1}))$ , it holds that

$$|\text{Tr} [\mathcal{OC}(\rho)] - \text{Tr} [\mathcal{OC}_{[L,L-j]}(\sigma)]| \leq \varepsilon \|O\|, \quad (\text{E19})$$

with probability at least  $1 - \delta$ . This concludes the proof of the Theorem. ■

## APPENDIX F: NOISE-INDUCED CONCENTRATION—PROOF OF PROPOSITION 1

In this section, we upper bound the magnitude of expectation values in typical random noisy circuits. For a traceless observable  $O$  and a sufficiently deep circuit subject to depolarizing noise, it is well known that when the depth is much larger than  $p^{-1}$ , the expectation values concentrate sharply around zero. In such regimes, one can effectively “guess zero” instead of running a classical simulation or the quantum device.

We show that a related—though weaker—phenomenon also occurs for typical circuits with arbitrary local noise, including non-unital channels.

*Supplemental Proposition F1.* For all traceless observables  $O$  it holds that

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \|\mathcal{C}^\dagger(O)\|_F^2 \leq \left[ (1-p)^{2L} + \frac{\|\mathbf{t}\|_\infty^2}{p(2-p)} \right] \|O\|_F^2. \quad (\text{F1})$$

Moreover, given a quantum states  $\rho$ , we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr} [\mathcal{OC}(\rho)]^2 \leq \left[ (1-p)^{2L} + \frac{\|\mathbf{t}\|_\infty^2}{p(2-p)} \right] \|O\|_F^2. \quad (\text{F2})$$

*Proof.* We begin by defining the following observable, which contains only the traceless components of the partially Heisenberg-evolved observables  $\mathcal{C}_j^\dagger \circ \dots \circ \mathcal{C}_L^\dagger(O)$ ,

$$\begin{aligned} O_j = \sum_{P_j, P_{j+1}, \dots, P_L \in \mathcal{P}_n \setminus \{I^{\otimes n}\}} \langle O | P_L \rangle \langle P_L | \widehat{\mathcal{C}}_L | P_{L-1} \rangle \\ \dots \langle P_{j+1} | \widehat{\mathcal{C}}_{j+1} | P_j \rangle P_j. \end{aligned} \quad (\text{F3})$$

By Supplemental Theorem E1, the expected Frobenius norm of these components decays exponentially with the number of layers they propagate through:

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \|O_{L-m}\|_F^2 \leq (1-p)^{2m} \|O\|_F^2. \quad (\text{F4})$$

Next, we express the partially Heisenberg-evolved observable by separating the identity component at each layer. For the initial observable,

$$O = \frac{\text{Tr}[O]}{2^n} I^{\otimes n} + O_L, \quad (\text{F5})$$

and applying one additional adjoint channel yields

$$\mathcal{C}_j^\dagger(O_j) = \frac{\text{Tr}[\mathcal{C}_j^\dagger(O_j)]}{2^n} I^{\otimes n} + O_{j-1}. \quad (\text{F6})$$

Iterating this decomposition over the last  $t$  layers gives

$$\text{Tr}[\mathcal{C}^\dagger(O)] = \text{Tr}[\mathcal{C}_1^\dagger \circ \dots \circ \mathcal{C}_L^\dagger(O)] \quad (\text{F7})$$

$$= \text{Tr}[\mathcal{C}_1^\dagger \circ \dots \circ \mathcal{C}_L^\dagger(O_L)] + \text{Tr}[O] \quad (\text{F8})$$

$$\begin{aligned} &= \frac{\text{Tr}[\mathcal{C}_1^\dagger \circ \dots \circ \mathcal{C}_L^\dagger(I^{\otimes n})]}{2^n} \text{Tr}[\mathcal{C}_L^\dagger(O_L)] \\ &\quad + \text{Tr}[\mathcal{C}_1^\dagger \circ \dots \circ \mathcal{C}_{L-1}^\dagger(O_{L-1})] + \text{Tr}[O] \end{aligned} \quad (\text{F9})$$

$$= \dots = \text{Tr}[O] + \sum_{j=1}^L \text{Tr}[\mathcal{C}_j^\dagger(O_j)]. \quad (\text{F10})$$

This identity expresses the full trace entirely in terms of the layerwise contributions  $\text{Tr}[\mathcal{C}_j^\dagger(O_j)]$ .

