Simulating quantum circuits with arbitrary local noise using Pauli Propagation

Armando Angrisani, 1,* Antonio A. Mele, 2,3,† Manuel S. Rudolph, 1 M. Cerezo, 4 and Zoë Holmes 1 Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland 2 Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany 3 Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA 4 Information Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

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 6×6 lattice of qubits under the effects of amplitude damping and dephasing noise, as well as real-time dynamics on an 11×11 lattice of qubits affected by amplitude damping.

I. INTRODUCTION

Recent years have seen a healthy back-and-forth between experimental teams attempting to implement non-classically 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 towards 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 towards the maximally mixed state. This makes any non-trivial 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

^{*} armando.angrisani@epfl.ch

[†] a.mele@fu-berlin.de

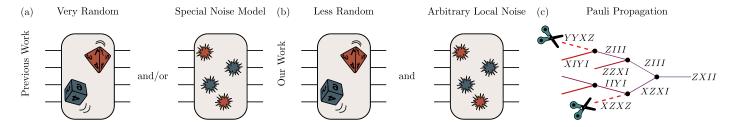


Figure 1. Schematic summary of our main results. a) Prior work has provided guarantees for simulating a quantum circuit either for highly random quantum circuits or assuming a special noise model. b) Here we show that the combination of a very general noise model (arbitrary local incoherent noise) with 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 Paulis operators on a given branch [13, 16, 30] in contrast to *current* Pauli weight [6, 16, 39].

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 towards 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-circuits 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:

- 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-unitary 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.
- 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 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.

• 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 unitary 1-designs. For dephasing noise, we assume the gates satisfy an approximate 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 unitary 2-designs.

Finally, we complement our theoretical investigation with numerical simulations of a periodic 6×6 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 11×11 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

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

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

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

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

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

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

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

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 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.

- Depolarizing-like noise: This corresponds to unital channels satisfying $\|\mathbf{D}\|_{\infty} < 1$, i.e., to noise channels that drive any input state towards the maximally mixed state. The most prominent example of depolarizing-like noise is the depolarizing channel, where $D_X = D_Y = D_Z$.
- 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 towards such diameter.

Arbitrary Local Noise

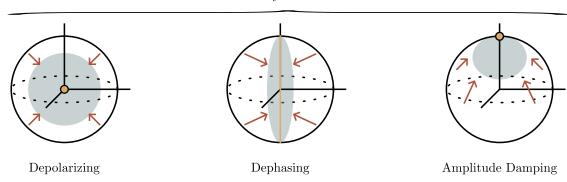


Figure 2. **Overview of noise models.** Here we provide a geometric sketch of three commonly considered families of local noise: a) Depolarizing noise, which drives any input state towards the maximally mixed state, b) Dephasing-like noise, which drives states to z-axis connecting poles of the Bloch 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.

• 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 lead the quantum state towards a *fixed point* different than maximally mixed state. A common example of non-unital noise is the amplitude-damping channel, characterized by $\mathbf{D} = (\sqrt{1-\gamma}, \sqrt{1-\gamma}, 1-\gamma)$ 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, ρ is an arbitrary initial n-qubit quantum state, O is a bounded Hermitian observable such that $||O|| \leq 1$ (where ||...|| denotes the spectral norm), and \mathcal{C} is an L-layered noisy random circuit of the form

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

Above, the final layer $\mathcal{V}^{\text{single}}$ consists of single-qubit gates, 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 $\mathcal{U}_1, \ldots, \mathcal{U}_L$ 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.

- Locally unbiased circuits: If the single-qubit gates are sampled from 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.
- Approximately locally scrambling circuits: If the single-qubit gates are sampled from 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.

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 significant 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 sub-logarithmic 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 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 poly-logarithmic 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]. Refs. [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].

Local noise model	Average-case assumptions	Runtime
Noiseless	locally scrambling layers	$n^{\mathcal{O}(\log(n))}$ [39]
Depolarizing	random input state	poly(n) [16]
Spontaneous emission	emissions occurring in random directions, random input state	$M \cdot \text{poly}(n)$ [16]
Depolarizing-like [15, 29], amplitude damping [47]	circuit with Clifford gates and random Pauli rotations	$M \cdot \operatorname{poly}(n)$
Arbitrary non-unital noise	locally unbiased layers	$M \cdot \text{poly}(n)$ [Thm. 1]
Arbitrary unital noise (including dephasing-like)	approximately locally scrambling layers	poly(n) [Thm. 2]

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.

IV. RESULTS

Theoretical guarantees. 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 Section 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 the previous section.

Theorem 1 (Non-unital noise, informal). Let O be an observable expressed as linear combination of M Pauli operators, and let ρ be an initial state. Assume the Pauli coefficients of O and ρ 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 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 $M \cdot \text{poly}(n, \epsilon^{-1}, \delta^{-1})$ within error $\epsilon ||O||$ and with probability $1 - \delta$ over the circuit randomness.

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 4). 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 ρ be an initial state. Assume the Pauli coefficients of O and ρ 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. Then the noisy expectation value of any observable O can be estimated classically in time $\operatorname{poly}(n, \epsilon^{-1}, \delta^{-1})$ within error $\epsilon ||O||$ and with probability $1 - \delta$ over the circuit randomness.

Finally, we prove 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 \mathcal{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 observables and consider an L-layered circuit C randomly sampled from \mathcal{D}_{circ} where \mathcal{D}_{circ} is a distribution over noisy circuits satisfying the assumptions of Theorems 1 or 2. Consider the "truncated" noisy circuit

$$\mathcal{C}_{[L,L-j]} := \mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_L \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_{L-1} \circ \cdots \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_{L-j}, \tag{7}$$

for a suitable $j \in \mathcal{O}(\log(1/\epsilon))$. With high probability over the choice of \mathcal{C} , it holds that

$$\left| \text{Tr}[OC(\rho)] - \text{Tr}\left[OC_{[L,L-j]}(\sigma)\right] \right| \leqslant \epsilon ||O||, \tag{8}$$

where σ is an arbitrary state.

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 [54]. However, the bounds derived in this paper are significantly stronger, as they establish a polynomial runtime for noisy circuits across all architectures.

Numerical implementation. We complement our theoretical analysis with precise numerical estimates of the average simulation error for a Hamiltonian variational ansatz [55] and a large-scale simulation of noisy quantum dynamics. All numerical results were collected with the PauliPropagation.jl package.

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 6×6 latice on 36 qubits. This corresponds to an 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, 56], due to the fact that such circuits can experience 'absence of barren plateaus' [37], highlighting the need for non-trivial classical algorithms to tackle such cases. In Appendix E, 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 Refs. [37].

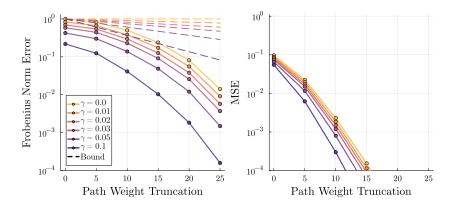


Figure 3. Error of simulating circuits with amplitude damping noise. a) Frobenius norm error and b) mean square error (MSE) for different path-weight truncation values and noise strengths in a periodic 6×6 lattice of qubits. The observable is a Pauli-Z operator in the middle of the lattice with a circuit 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 Thm. 5, 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 5.

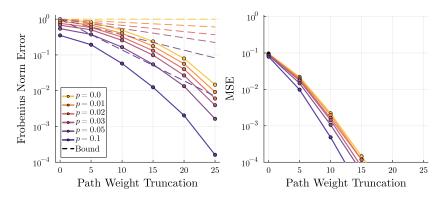


Figure 4. **Error of simulating circuits with dephasing noise**. a) Frobenius norm error and b) mean square error (MSE) for the same 36-qubit simulation as in Fig. 3. The theoretical bound, established in Thm. 5, is calculated via $\left(\frac{1+(1-2p)^2}{2}\right)^k$, where k is the path-weight truncation order and $\frac{1+(1-2p)^2}{2}$ is the mean squared contraction coefficient computed in Example 5.

Our bounds on the Frobenius norm error of the backpropagated observable as a function of the path-weight truncation order are well satisfied (see panels (a)). This norm also indirectly bounds the error of expectation values with random initial states. We show, however, that the mean square error (MSE) of expectation values with the ubiquitous all-zero initial state is orders of magnitude lower at the same path-weight truncation order (see panels (b)). These error estimates were obtained via the Monte Carlo approach provided in Theorem 4 in the Appendices. 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. [57], which compares the converging Pauli propagation results to iPEPS simulation data. Namely, we consider an 11×11 square qubit lattice with a rotated transverse-field Ising Hamiltonian,

$$H = -J \sum_{\langle i,j \rangle} X_i X_j - h \sum_i Z_i \,, \tag{9}$$

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 [58]. 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.

Ref. [57] 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 same XY-Pauli truncation with the most challenging coefficient truncation threshold of 2^{-23} , which brings the noise-free system close to apparent convergence with more than 1TB of memory usage. Then, we insert local amplitude damping noise into the quantum circuit and simulate the evolution with path-weight truncation on a laptop.

To probe whether our substantially less resource-intensive noisy simulations are reliable we continually loosen the algorithmic truncations and observe convergence to expectation values. Based on the results with path-weight truncation values of 20 (dotted) and 25 (dashed) in Fig. 5, which diverge at time t = 0.4 and t = 0.6, respectively, we believe that, given the exponential convergence present in Fig. 3, the simulation of the least-noisy system with $\gamma = 0.05$ with pathweight truncation of 30 can be trusted to approximately time t = 0.8. Our peak memory usage was around 12GB, and the full simulation took approximately one hour. However, systems affected by more noise ($\gamma = 0.1$ and 0.2) appear to require lower truncation values and are likely more accurate until the final time of t = 0.92.