Now consider the mixed terms in the second moment. Because  $\mathcal{C}_j$  and  $\mathcal{C}_{j'}$  are independently sampled, we have for  $j \neq j'$ :

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[\mathcal{C}_j^\dagger(O_j)] \text{Tr}[\mathcal{C}_{j'}^\dagger(O_{j'})] = \left( \sum_{P_j, P_{j+1}, \dots, P_L \neq I^{\otimes n}} \langle\langle O | P_L \rangle\rangle \cdots \langle\langle P_{j+1} | \hat{\mathcal{C}}_{j+1} | P_j \rangle\rangle \text{Tr}[\mathcal{C}_j^\dagger(P_j)] \right) \quad (\text{F11})$$

$$\times \left( \sum_{P_{j'}, P_{j'+1}, \dots, P_L \neq I^{\otimes n}} \langle\langle O | P_L \rangle\rangle \cdots \langle\langle P_{j'+1} | \hat{\mathcal{C}}_{j'+1} | P_{j'} \rangle\rangle \text{Tr}[\mathcal{C}_{j'}^\dagger(P_{j'})] \right) = 0. \quad (\text{F12})$$

Thus all cross-layer correlations vanish. For the diagonal terms, expanding in the Pauli basis gives

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[\mathcal{C}_j^\dagger(O_j)]^2 = \left( \sum_{P_j} \langle\langle O_j | P_j \rangle\rangle \text{Tr}[\mathcal{C}_j^\dagger(P_j)] \right)^2 \quad (\text{F13})$$

$$= \sum_{P_j} \langle\langle O_j | P_j \rangle\rangle^2 \text{Tr}[\mathcal{C}_j^\dagger(P_j)]^2 \\ \leq \|O_j\|_F^2 \|\mathbf{t}\|_\infty^2, \quad (\text{F14})$$

where  $\|\mathbf{t}\|_\infty$  bounds the magnitude of the trace of each Pauli under  $\mathcal{C}_j^\dagger$ .

Combining the orthogonality of layers with the above bounds yields

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[\mathcal{C}^\dagger(O)]^2 = \sum_{j=1}^L \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[\mathcal{C}_j^\dagger(O_j)]^2 \\ \leq \|\mathbf{t}\|_\infty^2 \sum_{j=1}^L \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \|O_j\|_F^2 \quad (\text{F15})$$

$$\leq \|O\|_F^2 \|\mathbf{t}\|_\infty^2 \sum_{m=0}^{\infty} (1-p)^{2m} \\ = \frac{\|O\|_F^2 \|\mathbf{t}\|_\infty^2}{p(2-p)}, \quad (\text{F16})$$

where in the last step we used the fact that  $\sum_{m=0}^{\infty} (1-p)^{2m} = 1/p(2-p)$ . Finally, using Supplemental Lemma A3, we obtain

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \text{Tr}[OC(\rho)]^2 \leq \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \|\mathcal{C}^\dagger(O)\|_F^2 \\ \leq \left( (1-p)^{2L} + \frac{\|\mathbf{t}\|_\infty^2}{p(2-p)} \right) \|O\|_F^2. \quad (\text{F17})$$

This completes the proof.  $\blacksquare$

For circuit depth  $\gg p^{-1}$ , the magnitude of expectation values satisfies

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} |\text{Tr}[OC(\rho)]| \lesssim \|\mathbf{t}\|_\infty p^{-1/2} \|O\|_F. \quad (\text{F18})$$

For amplitude-damping noise, one has  $\|\mathbf{t}\|_\infty = \gamma$  and  $p = \Theta(\gamma)$ , so the bound becomes  $O(\gamma^{1/2})$ . Hence, if the target accuracy  $\varepsilon$  satisfies  $\varepsilon \gg \gamma^{1/2}$ , and the circuit is sufficiently deep, then it is again sufficient to simply “guess zero” rather than perform the classical simulation or execute the quantum circuit.

## APPENDIX G: NECESSITY OF THE RANDOM CIRCUIT ASSUMPTION

We have stated that we need to require *some* average-case assumption if we want to simulate noisy circuits for any possibly non-unital noise. However, one might question why this assumption is specifically made for random circuits rather than, for instance, random input states with fixed circuits, as done in Ref. [16]. The reason is that we can show that relying solely on the randomness in the input state is impossible, assuming  $\text{BPP} \neq \text{BQP}$ .

This can be seen by still leveraging Ref. [35], which shows how to perform fault-tolerant quantum computation in a fully quantum-coherent fashion without mid-circuit measurements and fresh auxiliary qubits through a careful compilation strategy, when the noise is non-unital. The idea is to use such a compilation strategy after waiting for a small time during which the noise drives the system (initially in its random state) toward its fixed point. We provide the detailed proof below.

*Supplemental Theorem G1.* If  $\text{BPP} \neq \text{BQP}$ , there exists no efficient classical algorithm to simulate non-unital noisy quantum circuits using fixed circuits and only a randomness assumption over the input state choice (i.e., with high probability over the input state sampled from some distribution).

*Proof.* Consider a local non-unital noise  $\mathcal{N}$ , specifically a probabilistic reset channel  $\mathcal{N}$  that, with some probability  $p > 0$ , keeps the state invariant, and with probability  $1 - p$ , resets the state in the zero computational basis state.

Assume the existence of a classical algorithm capable of estimating expectation values of observables for noisy

quantum circuits under the assumption of randomness over the input states (with high probability over the input state distribution). We demonstrate how such an algorithm can simulate fault-tolerant quantum computation.

Let  $\mathcal{C}$  be any polysize noiseless quantum circuit. We show how to classically estimate the expectation value of  $Z_1$  with arbitrary accuracy  $\varepsilon$  (thus solving a BQP-complete problem) using the following procedure:

- (1) Assume an initial input state  $\rho_0$  sampled from a distribution (the specific distribution is irrelevant to the argument).
- (2) Let the system evolve for  $L = O(\log(n/\varepsilon))$  time steps without applying gates, allowing the noise to act independently across  $n$  qubits. The resulting

state is  $\mathcal{N}_L^{\otimes n}(\rho_0)$ , where

$$\mathcal{N}_L := \underbrace{\mathcal{N} \circ \mathcal{N} \circ \dots \circ \mathcal{N}}_{L \text{ times}}.$$

- (3) Apply  $\Phi_{\mathcal{C}}$ , the fault-tolerant implementation of  $\mathcal{C}$  as described in Ref. [35]. This implementation allows estimation of  $\text{Tr}(Z_1 C(|0^n\rangle\langle 0^n|))$  with accuracy  $\varepsilon/2$ .

The classical algorithm is promised to be able to estimate the  $Z_1$  expectation value of the state

$$\rho' = \Phi_{\mathcal{C}} \circ \mathcal{N}_L^{\otimes n}(\rho_0),$$

with precision  $\varepsilon/2$  and high probability over the choice of  $\rho_0$ . We now prove that this estimate also approximates  $\text{Tr}(Z_1 C(|0^n\rangle\langle 0^n|))$  to within  $\varepsilon$ . We begin with

$$|\text{Tr}(Z_1 C(|0^n\rangle\langle 0^n|)) - \text{Tr}(Z_1 \rho')| \leq |\text{Tr}(Z_1 \Phi_{\mathcal{C}}(|0^n\rangle\langle 0^n|)) - \text{Tr}(Z_1 \rho')| + |\text{Tr}(Z_1 C(|0^n\rangle\langle 0^n|)) - \text{Tr}(Z_1 \Phi_{\mathcal{C}}(|0^n\rangle\langle 0^n|))| \quad (\text{G1})$$

$$\leq |\text{Tr}(Z_1 \Phi_{\mathcal{C}}(|0^n\rangle\langle 0^n|)) - \text{Tr}(Z_1 \rho')| + \varepsilon/2. \quad (\text{G2})$$