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. 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.

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 $P = P^{(1)} \otimes P^{(2)} \otimes \cdots \otimes P^{(n)} \in \mathcal{P}_n$, we define its support, denoted as $\operatorname{supp}(P)$, as the subset of qubits upon which P acts non-trivially, and we define its Pauli weight, denoted as |P| as the cardinality of its support or, equivalently, as the number of non-identity 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

$$||O||_{\mathcal{F}}^2 := \sum_{P \in \mathcal{P}_n} a_P^2. \tag{10}$$

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

$$C := \underbrace{\mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_{L}}_{\mathcal{C}_{L}} \circ \underbrace{\mathcal{N}^{\otimes n} \circ \mathcal{U}_{L-1}}_{\mathcal{C}_{L-1}} \circ \cdots \circ \underbrace{\mathcal{N}^{\otimes n} \circ \mathcal{U}_{1}}_{\mathcal{C}_{1}} = \mathcal{C}_{L} \circ \mathcal{C}_{L-1} \circ \cdots \circ \mathcal{C}_{1}. \tag{11}$$

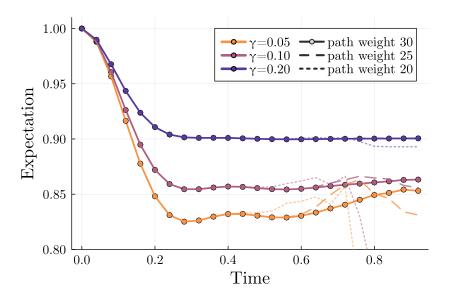


Figure 5. Simulation of noisy real-time quantum dynamics. We simulate the expectation value of a Pauli-Z observable in the middle of an 11×11 arrangement of 121 qubits evolving under a rotated transverse-field Ising model at a quantum critical point. In addition to the path-weight truncations 20, 25 and 30, we truncate small coefficients below 2^{-23} and propagating Pauli operators with more than 5 X or Y Paulis. This setup was used in Ref. [57] to accurately simulate the noise-free case with significantly more computational resources.

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 Appendices.

Our technique consists in computing an approximate Heisenberg evolution of the observable via a truncated Paulipath 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 $C^{\dagger}(O)$ as a sum across Pauli paths as follows

$$C^{\dagger}(O) = \sum_{P_0, P_1, \dots, P_L \in \mathcal{P}_n} \langle \langle O|P_L \rangle \rangle \langle \langle P_L|\widehat{\mathcal{C}}_L|P_{L-1} \rangle \rangle \langle \langle P_{L-1}|\widehat{\mathcal{C}}_{L-1}|P_{L-2} \rangle \dots \langle \langle P_1|\widehat{\mathcal{C}}_1|P_0 \rangle \rangle P_0, \tag{12}$$

where we employed the normalized vectorized notation $\langle\!\langle O|P_L\rangle\!\rangle := \frac{1}{2^n} \text{Tr}[OP_L], \ \langle\!\langle P_j|\widehat{\mathcal{C}}_j|P_{j-1}\rangle\!\rangle := \frac{1}{2^n} \text{Tr}[P_j\mathcal{C}_j(P_{j-1})].$ We can also rearrange Eq. (12) as follows by introducing a more compact notation:

$$C^{\dagger}(O) := \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle \!\langle O | P_L \rangle \!\rangle \, \Phi_{\gamma}(C) P_0, \tag{13}$$

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

$$|\gamma| := \sum_{j=1}^{L} |P_j|. \tag{14}$$

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

$$C_k^{\dagger}(O) := \sum_{\substack{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1} \\ |\gamma| < k}} \langle\!\langle O | P_L \rangle\!\rangle \, \Phi_{\gamma}(\mathcal{C}) P_0. \tag{15}$$

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

The Pauli propagation algorithm.

The approximate Heisenberg evolved observable $C_k^{\dagger}(O)$ can be computed with an iterative process. Assume without loss of generality that O consists in 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 $C_L^{\dagger}(O)$ in the Pauli basis:

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

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. (15), 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 Ref. [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 $C_k^{\dagger}(O)$ 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:

$$\mathcal{N}_p^{(\text{depo})\otimes n}(P) = \mathcal{N}_p^{\dagger(\text{depo})\otimes n}(P) = (1-p)^{|P|}P,\tag{16}$$

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 $\left\|\mathcal{N}_p^{(\text{depo})\dagger\otimes n}(P)\right\|_{\mathrm{F}}=(1-p)^{|P|}$. Moreover, the Frobenius norm of an observable $O=\sum_{P\in\mathcal{P}_n}a_PP$ takes the following form after the application of the depolarizing channel

$$\left\| \mathcal{N}_p^{\dagger(\text{depo}) \otimes n}(O) \right\|_{\mathcal{F}}^2 = \sum_{P \in \mathcal{P}_n} a_P^2 (1-p)^{2|P|}. \tag{17}$$

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 Lemma 9, when averaged over random gates, any incoherent noise channel contracts the Frobenius norm of Pauli operators under Heisenberg evolution, as we have

$$\mathbb{E}_{V \sim \mathcal{D}} \| \mathcal{N}^{\dagger \otimes n}(V^{\dagger} P V) \|_{\mathcal{F}}^{2} \leqslant \left(\chi_{\mathcal{D}}^{2}(\mathcal{N}) \right)^{|P|} \tag{18}$$

where $\chi^2_{\mathcal{D}}(\mathcal{N}) := \frac{\|\mathbf{D}\|_2^2 + \|\mathbf{t}\|_2^2}{3}$ 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

$$\mathbb{E}_{V \sim \mathcal{D}} \left\| \mathcal{N}^{\dagger \otimes n}(V^{\dagger} O V) \right\|_{\mathcal{F}}^{2} \leqslant \sum_{P \in \mathcal{P}_{n}} a_{P}^{2} \left(\chi_{\mathcal{D}}^{2}(\mathcal{N}) \right)^{|P|}, \tag{19}$$

which also matches the effect of local depolarizing noise of Eq. (17).

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:

$$\mathcal{N} := \mathcal{N}_p^{(\text{depo})} \circ \tilde{\mathcal{N}}_p, \tag{20}$$

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

$$p := 1 - \chi_{\mathcal{D}}(\mathcal{N}). \tag{21}$$

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

$$\mathbb{E}\left|\operatorname{Tr}\left[\left(\mathcal{C}_{k}^{\dagger}(O)-\mathcal{C}^{\dagger}(O)\right)\rho\right]\right| \in \exp(-\Omega\left(k\right)). \tag{22}$$

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 $C_k^{\dagger}(O)$ can be computed in poly(n)-time provided that $k \in \mathcal{O}(\log(n))$ and the observable O is a linear combination of at most 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.

• 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 $poly(n^{-1})||H||_F$ 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

$$H = \sum_{\substack{P \in \{I,Z\}^{\otimes n} \\ |P| < \ell}} P. \tag{23}$$

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

• 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].

- 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 [59]. 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.
- Alternative noise models. Despite its generality, our noise models also comes 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 [60, 61]). The classical simulability of such noisy circuits is an important direction for future research.

VII. CODE AVAILABILITY

The numerical simulations in this work were performed with the open-source PauliPropagation.jl package. The extensions necessary to employ path-weight truncation will soon be publicly available, but can be shared upon request.

VIII. ACKNOWLEDGMENTS

The authors thank Lennart Bittel, Sumeet Khatri, Victor Martinez, Thomas Schuster, Daniel Stilck França and Chu Zhao for valuable discussions and feedback. AA and ZH acknowledge support from the Sandoz Family Foundation-Monique de Meuron program for Academic Promotion. AAM acknowledges support by the German Federal Ministry for Education and Research (BMBF) under the project FermiQP. AAM 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 number 20230049DR. MSR acknowledges funding from the 2024 Google PhD Fellowship. MC 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. MR and ZH acknowledge support of the NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 205602).

^[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick,

- A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, 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, Physical Review Letters 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 preprint arXiv:2308.09109 (2023).
- [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, Science Advances 10, 10.1126/sciadv.adk4321 (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, arXiv preprint arXiv:2306.15970 https://doi.org/10.48550/arXiv.2306.15970 (2023).
- [10] E. G. D. Torre and M. M. Roses, Dissipative mean-field theory of ibm utility experiment, arXiv preprint arXiv:2308.01339 https://doi.org/10.48550/arXiv.2308.01339 (2023).
- [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 preprint arXiv:2308.03082 https://doi.org/10.48550/arXiv.2308.03082 (2023).
- [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, arXiv preprint arXiv:2308.05077 https://doi.org/10.48550/arXiv.2308.05077 (2023).
- [13] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, A polynomial-time classical algorithm for noisy random circuit sampling, Proceedings of the 55th Annual ACM Symposium on Theory of Computing, 945 (2023).
- [14] A. Müller-Hermes, D. Stilck França, and M. M. Wolf, Relative entropy convergence for depolarizing channels, Journal of Mathematical Physics 57, 022202 (2016).
- [15] E. Fontana, M. S. Rudolph, R. Duncan, I. Rungger, and C. Cîrstoiu, Classical simulations of noisy variational quantum circuits, arXiv preprint arXiv:2306.05400 (2023).
- [16] T. Schuster, C. Yin, X. Gao, and N. Y. Yao, A polynomial-time classical algorithm for noisy quantum circuits, arXiv preprint arXiv:2407.12768 https://doi.org/10.48550/arXiv.2407.12768 (2024).
- [17] M. Raginsky, Scaling and renormalization in fault-tolerant quantum computers, Quantum Information Processing 2, 249 (2003).
- [18] C. King and M. Ruskai, Minimal entropy of states emerging from noisy quantum channels, IEEE Trans. Inf. Th. 47, 192–209 (2001).
- [19] C. King, The capacity of the quantum depolarizing channel, IEEE Transactions on Information Theory 49, 221 (2002).
- [20] B. M. Terhal and G. Burkard, Fault-tolerant quantum computation for local non-markovian noise, Physical Review A—Atomic, Molecular, and Optical Physics 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 preprint quant-ph/9611028 (1996).
- [25] C. Hirche, C. Rouzé, and D. S. 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, Nature Communications 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, Physical Review A 99, 062337 (2019).