It suffices to show that  $|\text{Tr}(Z_1 \Phi_{\mathcal{C}}(|0^n\rangle\langle 0^n|)) - \text{Tr}(Z_1 \rho')| \leq \varepsilon/2$  to complete the proof. Using Hölder's inequality, it is enough to bound

$$\|\Phi_{\mathcal{C}} \circ \mathcal{N}_L^{\otimes n}(\rho_0) - \Phi_{\mathcal{C}}(|0^n\rangle\langle 0^n|)\|_1. \quad (\text{G3})$$

By the data-processing inequality [72], we have

$$\|\Phi_{\mathcal{C}} \circ \mathcal{N}_L^{\otimes n}(\rho_0) - \Phi_{\mathcal{C}}(|0^n\rangle\langle 0^n|)\|_1 \leq \|\mathcal{N}_L^{\otimes n}(\rho_0) - |0^n\rangle\langle 0^n|\|_1. \quad (\text{G4})$$

We can write

$$\mathcal{N}_L^{\otimes n}(\rho_0) = \text{Prob}^{(L)}(\text{All reset}) |0^n\rangle\langle 0^n| + (1 - \text{Prob}^{(L)}(\text{All reset})) \rho_{\text{rest}}, \quad (\text{G5})$$

where  $\text{Prob}^{(L)}(\text{All reset})$  is the probability that every qubit is reset at time step  $L$ , and  $\rho_{\text{rest}}$  is a quantum state representing the case in which not all qubits are reset. Then, because of triangle-inequality, we have

$$\|\mathcal{N}_L^{\otimes n}(\rho_0) - |0^n\rangle\langle 0^n|\|_1 \leq 2(1 - \text{Prob}^{(L)}(\text{All reset})) = 2\text{Prob}^{(L)}(\text{Exists one not reset}). \quad (\text{G6})$$

The probability of at least one qubit not resetting is

$$\text{Prob}^{(L)}(\text{Exists one not reset}) \leq n\text{Prob}^{(L)}(\text{1st qubit not reset}) = n(1 - p)^L. \quad (\text{G7})$$

By choosing  $L = O(\log(n/\varepsilon))$ , we ensure

$$\|\mathcal{N}_L^{\otimes n}(\rho_0) - |0^n\rangle\langle 0^n|\|_1 \leq \varepsilon/2. \quad (\text{G8})$$

This concludes the proof. ■

## APPENDIX H: ADDITIONAL NUMERICAL RESULTS

As non-unital noise (e.g., amplitude damping) can prevent barren plateaus [37,38], simulating deep noisy random circuits with non-unital noise is arguably more interesting than simulating circuits in noiseless or depolarizing

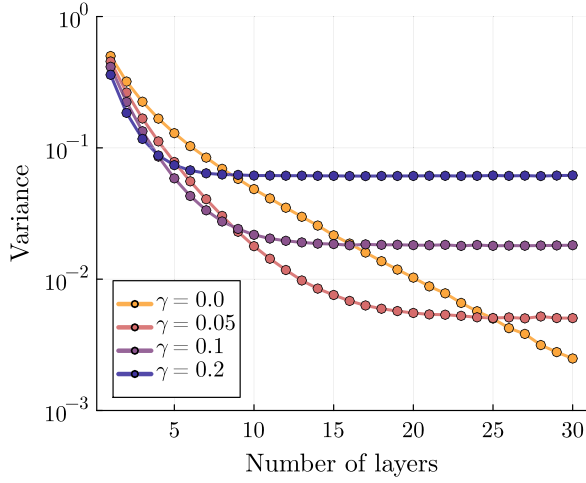


FIG. 6. Absence of barren plateaus with amplitude-damping noise. The variance, estimated via Supplemental Theorem A1 with  $10^6$  randomly sampled Pauli paths, of a Z expectation value in the middle of a 1D lattice is shown for a system of 60 qubits. Here, we used the same ansatz as in Figs. 3 and 4, specifically a 1D bricklayer topology of RX-RZ-RZZ gates. The parameter  $\gamma$  in the plot refers to the amplitude-damping rate.

scenarios, where simply outputting zero is often sufficient [26,59]. For amplitude damping (or any non-unital noise), this strategy fails due to the absence of barren plateaus, established in Ref. [37] for circuits with local 2-design gates. This leads to observable expectation values exhibiting large variances, making the zero-output approach ineffective. Our large-scale numerical simulations, shown in Fig. 6 confirm that amplitude damping induces the absence of barren plateaus in our ansatz, which is the same we used in Figs. 3 and 4 and which does not satisfy the local 2-design assumptions used in Ref. [37]. These results highlight the necessity for non-trivial classical algorithms

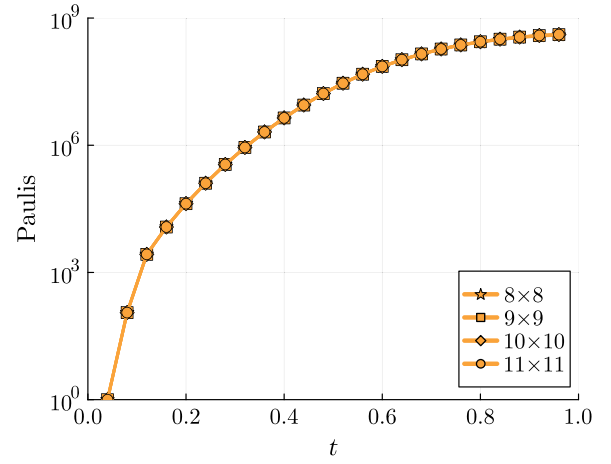


FIG. 7. Complexity scaling with the number of qubits. We show the dependence (or lack thereof) of the number of Pauli strings generated at  $k = 40$  path weight truncation.

to accurately simulate such noisy circuits, even at high depths.

We further provide a study of how the computational complexity of Fig. 5 scales with the number of qubits. In Fig. 7 we show how—in this case—the number of Pauli strings virtually does not depend on the number of qubits ranging from  $8 \times 8$  to  $11 \times 11$  grids. This is because the finite-time dynamics proposed in Ref. [60] (which we follow here) aim to probe infinite-size properties, and consequently the Pauli operator dynamics do not significantly interact with the open boundaries. The almost identical number of Pauli strings highlights that our memory complexity, and similarly, the runtime, is barely affected by the system size. Small differences in the precise number not visible on this scale are due to subtle interactions between Pauli strings with small coefficients at the boundaries and the very low coefficient truncation threshold of  $2^{-23}$ . The

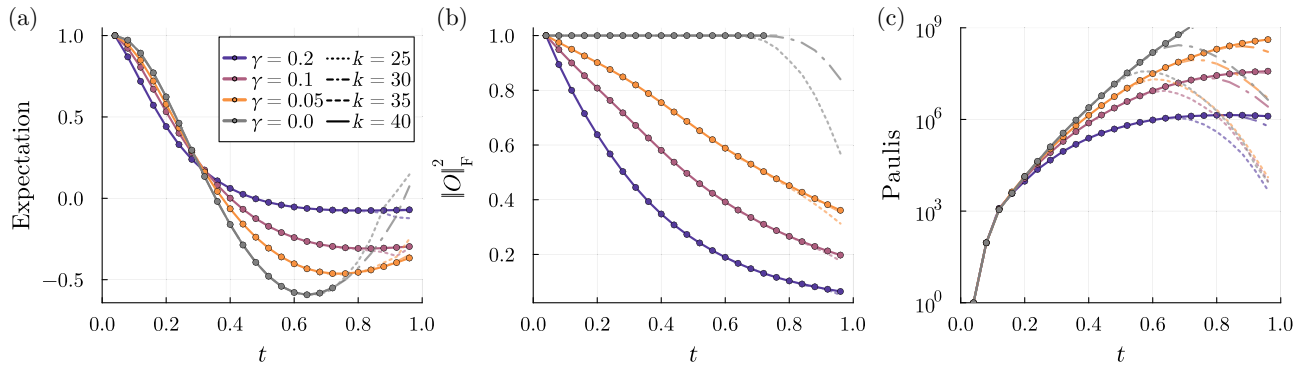


FIG. 8. Further simulation of noisy real-time quantum dynamics. We simulate a system that is a variant of Fig. 5 on an  $11 \times 11$  arrangement of 121 qubits. The evolution is still following the rotated transverse-field Ising model, but the observable is an X-Pauli in the middle of the grid instead of a Z-Pauli, and the initial state is the  $|+\rangle$  state instead of the  $|0\rangle$  state. The computational scaling is observed to be very similar.

more significant differences are that (i) at larger system sizes we apply almost twice as many gates, and (ii) we require linearly more bits to store the same number of Pauli strings, resulting in a linear factor slowdown. The same is true for sizes beyond  $11 \times 11$ .