- [29] Y. Shao, F. Wei, S. Cheng, and Z. Liu, Simulating quantum mean values in noisy variational quantum algorithms: A polynomial-scale approach, arXiv preprint arXiv:2306.05804 (2023).
- [30] G. González-García, J. I. Cirac, and R. Trivedi, Pauli path simulations of noisy quantum circuits beyond average case, arXiv preprint arXiv:2407.16068 (2024).
- [31] C. Cirstoiu, A fourier analysis framework for approximate classical simulations of quantum circuits, arXiv preprint arXiv:2410.13856 10.48550/arXiv.2410.13856 (2024).
- [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, Advances in Physics 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 preprint arXiv:1301.1995 (2013).
- [36] B. Fefferman, S. Ghosh, M. Gullans, K. Kuroiwa, and K. Sharma, Effect of non-unital noise on random circuit sampling, arXiv preprint arXiv:2306.16659 (2023).
- [37] A. A. Mele, A. Angrisani, S. Ghosh, S. Khatri, J. Eisert, D. S. França, and Y. Quek, Noise-induced shallow circuits and absence of barren plateaus, arXiv preprint arXiv:2403.13927 (2024).
- [38] P. Singkanipa and D. A. Lidar, Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and fixed points, arXiv preprint arXiv:2402.08721 (2024).
- [39] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, Classically estimating observables of noiseless quantum circuits, arXiv preprint arXiv:2409.01706 (2024).
- [40] A. Deshpande, M. Hinsche, S. Najafi, K. Sharma, R. Sweke, and C. Zoufal, Dynamic parameterized quantum circuits: expressive and barren-plateau free, arXiv preprint arXiv:2411.05760 10.48550/arXiv.2411.05760 (2024).
- [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, Nature Communications 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, Nature Communications 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, A lower bound on the space overhead of fault-tolerant quantum computation, in 13th Innovations in Theoretical Computer Science Conference (ITCS 2022) (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022).
- [45] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, arXiv preprint arXiv:2210.11505 (2022).
- [46] G. Crognaletti, M. Grossi, and A. Bassi, Estimates of loss function concentration in noisy parametrized quantum circuits, arXiv preprint arXiv:2410.01893 (2024).
- [47] V. Martinez, A. Angrisani, E. Pankovets, O. Fawzi, and D. Stilck França, Efficient simulation of parametrized quantum circuits under non-unital noise through pauli backpropagation (2025).
- [48] T. Begušić, K. Hejazi, and G. K. Chan, Simulating quantum circuit expectation values by Clifford perturbation theory, arXiv preprint arXiv:2306.04797 (2023).
- [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 preprint arXiv:2411.19896 10.48550/arXiv.2411.19896 (2024).
- [50] R. Zhang, Y. Shao, F. Wei, S. Cheng, Z. Wei, and Z. Liu, Clifford perturbation approximation for quantum error mitigation, arXiv preprint arXiv:2412.09518 (2024).
- [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, Physical Review Research 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, arXiv preprint arXiv:2408.12739 (2024).
- [53] A. Tanggara, M. Gu, and K. Bharti, Classically spoofing system linear cross entropy score benchmarking, arXiv preprint arXiv:2405.00789 (2024).
- [54] The results in Ref. [39] were originally established for noiseless circuits but can be extended straightforwardly to noisy circuits.
- [55] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Physical Review A 92, 042303 (2015).
- [56] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications 9, 1 (2018).
- [57] T. Begušić and G. K. Chan, Real-time operator evolution in two and three dimensions via sparse pauli dynamics, arXiv preprint arXiv:2409.03097 (2024).
- [58] H. W. J. Blöte and Y. Deng, Cluster monte carlo simulation of the transverse Ising model, Phys. Rev. E 66, 066110 (2002).

- [59] 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, Nature Reviews Physics 3, 625–644 (2021).
- [60] Z. Zhou, R. Sitler, Y. Oda, K. Schultz, and G. Quiroz, Quantum crosstalk robust quantum control, Physical Review Letters 131, 210802 (2023).
- [61] 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, arXiv preprint arXiv:2311.10661 10.48550/arXiv.2311.10661 (2023).
- [62] J. Watrous, The Theory of Quantum Information (Cambridge University Press, 2018).
- [63] A. A. Mele, Introduction to haar measure tools in quantum information: A beginner's tutorial, Quantum 8, 1340 (2024).
- [64] W.-T. Kuo, A. Akhtar, D. P. Arovas, and Y.-Z. You, Markovian entanglement dynamics under locally scrambled quantum evolution, Physical Review B 101, 224202 (2020).
- [65] H.-Y. Hu, S. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, Physical Review Research 5, 023027 (2023).
- [66] H.-Y. Huang, S. Chen, and J. Preskill, Learning to predict arbitrary quantum processes, arXiv preprint arXiv:2210.14894 (2022).
- [67] M. Beth Ruskai, S. Szarek, and E. Werner, An analysis of completely-positive trace-preserving maps on M2, Linear Algebra and its Applications 347, 159 (2002).
- [68] R. Bhatia, *Positive Definite Matrices* (Princeton University Press, 2015).

Appendix

Appendix A: Preliminaries

1. Notation

We briefly introduce the notations employed in this paper. Given a positive integer n, we denote the set $[n] := \{1, 2, ..., n\}$. Given a distribution \mathcal{D} over the set \mathcal{X} and a measurable function $F: \mathcal{X} \to \mathbb{R}$, 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 vec : $\mathcal{B}(\mathcal{H}_n) \to \mathcal{H}_n^{\otimes 2}$, 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 $A = \sum_{i,j \in \{0,1\}^n} A_{i,j} |i\rangle\langle j|$, then the vectorization of A is given by $\text{vec}(A) = \sum_{i,j \in \{0,1\}^n} A_{i,j} |i\rangle \otimes |j\rangle$. We also adopt notation $|A\rangle := \frac{1}{\sqrt{2^n}} \text{vec}(A)$, 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:

$$|ABC\rangle\rangle = A \otimes C^T |B\rangle\rangle, \tag{A1}$$

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

Linear superoperators. We denote by $\mathcal{L}_{n\to n}$ the set of all linear superoperators of the form $\Phi: \mathcal{B}(\mathcal{H}_n) \to \mathcal{B}(\mathcal{H}_n)$. Notably, quantum channels are linear superoperators which are also completely positive and trace-preserving. Given $\Phi \in \mathcal{L}_{n\to n}$, it is important to note that the map $|\Phi(X)\rangle$ is linear in $|X\rangle$ for all $|X\rangle \in \mathcal{H}_n^{\otimes 2}$. 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 Φ on an operator X can be written as:

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

Furthermore, for every linear superoperator $\Phi: \mathcal{B}(\mathcal{H}_n) \to \mathcal{B}(\mathcal{H}_n)$, there exists [62] 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 $|\Phi(X)\rangle = \sum_{i=1}^{4^n} A_i \otimes B_i^* |X\rangle$, which implies $\widehat{\Phi} = \sum_{i=1}^{4^n} A_i \otimes B_i^*$, that we denote as the matrix form of the linear superoperator.

We will always use the *hat* symbol over a linear superoperator $(\widehat{\Phi})$, to indicate the matrix form of the linear superoperator. It is also easy to observe that the matrix form of composition of two linear superoperators is equal to the product of the associated matrix form: Let Φ_1, Φ_2 be two linear superoperators, then we have $\widehat{\Phi_2 \circ \Phi_1} = \widehat{\Phi}_2 \cdot \widehat{\Phi}_1$.

For every linear superoperator $\Phi: \mathcal{B}(\mathcal{H}_n) \to \mathcal{B}(\mathcal{H}_n)$, written as $\Phi(X) = \sum_{i=1}^{4^n} A_i X B_i^{\dagger}$, its adjoint linear superoperator (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 superoperator Φ^{\dagger} is equal to the adjoint of the matrix form of the linear superoperator, 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 $P, Q \in \mathcal{P}_n$,

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

Given a linear map $\mathcal{C}: \mathcal{B}(\mathcal{H}_n) \to \mathcal{B}(\mathcal{H}_n)$ we denote the transition amplitude associated to two Pauli operators $P, Q \in \mathcal{P}_n$ as follows

$$\langle\!\langle P|\widehat{\mathcal{C}}|Q\rangle\!\rangle = \langle\!\langle P|\mathcal{C}(Q)\rangle\!\rangle = \langle\!\langle \mathcal{C}^{\dagger}(P)|Q\rangle\!\rangle.$$
 (A4)

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

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

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

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

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* (PTM) of the channel.

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

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

and we use |P| to denote the Pauli weight of P, which is defined as the number of non-identity single-qubit Pauli operators in the Pauli tensor decomposition of P, i.e., |P| = |supp(P)|.

Matrix norms. For a matrix $A \in \mathcal{B}(\mathcal{H}_n)$, the Schatten p-norm is defined as $||A||_p := \left(\operatorname{Tr}\left[(\sqrt{A^{\dagger}A})^p\right]\right]^{1/p}$, 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 $p \to \infty$.

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.,

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

Given a Hermitian operator O, we define its squared normalized Frobenius norm $||O||_{\rm F}^2$ as

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

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

$$\|O\|_2^2 = \text{Tr}\big[\mathbb{F}O^{\otimes 2}\big] \tag{A10}$$

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

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

2. Ensembles of superoperators and states

In this section, we define several ensembles of superoperators 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. [63].

Definition 1 (Haar measure and t-designs). The Haar measure μ_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

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

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 of the t-th order, that is

$$\mathbb{E}_{U \sim \mathcal{D}} \left[U^{\otimes t} O U^{\dagger \otimes t} \right] = \mathbb{E}_{U \sim \mu_H} \left[U^{\otimes t} O U^{\dagger \otimes t} \right], \tag{A13}$$

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

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

$$\mathbb{E}_{U \sim \mathcal{D}} \left[U^{\otimes t} \otimes U^{* \otimes t} \right] = \mathbb{E}_{U \sim \mu_H} \left[U^{\otimes t} \otimes U^{* \otimes t} \right]. \tag{A14}$$

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

Example 1 (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

$$\mathbb{E}_{U \sim \mathcal{D}} \left[UOU^{\dagger} \right] = \mathbb{E}_{U \sim \mu_H} \left[UOU^{\dagger} \right] = \frac{I}{2^n} \operatorname{Tr}[O]. \tag{A15}$$

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

$$\frac{1}{4^n} \sum_{P \in \mathcal{P}_n} [POP] = \frac{I}{2^n} \operatorname{Tr}[O]. \tag{A16}$$

The above identity is also commonly referred 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

$$\mathbb{E}_{U \sim \mathcal{D}} \left[U^{\otimes 2} O U^{\dagger \otimes 2} \right] = \mathbb{E}_{U \sim \mu_H} \left[U^{\otimes 2} O U^{\dagger \otimes 2} \right] = c_{\mathbb{I}, O} \mathbb{I} + c_{\mathbb{F}, O} \mathbb{F}. \tag{A17}$$

where

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

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

In the following, we define the class of *locally unbiased* distributions over super-operators. 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 2 (Locally unbiased distribution). A distribution \mathcal{D} over linear superoperators $\mathcal{L}_{n\to n}$ 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 $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$ over $\mathbb{U}(2)$ such that (i) all \mathcal{D}_i are unitary 1-designs, and (ii) the following identity holds

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

Locally unbiased distributions generalize the following notion of *Pauli-invariance* which was previously proposed in Ref. [13] (in the sense that Pauli-invariance implies local unbiasedness).

Definition 3 (Pauli-invariant distribution). A distribution \mathcal{D} over linear superoperators $\mathcal{L}_{n\to n}$ is Pauli-invariant if it is invariant under right-multiplication of random Pauli operators up to the second moment, i.e., if

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

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

Lemma 1 (Orthogonality). Let \mathcal{D} be a distribution over linear superoperators $\mathcal{L}_{n\to n}$.

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

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

In particular, for all observables O, we have

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

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

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

In particular, for all observables O, we have

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

Proof. We start by proving the first statement. Recall that, by Definition 2, we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left[\widehat{\mathcal{C}}^{\otimes 2} \right] = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \mathbb{E}_{V \sim \bigotimes_{i=1}^{n} \mathcal{D}_{i}} \left[\widehat{\mathcal{C}}^{\otimes 2} V^{\otimes 2} \otimes V^{* \otimes 2} \right], \tag{A25}$$

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 $\operatorname{supp}(P) \neq \operatorname{supp}(Q)$, there exists an index $j \in [n]$ such that $P_j \neq I$ and $Q_j = I$. From the previous, we have that

$$\mathbb{E}_{V_i \sim \mathcal{D}_i} \left[V_i^{\dagger \otimes 2} (P_j \otimes Q_j) V_i^{\otimes 2} \right] = \mathbb{E}_{V_i \sim \mathcal{D}_i} \left[V_i^{\dagger} P_j V_j \right] \otimes I = 0, \tag{A26}$$

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

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

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

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

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

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

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

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 super-operators $\mathcal{L}_{n\to n}$. First, let us use the invariance under right-multiplication of random Pauli

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

From here we can see that it suffices to show that

$$\sum_{R \in \mathcal{P}_n} RPR \otimes RQR = 0. \tag{A33}$$

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

$$\sum_{R \in \mathcal{P}_n} RPR \otimes RQR = \sum_{R \in \mathcal{P}_n} (-1)^{\langle P, R \rangle + \langle Q, R \rangle} P \otimes Q = \sum_{R \in \mathcal{P}_n} (-1)^{\langle PQ, R \rangle} P \otimes Q = 0, \tag{A34}$$

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 $\mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left[\mathcal{C}^{\otimes 2}(P \otimes Q) \right] = 0$ implies the Pauli invariance property. Let O be an arbitrary Hermitian operator. We have

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

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

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

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

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

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

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

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 2 (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

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

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

$$= -\frac{1}{2}(X \otimes Y) + \frac{1}{2}\left(\cos^2(\theta)X \otimes Y + \sin^2(\theta)Y \otimes X + \cos(\theta)\sin(\theta)(X \otimes X - Y \otimes Y)\right),\tag{A43}$$

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

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

Definition 4 (Approximate local scrambling). Let $\eta \in [0,1)$. A distribution \mathcal{D} over $\mathbb{U}(2)$ is an η -approximate scrambler if it satisfies the following properties

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

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

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

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

where V_1, V_2, \ldots, 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, 64–66].

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

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

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

$$\|\mathbf{a}\|_{2}^{2} = 1 \quad and \quad \|\mathbf{a}\|_{\infty}^{2} \leqslant \frac{1+2\eta}{3},$$
 (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

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

Equivalently, the min-entropy can be expressed in terms of the max-relative entropy between p_a and the uniform distribution $p_u := (1/3, 1/3, 1/3)$

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

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 η -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 η -approximate scrambler, as formalized in the following Lemma.

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

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

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

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

$$\mathbb{E}_{U \sim \mathcal{D}} \left[(UV)^{\dagger \otimes 2} (P \otimes Q) (UV)^{\otimes 2} \right] = V^{\dagger \otimes 2} \mathbb{E}_{U \sim \mathcal{D}} \left[U^{\dagger \otimes 2} (P \otimes Q) U^{\otimes 2} \right] V^{\otimes 2}, \tag{A53}$$

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 1. We have

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

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

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

$$\sum_{R \in \{X,Y,Z\}} \langle\!\langle Q|V \otimes V^*|R \rangle\!\rangle^2 = \|V^{\dagger}QV\|_{F}^2 = \|Q\|_{F}^2 = 1, \tag{A56}$$

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 $\mathcal{K}: \{X,Y,Z\} \to \{X,Y,Z\}$ defined as follows

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

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

$$\mathcal{K}(p_{\boldsymbol{a}}) = p_{\boldsymbol{b}} \coloneqq (b_X^2, b_Y^2, b_Z^2),\tag{A58}$$

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

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

$$\mathcal{K}(p_{\boldsymbol{u}}) = p_{\boldsymbol{u}}.\tag{A60}$$

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

$$D_{\max}(p_{\boldsymbol{b}}||p_{\boldsymbol{u}}) = D_{\max}(\mathcal{K}(p_{\boldsymbol{a}})||\mathcal{K}(p_{\boldsymbol{u}})) \tag{A61}$$

$$\leq D_{\max}(p_{\boldsymbol{a}}||p_{\boldsymbol{u}}) \leq \log(1+2\eta),$$
 (A62)

which implies Eq. (A52).

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 1. However, we will demonstrate below that there are other practically relevant distributions over $\mathbb{U}(2)$ that are η -approximate scramblers with $\eta < 1$, highlighting the broad applicability of our analysis.

Example 3 (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 ϕ and θ 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 $P \neq Q$,

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

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

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

$$\mathbb{E}_{\theta,\phi}(R_Z(\theta)^{\dagger}R_X(\phi)^{\dagger}YR_X(\phi)R_Z(\theta))^{\otimes 2} = \frac{1}{4}\left(X^{\otimes 2} + Y^{\otimes 2}\right) + \frac{1}{2}Z^{\otimes 2},\tag{A65}$$

$$\mathbb{E}_{\theta,\phi}(R_Z(\theta)^{\dagger}R_X(\phi)^{\dagger}ZR_X(\phi)R_Z(\theta))^{\otimes 2} = \frac{1}{4}\left(X^{\otimes 2} + Y^{\otimes 2}\right) + \frac{1}{2}Z^{\otimes 2}.$$
 (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 η -approximate scramblers for $\eta < 1$, nor do they form a unitary 1-design.

Example 4 (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 $R_Z(\theta)ZR_Z(\theta)^{\dagger}=Z$ for all θ . In particular, this implies

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

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

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

$$\|\mathbb{E}_{\rho \sim \mathcal{D}}[\rho]\|_{\infty} \leqslant \frac{c}{2^n}.\tag{A69}$$

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 ρ

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

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.

Lemma 3 (Lemma 3 in Ref. [16]). Let O be an observable and \mathcal{D} be a low-average ensemble over quantum states with purity c. We have

$$\mathbb{E}_{\rho \sim \mathcal{D}} \operatorname{Tr}[O\rho]^2 \leqslant c \|O\|_{\mathcal{F}}^2. \tag{A71}$$

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 $\langle P|\widehat{\mathcal{N}}|Q\rangle\rangle$, where $P,Q\in\mathcal{P}_1$. Since \mathcal{N} is trace-preserving, it satisfies $\langle I|\widehat{\mathcal{N}}|Q\rangle\rangle=0$ for all $Q\in\{X,Y,Z\}$ and $\langle I|\widehat{\mathcal{N}}|I\rangle\rangle=1$. 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].