To demonstrate that the computational scaling of our method is not drastically sensitive to the particular choice of initial state or the local observable, we replicate Fig. 5 with an initial  $|+\rangle$  state and an X-Pauli observable. See Fig. 8 for those results. Due to the initial state and the observable now being orthogonal to the effect of the amplitude-damping noise (resetting qubits to  $|0\rangle$ ), expectation values drift closer to 0, and the Frobenius norm of the evolving observable decays monotonically. However, notably, the computational resources required are very similar, and we see noisier systems converge more quickly and with fewer resources than the noise-free system.

- 
- [1] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
  - [2] X. Gao, M. Kalinowski, C.-N. Chou, M. D. Lukin, B. Barak, and S. Choi, Limitations of linear cross-entropy as a measure for quantum advantage, *PRX Quantum* **5**, 010334 (2024).
  - [3] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins *et al.*, Quantum computational advantage with a programmable photonic processor, *Nature* **606**, 75 (2022).
  - [4] C. Oh, L. Jiang, and B. Fefferman, Spoofing cross-entropy measure in boson sampling, *Phys. Rev. Lett.* **131**, 010401 (2023).
  - [5] 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* **618**, 500 (2023).
  - [6] M. S. Rudolph, E. Fontana, Z. Holmes, and L. Cincio, Classical surrogate simulation of quantum systems with LOWESA, *arXiv:2308.09109*.
  - [7] T. Begušić, J. Gray, and G. K.-L. Chan, Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance, *Sci. Adv.* **10**, eadk4321 (2024).
  - [8] J. Tindall, M. Fishman, E. M. Stoudenmire, and D. Sels, Efficient tensor network simulation of IBM's Eagle kicked Ising experiment, *PRX Quantum* **5**, 010308 (2024).
  - [9] 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).
  - [10] E. G. D. Torre and M. M. Roses, Dissipative mean-field theory of IBM utility experiment, *arXiv:2308.01339*.
  - [11] H.-J. Liao, K. Wang, Z.-S. Zhou, P. Zhang, and T. Xiang, Simulation of IBM's kicked Ising experiment with projected entangled pair operator, *arXiv:2308.03082*.
  - [12] 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).
  - [13] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2023), p. 945.
  - [14] A. Müller-Hermes, D. Stilck França, and M. M. Wolf, Relative entropy convergence for depolarizing channels, *J. Math. Phys.* **57**, 022202 (2016).
  - [15] E. Fontana, M. S. Rudolph, R. Duncan, I. Rungger, and C. Cîrstoiu, Classical simulations of noisy variational quantum circuits, *npj Quantum Inf.* **11**, 1 (2025).
  - [16] T. Schuster, C. Yin, X. Gao, and N. Y. Yao, A polynomial-time classical algorithm for noisy quantum circuits, *Phys. Rev. X* **15**, 041018 (2025).
  - [17] M. Raginsky, Scaling and renormalization in fault-tolerant quantum computers, *Quantum Inf. Process.* **2**, 249 (2003).
  - [18] C. King and M. B. Ruskai, Minimal entropy of states emerging from noisy quantum channels, *IEEE Trans. Inf. Theory* **47**, 192 (2001).
  - [19] C. King, The capacity of the quantum depolarizing channel, *IEEE Trans. Inf. Theory* **49**, 221 (2003).
  - [20] B. M. Terhal and G. Burkard, Fault-tolerant quantum computation for local non-Markovian noise, *Phys. Rev. A: At. Mol. Opt. Phys.* **71**, 012336 (2005).
  - [21] S. Khatri, K. Sharma, and M. M. Wilde, Information-theoretic aspects of the generalized amplitude-damping channel, *Phys. Rev. A* **102**, 012401 (2020).
  - [22] D. García-Martín, M. Larocca, and M. Cerezo, Effects of noise on the overparametrization of quantum neural networks, *Phys. Rev. Res.* **6**, 013295 (2024).
  - [23] E. Fontana, M. Cerezo, A. Arrasmith, I. Rungger, and P. J. Coles, Non-trivial symmetries in quantum landscapes and their resilience to quantum noise, *Quantum* **6**, 804 (2022).
  - [24] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of noisy reversible computation, *arXiv:quant-ph/9611028*.
  - [25] C. Hirche, C. Rouzé, and D. Stilck França, On contraction coefficients, partial orders and approximation of capacities for quantum channels, *Quantum* **6**, 862 (2022).
  - [26] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, *Nat. Commun.* **12**, 1 (2021).
  - [27] S. Wang, P. Czarnik, A. Arrasmith, M. Cerezo, L. Cincio, and P. J. Coles, Can error mitigation improve trainability of noisy variational quantum algorithms? *Quantum* **8**, 1287 (2024).
  - [28] P. Rall, D. Liang, J. Cook, and W. Kretschmer, Simulation of qubit quantum circuits via Pauli propagation, *Phys. Rev. A* **99**, 062337 (2019).
  - [29] Y. Shao, F. Wei, S. Cheng, and Z. Liu, Simulating noisy variational quantum algorithms: A polynomial approach, *Phys. Rev. Lett.* **133**, 120603 (2024).