Lemma 4 (Normal form of a quantum channel [18, 67]). Any single-qubit quantum channel \mathcal{N} can be written in the so called 'normal' form:

$$\mathcal{N}(\cdot) = U\mathcal{N}'(V^{\dagger}(\cdot)V)U^{\dagger},\tag{A72}$$

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

$$\langle\!\langle P|\widehat{\mathcal{N}'}|Q\rangle\!\rangle := \begin{cases} 1 & \text{if } P = Q = I, \\ D_P & \text{if } P = Q \neq I, \\ t_P & \text{if } P \neq I \text{ and } Q = I, \\ 0 & \text{otherwise.} \end{cases}$$
(A73)

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.

- 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.
- 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:
 - **Depolarizing class.** A channel \mathcal{N} belong 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.
 - **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 ρ 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.

• 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 ρ 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.

Lemma 5 (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

$$\Upsilon(\mathbf{D}, \mathbf{t}) := \max_{\|\mathbf{a}\|_{2}^{2} = 1} \left\{ \sum_{Q \in \{X, Y, Z\}} a_{Q}^{2} D_{Q}^{2} + \left(\sum_{Q \in \{X, Y, Z\}} a_{Q} t_{Q} \right)^{2} \right\}.$$
(A74)

We have that

$$\Upsilon(\mathbf{D}, \mathbf{t}) \leqslant 1. \tag{A75}$$

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. (A74), i.e.

$$(b_X, b_Y, b_Z) := \underset{\|a\|_2^2 = 1}{\operatorname{arg \, max}} \sum_{Q \in \{X, Y, Z\}} a_Q^2 D_Q^2 + \left(\sum_{Q \in \{X, Y, Z\}} a_Q t_Q\right)^2. \tag{A76}$$

We define the observable O and the state ρ as follows:

$$O = \sum_{P \in \{X,Y,Z\}} |b_P| \cdot \operatorname{sgn}(t_P \cdot D_P) P \quad \text{and} \quad \rho = \frac{I+O}{2}, \tag{A77}$$

where 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

$$||O||_{\infty} = \max_{\sigma} |\text{Tr}[O\sigma]| = \max_{\substack{r \in \mathbb{R}^3 \\ ||r||_2^2 = 1}} \left| \sum_{P \in \{X, Y, Z\}} |b_P| r_P \cdot \text{sgn}(t_P \cdot D_P) \right| \le \max_{\substack{r \in \mathbb{R}^3 \\ ||r||_2^2 = 1}} ||\mathbf{b}||_2 ||\mathbf{r}||_2 = 1, \tag{A78}$$

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. (A76). Recall that the channel \mathcal{N}' acts as

$$\mathcal{N}'\left(\frac{1}{2}\left(I + \sum_{P \in \{X,Y,Z\}} r_P P\right)\right) = \frac{I}{2} + \frac{1}{2} \sum_{P \in \{X,Y,Z\}} (r_P D_P + t_P) P. \tag{A79}$$

Replacing the coefficients r_P with the appropriate values for the state ρ , we obtain that

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

$$= \sum_{P\in\{X,Y,Z\}} b_P^2 D_P + |b_P t_P| \operatorname{sgn}(D_P).$$
(A81)

By Hölder's inequality, we have that

$$1 = ||O||_{\infty} ||\rho||_{1} \geqslant |\text{Tr}[O\mathcal{N}'(\rho)]| = \sum_{P \in \{X, Y, Z\}} b_{P}^{2} D_{P} + |b_{P} t_{P}| \text{sgn}(D_{P})$$
(A82)

$$= \sum_{P \in \{X,Y,Z\}} b_P^2 |D_P| + \sum_{P \in \{X,Y,Z\}} |b_P t_P| \tag{A83}$$

$$\geqslant \sum_{P \in \{X,Y,Z\}} b_P^2 D_P^2 + \left(\sum_{P \in \{X,Y,Z\}} b_P t_P\right)^2, \tag{A84}$$

where the second-to-last step follows from the fact that the coefficients D_P have all the same sign (cf. Lemma 4) 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

$$\|\mathbf{D}\|_{\infty}^{2} \in (0,1) \implies \forall P \in \{X, Y, Z\} : D_{P}^{2} < |D_{P}|$$
 (A85)

$$\implies \sum_{P \in \{X,Y,Z\}} b_P^2 D_P^2 < \sum_{P \in \{X,Y,Z\}} b_P^2 |D_P|, \tag{A86}$$

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 $||b||_2^2 = 1$. Moreover, by Cauchy-Schwarz inequality, we have

$$\left(\sum_{P \in \{X,Y,Z\}} b_P t_P\right)^2 \le \|\mathbf{t}\|_2^2 \|\boldsymbol{b}\|_2^2 = \|\mathbf{t}\|_2^2. \tag{A87}$$

and therefore

$$\|\mathbf{t}\|_{2}^{2} \in (0,1) \implies \left(\sum_{P \in \{X,Y,Z\}} b_{P} t_{P}\right)^{2} < \sum_{P \in \{X,Y,Z\}} |b_{P} t_{P}|,$$
 (A88)

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

$$\sum_{P \in \{X,Y,Z\}} b_P^2 D_P^2 + \left(\sum_{P \in \{X,Y,Z\}} b_P t_P\right)^2 < 1,\tag{A89}$$

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 $\mathcal C$ be an L-layered quantum channel

$$C = C_L \circ C_{L-1} \cdots \circ C_1. \tag{A90}$$

Then, its matrix (vectorized) form is:

$$\widehat{\mathcal{C}} = \widehat{\mathcal{C}}_L \widehat{\mathcal{C}}_{L-1} \dots \widehat{\mathcal{C}}_1. \tag{A91}$$

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

$$\Phi_{\gamma}(\mathcal{C}) := \langle \langle P_L | \widehat{\mathcal{C}}_L | P_{L-1} \rangle \rangle \times \langle \langle P_{L-1} | \widehat{\mathcal{C}}_{L-1} | P_{L-2} \rangle \rangle \times \dots \times \langle \langle P_1 | \widehat{\mathcal{C}}_1 | P_0 \rangle \rangle$$
(A92)

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

$$C(\rho) = \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle \langle \rho | P_0 \rangle \rangle \Phi_{\gamma}(\mathcal{C}) P_L, \tag{A93}$$

$$\mathcal{C}(\rho) = \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle \langle \rho | P_0 \rangle \rangle \Phi_{\gamma}(\mathcal{C}) P_L, \tag{A93}$$

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

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

$$C_{\mathcal{A}}(\cdot) := \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{A}} \langle \langle \cdot | P_0 \rangle \rangle \Phi_{\gamma}(\mathcal{C}) P_L.$$
(A95)

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

$$C_{\mathcal{A}}^{\dagger}(\cdot) = \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{A}} \langle \langle \cdot | P_L \rangle \rangle \Phi_{\gamma}(\mathcal{C}) P_0.$$
(A96)

Then we have

$$C_{\mathcal{A}}(\rho) = \sum_{\gamma = (P_0, P_L, P_L) \in \mathcal{A}} \langle \langle \rho | P_0 \rangle \rangle \Phi_{\gamma}(\mathcal{C}) P_L, \tag{A97}$$

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

$$C_{\mathcal{A}}^{\dagger}(O) = \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{A}} \langle \langle O | P_L \rangle \rangle \Phi_{\gamma}(\mathcal{C}) P_0.$$
(A98)

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

Definition 6 (Path support and path weight). Let $\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}$ be a Pauli path.

• The support of γ is the vector of the supports of its components:

$$\operatorname{supp}(\gamma) = (\operatorname{supp}(P_0), \operatorname{supp}(P_1), \dots, \operatorname{supp}(P_L)). \tag{A99}$$

• The weight of γ is the sum of the Pauli weight of its components:

$$|\gamma| = \sum_{i=1}^{L} |P_i|. \tag{A100}$$

We will also make extensive use the following orthogonality relationships.

Lemma 6 (Orthogonality of Pauli paths). Let $C = C_L \circ C_{L-1} \circ \cdots \circ C_1$ be an L-layered quantum channel sampled from a distribution \mathcal{D}_{circ} .

1. Assume that all layers 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:

$$\operatorname{supp}(\gamma) \neq \operatorname{supp}(\gamma') \implies \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\operatorname{circ}}} \left[\Phi_{\gamma}(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C}) \right] = 0. \tag{A101}$$

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

$$\gamma \neq \gamma' \implies \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left[\Phi_{\gamma}(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C}) \right] = 0.$$
 (A102)

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

$$\Phi_{\gamma}(\mathcal{C})\Phi_{\gamma'}(\mathcal{C}) \tag{A103}$$

$$= \langle \langle P_L | \widehat{\mathcal{C}}_L | P_{L-1} \rangle \rangle \langle \langle P_L' | \widehat{\mathcal{C}}_L | P_{L-1}' \rangle \rangle \langle \langle P_{L-1} | \widehat{\mathcal{C}}_{L-1} | P_{L-2} \rangle \rangle \langle \langle P_{L-1}' | \widehat{\mathcal{C}}_{L-1} | P_{L-2}' \rangle \rangle \dots \langle \langle P_1 | \widehat{\mathcal{C}}_1 | P_0 \rangle \rangle \langle \langle P_1' | \widehat{\mathcal{C}}_1 | P_0' \rangle \rangle$$
(A104)

$$= \langle \langle P_L \otimes P_L' | \widehat{\mathcal{C}}_L^{\otimes 2} | P_{L-1} \otimes P_{L-1}' \rangle \langle \langle P_{L-1} \otimes P_{L-1}' | \widehat{\mathcal{C}}_{L-1}^{\otimes 2} | P_{L-2} \otimes P_{L-2}' \rangle \dots \langle \langle P_1 \otimes P_1' | \widehat{\mathcal{C}}_1^{\otimes 2} | P_0 \otimes P_0' \rangle \rangle. \tag{A105}$$