- [30] G. González-García, J. I. Cirac, and R. Trivedi, Pauli path simulations of noisy quantum circuits beyond average case, *Quantum* **9**, 1730 (2025).
- [31] C. Cirstoiu, A Fourier analysis framework for approximate classical simulations of quantum circuits, [arXiv:2410.13856](#).
- [32] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).
- [33] L. Chirolli and G. Burkard, Decoherence in solid-state qubits, *Adv. Phys.* **57**, 225 (2008).
- [34] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. Allman, C. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer *et al.*, Demonstration of the trapped-ion quantum CCD computer architecture, *Nature* **592**, 209 (2021).
- [35] M. Ben-Or, D. Gottesman, and A. Hassidim, Quantum refrigerator, [arXiv:1301.1995](#).
- [36] B. Fefferman, S. Ghosh, M. Gullans, K. Kuroiwa, and K. Sharma, Effect of nonunitary noise on random-circuit sampling, *PRX Quantum* **5**, 030317 (2024).
- [37] A. A. Mele, A. Angrisani, S. Ghosh, S. Khatir, J. Eisert, D. Stilck França, and Y. Quek, Noise-induced shallow circuits and absence of barren plateaus, *Nat. Phys.* **1** (2026).
- [38] P. Singkanipa and D. A. Lidar, Beyond unitary noise in variational quantum algorithms: noise-induced barren plateaus and limit sets, *Quantum* **9**, 1617 (2025).
- [39] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, Classically estimating observables of noiseless quantum circuits, *Phys. Rev. Lett.* **135**, 170602 (2025).
- [40] A. Deshpande, M. Hinsche, S. Najafi, K. Sharma, R. Sweke, and C. Zoufal, Dynamic parameterized quantum circuits: Expressive and barren-plateau free, [arXiv:2411.05760](#).
- [41] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 1 (2014).
- [42] 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).
- [43] H.-Y. Huang, S. Chen, and J. Preskill, Learning to predict arbitrary quantum processes, *PRX Quantum* **4**, 040337 (2023).
- [44] O. Fawzi, A. Müller-Hermes, and A. Shayeghi, in *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2022).
- [45] Y. Quek, D. Stilck França, S. Khatir, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, *Nat. Phys.* **20**, 1648 (2024).
- [46] G. Crognaletti, M. Grossi, and A. Bassi, Estimates of loss function concentration in noisy parametrized quantum circuits, [arXiv:2410.01893](#).
- [47] V. Martinez, A. Angrisani, E. Pankovets, O. Fawzi, and D. Stilck França, Efficient simulation of parametrized quantum circuits under nonunitary noise through Pauli backpropagation, *Phys. Rev. Lett.* **134**, 250602 (2025).
- [48] T. Begušić, K. Hejazi, and G. K. Chan, Simulating quantum circuit expectation values by Clifford perturbation theory, *J. Chem. Phys.* **162**, 154110 (2025).
- [49] S. Lerch, R. Puig, M. Rudolph, A. Angrisani, T. Jones, M. Cerezo, S. Thanasilp, and Z. Holmes, Efficient quantum-enhanced classical simulation for patches of quantum landscapes, [arXiv:2411.19896](#).
- [50] R. Zhang, Y. Shao, F. Wei, S. Cheng, Z. Wei, and Z. Liu, Clifford perturbation approximation for quantum error mitigation, [arXiv:2412.09518](#).