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

$$\mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} \left[\mathcal{C}_{j+1}^{\otimes 2} (P_j \otimes P_j') \right] = 0. \tag{A106}$$

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

$$\mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} \langle \langle P_{j+1} \otimes P'_{j+1} | \widehat{\mathcal{C}}_{j+1}^{\otimes 2} | P_j \otimes P'_j \rangle \rangle = \mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} \frac{1}{4^n} \operatorname{Tr} \left\{ \mathcal{C}_{j+1}^{\otimes 2} (P_j \otimes P'_j) (P_{j+1} \otimes P'_{j+1}) \right\}$$
(A107)

$$= \frac{1}{4^n} \operatorname{Tr} \left\{ \mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} [\mathcal{C}_{j+1}^{\otimes 2}(P_j \otimes P_j')] (P_{j+1} \otimes P_{j+1}') \right\} = 0, \tag{A108}$$

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

$$\mathbb{E}_{\mathcal{C}_L \sim \mathcal{D}_L} \left[\mathcal{C}_L^{\dagger \otimes 2} (P_L \otimes P_L') \right] = 0. \tag{A109}$$

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

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

$$\mathbb{E}_{\mathcal{C}_{j+1} \sim \mathcal{D}_{j+1}} \left[\mathcal{C}_{j+1}^{\otimes 2} (P_j \otimes P_j') \right] = 0. \tag{A110}$$

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

$$\mathbb{E}_{\mathcal{C}_L \sim \mathcal{D}_L} \left[\mathcal{C}_L^{\dagger \otimes 2} (P_L \otimes P_L') \right] = 0, \tag{A111}$$

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

a. Estimating second moments by sampling Pauli paths

Assume we are given as input an initial state ρ and an observable O. Then, the goal is to approximate the expectation value $\text{Tr}[O\mathcal{C}(\rho)]$, where \mathcal{C} is an L-layered quantum channel sampled according to a distribution \mathcal{D} . In this setting, it is customary to evaluate the quality of a classical approximation with respect to the mean squared error:

$$\Delta_{\mathcal{A}} := \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left(\text{Tr}[O\mathcal{C}(\rho)] - \text{Tr}[O\mathcal{C}_{\mathcal{A}}(\rho)] \right)^2 = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left(\text{Tr}[\mathcal{C}^{\dagger}(O)\rho] - \text{Tr}[\mathcal{C}^{\dagger}_{\mathcal{A}}(O)\rho] \right)^2. \tag{A112}$$

We observe that the mean squared error can be re-expressed as follows

$$\Delta_{\mathcal{A}} = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left(\sum_{\gamma = (P_0, P_1, \dots, P_L) \notin \mathcal{A}} \langle \langle O | P_L \rangle \rangle \Phi_{\gamma}(\mathcal{C}) \operatorname{Tr}[P_0 \rho] \right)^2.$$
(A113)

The above expression can be drastically simplified if the different Pauli paths are "orthogonal". That is, if given $\gamma =$ (P_0, P_1, \dots, P_L) and $\gamma' = (P'_0, P'_1, \dots, P'_L)$, the Pauli paths associated to O are orthogonal, i.e., they satisfy

$$\gamma \neq \gamma' \implies \langle\!\langle O^{\otimes 2} | P_L \otimes P_L' \rangle\!\rangle \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{circ}} \Phi_{\gamma}(\mathcal{C}) \Phi_{\gamma'}(\mathcal{C}) = 0. \tag{A114}$$

Under this assumption, which is for example satisfied for Pauli invariant distributions, Eq. (A113) can be simplified as

$$\Delta_{\mathcal{A}} = \sum_{\gamma = (P_0, P_1, \dots, P_L) \notin \mathcal{A}} \langle \langle O|P_L \rangle \rangle^2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \Phi_{\gamma}^2(\mathcal{C}) \operatorname{Tr}[P_0 \rho]^2.$$
(A115)

Lemma 7. Let \mathcal{D} be a distribution over L-layered quantum channels. Then for all observable O we have

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

Proof. For any Pauli $Q \in \mathcal{P}_n$ we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \mathcal{C}^{\dagger}(Q) \right\|_{\text{F}}^{2} = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \sum_{\substack{\gamma = (P_{0}, P_{1}, \dots, P_{L}) \in \mathcal{P}_{n}^{L+1} \\ P_{L} = Q}} \Phi_{\gamma}(\mathcal{C}) P_{0} \right\|_{\text{F}}^{2}$$
(A117)

$$= \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{circ}} \Phi_{\gamma}^2(\mathcal{C}) \|P_0\|_{F}^2 = \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{circ}} \Phi_{\gamma}^2(\mathcal{C}). \tag{A118}$$

Therefore, for an arbitrary observable O we have

$$\sum_{\mathcal{C} \sim \mathcal{D}_{circ}} \langle \langle O|P_L \rangle \rangle^2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{circ}} \Phi_{\gamma}^2(\mathcal{C})$$
(A119)

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

$$= \sum_{P_L \in \mathcal{P}_n} \langle \langle O|P_L \rangle \rangle^2 \sum_{\gamma=(P_0, P_1, \dots, P_{L-1}) \in \mathcal{P}_n^L} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{circ}} \Phi_{\gamma}^2(\mathcal{C})$$
(A119)

$$= \sum_{P_L \in \mathcal{P}_n} \langle \langle O | P_L \rangle \rangle^2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \mathcal{C}^{\dagger}(P_L) \right\|_F^2 = \|O\|_F^2 \max_{Q \in \mathcal{P}_n} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \mathcal{C}^{\dagger}(Q) \right\|_F^2$$
(A121)

$$\leq ||O||_{\mathcal{F}}^{2} ||\mathcal{C}^{\dagger}(Q)||_{\infty}^{2} \leq ||O||_{\mathcal{F}}^{2} ||Q||_{\infty} = ||O||_{\mathcal{F}}^{2},$$
 (A122)

where the first inequality follows from the relation $2^n \|A\|_F^2 = \|A\|_2^2 \leqslant 2^n \|A\|_\infty^2$ and the second inequality follows from the Russo-Dye theorem [68] (which states $\|\mathcal{C}^{\dagger}(A)\|_{\infty} \le \|A\|_\infty$ for any quantum channel \mathcal{C}), and in the final step we noted that the operator norm of any Pauli matrix is 1.

Theorem 4 (Monte Carlo estimates). Let $F: \mathcal{P}_n^{L+1} \to [0,1]$ be a function of the Pauli paths. Let v be a target value defined as

$$v := \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}} \langle \langle O|P_L \rangle \rangle^2 \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{circ}} \left[\Phi_{\gamma}^2(\mathcal{C}) \right] F(\gamma). \tag{A123}$$

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 $\mathbb{E}_{\mathcal{C}_j \sim \mathcal{D}_j} \langle\!\langle P|\mathcal{C}_j|Q\rangle\!\rangle^2$ in time T. Then for any $\epsilon, \delta \in (0,1]$, there exists a classical randomized algorithm that runs in time

$$TL\epsilon^{-2}\log(1/\delta) \tag{A124}$$

and outputs a value v' such that

$$|v' - v| \leqslant \epsilon ||O||_{\mathcal{F}}^2,\tag{A125}$$

with probability at least $1 - \delta$.

Proof. We consider the following estimation procedure.

1. Sample P_L with probability

$$\frac{\langle\!\langle O|P_L\rangle\!\rangle^2}{\sum_P \langle\!\langle O|P\rangle\!\rangle^2} = \frac{\text{Tr}[OP_L]^2}{2^n |\!|O|\!|_2^2} \,. \tag{A126}$$

2. for $i = L, \ldots, 1$, given P_i , sample P_{i-1} with probability

$$\frac{\mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_j} \langle \langle P_i | \mathcal{C}_i | P_{i-1} \rangle \rangle^2}{\sum_{P_{i-1}} \mathbb{E}_{\mathcal{C}_i} \langle \langle P_i | \mathcal{C}_i | P_{i-1} \rangle \rangle^2} = \frac{\mathbb{E}_{\mathcal{C}_i} \operatorname{Tr}[P_i \mathcal{C}_i(P_{i-1})]^2}{2^n \mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \left\| \mathcal{C}_i^{\dagger}(P_i) \right\|_2^2}.$$
(A127)

- 3. Given $\gamma = (P_1, P_2, \dots, P_L)$, output $F(\gamma)$.
- 4. Repeat from 1 to 3 for $2\log(2/\delta)/\epsilon^2$ times and compute the empirical mean.

It is easy to see that the procedure described above yields an unbiased estimator for the target value v. Moreover, by Chernoff-Hoeffding bound, the final outcome approximates the mean squared error with precision $\epsilon ||O||_F^2$ with probability at least $1 - \delta$.

In the following, we highlight some key applications of Theorem 4.

Observation 2. We observe that Theorem 4 allows for the estimation of many important quantities related to the distribution \mathcal{D}_{circ} . First, given a subset of paths $\mathcal{A} \subseteq \mathcal{P}_n^{L+1}$, the following choice of $F(\cdot)$ yields the mean squared error $\Delta_{\mathcal{A}}$

$$F(\gamma) = \begin{cases} 0 & \text{if } \gamma \in \mathcal{A} \\ \operatorname{Tr}[P_0 \rho]^2 \prod_{i=1}^L \mathbb{E}_{\mathcal{C}_i \sim \mathcal{D}_i} \left\| \mathcal{C}_i^{\dagger}(P_i) \right\|_2^2 & \text{if } \gamma \notin \mathcal{A} \end{cases}$$
 (A128)

We also remark that if $\mathcal{A} = \emptyset$, then $\Delta_{\mathcal{A}} = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\mathrm{circ}}} \operatorname{Tr}[O\mathcal{C}(\rho)]^2$ is the second moment of the observable O.

An alternative metric for evaluating the quality of the simulation is the expected (normalized) Frobenius distance between the Heisenberg evolved observable and its approximate version, i.e.

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \mathcal{C}^{\dagger}(O) - \mathcal{C}_{k}^{\dagger}(O) \right\|_{F}^{2} = \sum_{\gamma = (P_{0}, P_{1}, \dots, P_{L}) \notin \mathcal{A}} \left\langle \langle O|P_{L} \right\rangle^{2} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left[\Phi_{\gamma}^{2}(\mathcal{C}) \right]. \tag{A129}$$

Then the above quantity can be estimated via Theorem 4 by setting

$$F(\gamma) = \begin{cases} 0 & \text{if } \gamma \in \mathcal{A} \\ \prod_{i=1}^{L} \mathbb{E}_{\mathcal{C}_{i} \sim \mathcal{D}_{i}} \left\| \mathcal{C}_{i}^{\dagger}(P_{i}) \right\|_{2}^{2} & \text{if } \gamma \notin \mathcal{A} \end{cases}$$
 (A130)

5. Noisy circuit model

We introduce our noisy circuit model.

Definition 7 (Noisy Circuit Model). A noisy L-layered quantum circuit 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:

$$\mathcal{C} := \mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_L \circ \cdots \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_1 \circ \mathcal{N}^{\otimes n}, \tag{A131}$$

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. $V^{\text{single}} := V(\cdot)V^{\dagger}$, 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 $i=1,2,\ldots,L$, $\mathcal{U}_i=U_i(\cdot)U_i^{\dagger}$ 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, ..., L + 1, the distribution \mathcal{D}_i satisfies the following

$$\mathbb{E}_{U \sim \mathcal{D}_i} \left[U \otimes U^* \right] = \mathbb{E}_{U \sim \mathcal{D}_i'} \mathbb{E}_{W \sim \mathcal{D}_{\text{single}}^{\otimes n}} \left[UW \otimes U^* W^* \right] \tag{A132}$$

We denote by \mathcal{D}_{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. (A132) with the following more general expression:

$$\mathbb{E}_{U \sim \mathcal{D}_i} \left[U \otimes U^* \right] = \mathbb{E}_{U \sim \mathcal{D}_i'} \, \mathbb{E}_{W \sim \bigotimes_{j=1}^n \mathcal{D}_{\text{single},i}^{(j)}} \left[UW \otimes U^* W^* \right], \tag{A133}$$

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 $\bigotimes_{j=1}^{n} \mathcal{N}^{(j)}$.

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

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

and therefore

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

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

$$\left\| \mathcal{N}_p^{(\text{depo})\dagger \otimes n}(O) \right\|_{\text{F}}^2 \leqslant (1-p)^2 \|O\|_{\text{F}}^2. \tag{B3}$$

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. The scenario becomes significantly more intricate when the noise is non-unital. Let \mathcal{A}_q be the amplitude-damping channel with noise rate q. We have

$$\mathcal{A}_{q}^{\dagger \otimes 2}(ZZ + IZ + ZI) = (1 - q)^{2}ZZ + (1 - q^{2})(IZ + ZI) + (2q + q^{2})II.$$
 (B4)

Therefore, the corresponding Frobenius norm is

$$\|\mathcal{A}_{q}^{\dagger \otimes 2}(ZZ + IZ + ZI)\|_{F}^{2} = q^{2}(2+q)^{2} + (1-q)^{4} + 2(1-q^{2})^{2}.$$
 (B5)

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

$$\frac{\|\mathcal{A}_{q}^{\dagger \otimes 2}(ZZ + IZ + ZI)\|_{F}^{2}}{\|ZZ + IZ + ZI\|_{F}^{2}} = \frac{10}{3}.$$
 (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 8 (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 non-identity on A and identity elsewhere:

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

Given a single-qubit quantum channel N, we define the contraction coefficient of N as

$$\chi(\mathcal{N}) \coloneqq \max_{A \subseteq [n]} \max_{\substack{O \\ \|O_A\|_{\Gamma} \neq 0}} \left(\frac{\left\| \mathcal{N}^{\dagger \otimes n}(O_A) \right\|_{F}}{\left\| O_A \right\|_{F}} \right)^{1/|A|}. \tag{B8}$$

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

$$\chi_{\mathcal{D}}^{2}(\mathcal{N}) \coloneqq \max_{A \subseteq [n]} \max_{\substack{O \\ \|O_{A}\|_{F} \neq 0}} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \left(\frac{\left\| \mathcal{N}^{\dagger \otimes n}(W^{\dagger} O_{A} W) \right\|_{F}^{2}}{\left\| O_{A} \right\|_{F}^{2}} \right)^{1/|A|}. \tag{B9}$$

Observation 3. 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 $(U^{\dagger \otimes n}OU^{\otimes n})_A = U^{\dagger \otimes n}O_AU^{\otimes n}$, and second, the Frobenius norm is unitarily invariant. Thus, we obtain that

$$\chi(\mathcal{N}) = \max_{A \subseteq [n]} \max_{\substack{O \\ \|O_A\|_{\mathbf{n}} \neq 0}} \left(\frac{\left\| V^{\dagger \otimes n} \mathcal{N}'^{\dagger \otimes n} (U^{\dagger \otimes n} O_A U^{\otimes n}) V^{\otimes n} \right\|_{\mathbf{F}}}{\left\|O_A\right\|_{\mathbf{F}}} \right)^{1/|A|}$$
(B10)

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

Applying iteratively Lemma 5, 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.

Lemma 8 (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 **D** and **t**. We have

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

In particular, by Lemma 5, 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 prove the following claim.

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

$$\left\| \mathcal{N}'^{\dagger} \otimes \mathcal{I}_{k-1}(H) \right\|_{\mathcal{F}}^{2} \leqslant c \|H\|_{\mathcal{F}}^{2}, \tag{B13}$$

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 $\mathcal{N}'^{\dagger} \otimes \mathcal{I}_{k-1}(H)$

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

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

$$\|\mathcal{N}'^{\dagger} \otimes \mathcal{I}_{k-1}(H)\|_{F}^{2} = \sum_{Q \in \{I, X, Y, Z\} \otimes (k-1)} \left(\sum_{P \in \{X, Y, Z\}} D_{P}^{2} b_{P \otimes Q}^{2} + \left(\sum_{P \in \{X, Y, Z\}} b_{P \otimes Q} t_{P} \right)^{2} \right)$$
(B15)

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

$$\leqslant \Upsilon(\mathbf{D}, \mathbf{t}) \sum_{Q \in \{I, X, Y, Z\} \otimes (k-1)} \|\boldsymbol{b}_{Q}\|_{2}^{2} \coloneqq c \|\boldsymbol{H}\|_{\mathbf{F}}^{2}, \tag{B17}$$

where we denoted $\|\boldsymbol{b}_Q\|_2^2 := \sum_{P \in \{X,Y,Z\}} b_{P \otimes Q}^2$ and the only inequality follows from Lemma 5. This concludes the proof.

Given a subset of qubits $A \subseteq [n]$ with size |A| = k, let $O_A = \sum_{\text{supp}(P)=A} a_P P$ be a linear combination of Pauli operators supported on A. Given $j \in A$, we have that

$$\mathcal{I}_{[n]\backslash A} \otimes \mathcal{I}_{j} \otimes \mathcal{N}'^{\dagger \otimes (k-1)}(O_{A}) = \sum_{P \in \{X,Y,Z\}} \sum_{Q \in \{I,X,Y,Z\}^{\otimes (k-1)}} b_{P \otimes Q} P \otimes Q \otimes I^{\otimes (n-k)}$$
(B18)

for some suitable coefficients $b_{P\otimes Q}$. Then the desired result can be proven by applying iteratively Claim 1.

$$\left\| \mathcal{N}'^{\dagger \otimes n}(O_A) \right\|_{\mathrm{F}}^2 \le \left\| \mathcal{N}'^{\dagger} \otimes \mathcal{I}_{k-1}(\mathcal{I}_1 \otimes \mathcal{N}'^{\dagger \otimes (k-1)}(O_A)) \right\|_{\mathrm{F}}^2 \tag{B19}$$

$$\leqslant c \left\| \mathcal{I}_{1} \otimes \mathcal{N}'^{\dagger \otimes (k-1)}(O_{A}) \right\|_{F}^{2} \leqslant c \left\| \mathcal{I}_{1} \otimes \mathcal{N}'^{\dagger} \otimes \mathcal{I}_{k-2}(\mathcal{I}_{2} \otimes \mathcal{N}'^{\dagger \otimes (k-2)}(O_{A})) \right\|_{F}^{2}$$
(B20)

$$\leqslant c^2 \left\| \mathcal{I}_2 \otimes \mathcal{N}'^{\dagger \otimes (k-2)}(O_A) \right\|_{\mathcal{F}}^2 \leqslant \ldots \leqslant c^k \left\| O_A \right\|_{\mathcal{F}}^2. \tag{B21}$$

Noise within the dephasing class leads to norm contraction if the initial Pauli observable becomes scrambled into at least two other Pauli observables under the action of a unitary transformation. This intuition is formalized using the concept of approximate scrambling, as introduced in Definition 4.