- [51] K. Mitarai, Y. Suzuki, W. Mizukami, Y. O. Nakagawa, and K. Fujii, Quadratic Clifford expansion for efficient benchmarking and initialization of variational quantum algorithms, *Phys. Rev. Res.* **4**, 033012 (2022).
- [52] P. Bermejo, P. Braccia, M. S. Rudolph, Z. Holmes, L. Cincio, and M. Cerezo, Quantum convolutional neural networks are (effectively) classically simulable, *PRX Quantum* **7**, 020304 (2026).
- [53] A. Tanggara, M. Gu, and K. Bharti, Classically spoofing system linear cross entropy score benchmarking, [arXiv:2405.00789](#).
- [54] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp *et al.*, Does provable absence of barren plateaus imply classical simulability? *Nat. Commun.* **16**, 7907 (2025).
- [55] The results in Ref. [39] were originally established for noiseless circuits but can be extended straightforwardly to noisy circuits.
- [56] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).
- [57] See <https://github.com/MSRudolph/PauliPropagation.jl> for PauliPropagation.jl package.
- [58] M. S. Rudolph, T. Jones, Y. Teng, A. Angrisani, and Z. Holmes, Pauli propagation: A computational framework for simulating quantum systems, [arXiv:2505.21606](#).
- [59] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 1 (2018).
- [60] T. Begušić and G. K.-L. Chan, Real-time operator evolution in two and three dimensions via sparse Pauli dynamics, *PRX Quantum* **6**, 020302 (2025).
- [61] H. W. J. Blöte and Y. Deng, Cluster Monte Carlo simulation of the transverse Ising model, *Phys. Rev. E* **66**, 066110 (2002).
- [62] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [63] Z. Zhou, R. Sitler, Y. Oda, K. Schultz, and G. Quiroz, Quantum crosstalk robust quantum control, *Phys. Rev. Lett.* **131**, 210802 (2023).
- [64] J. Tuziemski, F. B. Maciejewski, J. Majsak, O. Slowik, M. Kotowski, K. Kowalczyk-Murynka, P. Podziemski, and M. Oszmaniec, Efficient reconstruction, benchmarking and validation of cross-talk models in readout noise in near-term quantum devices, *Open Syst. Inf. Dyn.* **32**, 2550019 (2025).
- [65] J. Watrous, *The Theory of Quantum Information* (Cambridge University, Cambridge, UK, 2018).

- [66] A. A. Mele, Introduction to Haar measure tools in quantum information: A beginner's tutorial, [Quantum](#) **8**, 1340 (2024).
- [67] K. Bu, D. E. Koh, R. J. Garcia, and A. Jaff , Classical shadows with Pauli-invariant unitary ensembles, [npj Quantum Inf.](#) **10**, 6 (2024).
- [68] W.-T. Kuo, A. A. Akhtar, D. P. Arovas, and Y.-Z. You, Markovian entanglement dynamics under locally scrambled quantum evolution, [Phys. Rev. B](#) **101**, 224202 (2020).
- [69] H.-Y. Hu, S. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, [Phys. Rev. Res.](#) **5**, 023027 (2023).
- [70] M. Beth Ruskai, S. Szarek, and E. Werner, An analysis of completely-positive trace-preserving maps on M2, [Linear Algebra Appl.](#) **347**, 159 (2002).
- [71] R. Bhatia, *Positive Definite Matrices* (Princeton University, Princeton, New Jersey, USA, 2015).
- [72] M. M. Wilde, *Quantum Information Theory* (Cambridge University, Cambridge, UK, 2013).