Lemma 9 (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 η -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

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

2. If \mathcal{N} belongs to the dephasing class, we have

$$\chi_{\mathcal{D}}^2(\mathcal{N}) \leqslant \eta + (1 - \eta) \frac{\|\mathbf{D}\|_2^2}{3}.$$
 (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 $V^{\dagger}W^{\dagger}PWV = \sum_{Q \in \{X,Y,Z\}} a_Q Q$. By the definition of η -approximate scrambler and Lemma 2, we have that

$$\max_{Q \in \{X,Y,Z\}} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} a_Q^2 = \max_{Q \in \{X,Y,Z\}} \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \langle \langle P|WU|Q \rangle \rangle^2 \leqslant \frac{1+2\eta}{3}, \tag{B24}$$

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

$$\mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \left\| \mathcal{N}^{\dagger}(W^{\dagger}PW) \right\|_{\mathcal{F}}^{2} = \mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \sum_{Q \in \{X,Y,Z\}} a_{Q}^{2} D_{Q}^{2} + \left(\sum_{Q \in \{X,Y,Z\}} a_{Q} t_{Q} \right)^{2}. \tag{B25}$$

Thus, for $\eta = 0$, we have

$$\mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \left\| \mathcal{N}^{\dagger}(W^{\dagger} P W) \right\|_{F}^{2} = \frac{1}{3} \left(\|\mathbf{D}\|_{2}^{2} + \|\mathbf{t}\|_{2}^{2} \right). \tag{B26}$$

Moreover, given $A\subseteq [n]$ a $O=\sum_{\substack{P\in\mathcal{P}_n\\ \mathrm{supp}(P)=A}}b_PP$, we have that

$$\mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \| \mathcal{N}^{\dagger}(W^{\dagger}OW) \|_{F}^{2} = \sum_{\substack{P \in \mathcal{P}_{n} \\ \text{supp}(P) = A}} b_{P}^{2} \left\{ \frac{\|\mathbf{D}\|_{2}^{2} + \|\mathbf{t}\|_{2}^{2}}{3} \right\}^{|A|} = \left\{ \frac{\|\mathbf{D}\|_{2}^{2} + \|\mathbf{t}\|_{2}^{2}}{3} \right\}^{|A|} \|O\|_{F}^{2}.$$
(B27)

Now assume that \mathcal{N} belongs to the dephasing class. Assume without loss of generality that $D_X = D_Y = D < 1$ and $D_Z = 1$. Denote $\nu_P := \mathbb{E}_{W \sim \mathcal{D}} a_P^2$ and $\nu_{\text{max}} := (1 + 2\eta)/3$. Then we have

$$\max_{\substack{\nu_X + \nu_Y + \nu_Z = 1 \\ \nu_P \in \{0, \nu_{\text{max}}\}}} \sum_{P \in \{X, Y, Z\}} \nu_P D_P^2 = \nu_{\text{max}} + (1 - \nu_{\text{max}}) D^2 = \eta + (1 - \eta) \frac{\|\mathbf{D}\|_2^2}{3}.$$
 (B28)

The contraction coefficients proposed in Def. 8 can be related to the average-case contraction of the Frobenius norm.

Lemma 10 (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

$$\mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \left\| \mathcal{N}^{\dagger \otimes n}(W^{\dagger} O W) \right\|_{\mathcal{F}}^{2} \leqslant \sum_{P \in \mathcal{P}_{n}} \left\{ \chi_{\mathcal{D}}^{2}(\mathcal{N}) \right\}^{|P|} a_{P}^{2}. \tag{B29}$$

In particular, we also have

$$\mathbb{E}_{W \sim \mathcal{D}^{\otimes n}} \left\| \mathcal{N}^{\dagger \otimes n}(W^{\dagger}OW) \right\|_{F}^{2} - 4^{-n} \text{Tr}[O]^{2} \leqslant \left\{ \chi_{\mathcal{D}}^{2}(\mathcal{N}) \right\} \left(\|O\|_{F}^{2} - 4^{-n} \text{Tr}[O]^{2} \right). \tag{B30}$$

Proof. In general, we can write

$$O = \sum_{w=1}^{n} \sum_{\substack{A \subseteq [n] \\ |A| = w}} \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P) = A}} a_P P.$$
(B31)

We have

$$\mathbb{E}_{U \sim \mathcal{D}}(U^{\dagger}OU)^{\otimes 2} = \mathbb{E}_{U \sim \mathcal{D}} \sum_{w=1}^{n} \sum_{\substack{A \subseteq [n] \\ |A| = w}} \left(\sum_{P \in A} a_{P} U^{\dagger} P U \right)^{\otimes 2}. \tag{B32}$$

Fix a subset A of size k,

$$\left\| \mathcal{N}^{\otimes n\dagger} \left(\sum_{P \in A} a_P U^{\dagger} P U \right) \right\|_{\mathcal{F}}^2 = \left\| \mathcal{N}^{\otimes k\dagger} \left(\sum_{P \in \{X,Y,Z\}^{\otimes k}} b_P P \right) \right\|_{\mathcal{F}}^2 \leqslant \left\{ \chi_{\mathcal{D}}^2(\mathcal{N}) \right\}^k \left\| O \right\|_{\mathcal{F}}^2, \tag{B33}$$

for some suitable coefficients b_P . Putting all together

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

$$\leqslant \sum_{w=1}^{n} \sum_{\substack{A \subseteq [n] \\ |A|=w}} \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P)=A}} \left\{ \chi_{\mathcal{D}}^2(\mathcal{N}) \right\}^w a_P^2 = \sum_{P \in \mathcal{P}_n} \left\{ \chi_{\mathcal{D}}^2(\mathcal{N}) \right\}^{|P|} a_P^2. \tag{B35}$$

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.

П n.

Example 5 (Contraction coefficients). Let $\mathcal{N}_{\gamma}^{(amp)}$ be the amplitude damping channel with noise strength γ , 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

$$\chi^{2}\left(\mathcal{N}_{\gamma}^{(\text{amp})}\right) \leqslant \max_{\|\boldsymbol{a}\|_{2}^{2}=1} \sum_{Q \in \{X,Y,Z\}} a_{Q}^{2} D_{Q}^{2} + \left(\sum_{Q \in \{X,Y,Z\}} a_{Q} t_{Q}\right)^{2} \tag{B36}$$

$$\leq \|\mathbf{D}\|_{\infty}^{2} + \|\mathbf{t}\|_{2}^{2} = 1 - \gamma + \gamma^{2},$$
 (B37)

where the first step follows from Lemma 8 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 3, we know that the \mathcal{D} is an η -approximate scrambler with $\eta = 1/4$. Therefore, Lemma 9 yields the following mean squared contraction coefficient:

$$\chi_{\mathcal{D}}^2 \left(\mathcal{N}_p^{\text{(deph)}} \right) = \frac{1 + (1 - 2p)^2}{2}.$$
(B38)

3. Effective depolarizing rate of noisy random circuits

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

$$\mathcal{N} = \mathcal{N}_p^{\text{(depo)}} \circ \tilde{\mathcal{N}}_p, \tag{B39}$$

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

$$\tilde{\mathcal{N}}_p(I) = \mathcal{N}(I)/(1-p) - Ip/(1-p), \qquad (B40)$$

$$\tilde{\mathcal{N}}_p(X) = \mathcal{N}(X)/(1-p),\tag{B41}$$

$$\tilde{\mathcal{N}}_p(Y) = \mathcal{N}(Y)/(1-p), \tag{B42}$$

$$\tilde{\mathcal{N}}_p(Z) = \mathcal{N}(Z)/(1-p). \tag{B43}$$

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

$$\tilde{\mathcal{N}}_{n}^{\dagger}(I) = I \tag{B44}$$

$$\tilde{\mathcal{N}}_p^{\dagger}(X) = \mathcal{N}^{\dagger}(X)/(1-p), \tag{B45}$$

$$\tilde{\mathcal{N}}_{p}^{\dagger}(Y) = \mathcal{N}^{\dagger}(Y)/(1-p), \tag{B46}$$

$$\tilde{\mathcal{N}}_p^{\dagger}(Z) = \mathcal{N}^{\dagger}(Z)/(1-p). \tag{B47}$$

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 γ_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 non-physical map, $\tilde{\mathcal{N}}_p^{\dagger}$, does not increase the Frobenius norm of any operator on average over random noise channel directions, provided that $p \leqslant \gamma_s/2$.

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. B39 is appropriately chosen based on the contraction coefficient $\chi_{\mathcal{D}}(\mathcal{N})$.

Lemma 11. 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

$$\mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \left\| \tilde{\mathcal{N}}_p^{\dagger \otimes n}(V^{\dagger} O V) \right\|_{\mathcal{F}}^2 \leqslant \|O\|_{\mathcal{F}}^2, \tag{B48}$$

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

$$\mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \left\| \tilde{\mathcal{N}}_{p}^{\dagger \otimes n}(V^{\dagger}OV) \right\|_{F}^{2} = \mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \sum_{k=1}^{n} \sum_{\substack{A \subseteq [n] \\ |A| = k}} \left\| \sum_{\substack{P \in \mathcal{P}_{n} \\ \operatorname{supp}(P) = A}} a_{P} \tilde{\mathcal{N}}_{p}^{\dagger \otimes n}(V^{\dagger}PV) \right\|_{F}^{2}$$
(B49)

$$= \mathbb{E}_{V \sim \mathcal{D}^{\otimes n}} \sum_{k=1}^{n} \sum_{\substack{A \subseteq [n] \\ |A| = k}} \frac{1}{(1-p)^{2k}} \left\| \sum_{\substack{P \in \mathcal{P}_n \\ \text{supp}(P) = A}} a_P \mathcal{N}_p^{\dagger \otimes n}(V^{\dagger} P V) \right\|_{\mathcal{F}}^2$$
(B50)

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

$$= ||V^{\dagger}OV||_{\mathcal{F}}^{2} = ||O||_{\mathcal{F}}^{2}, \tag{B52}$$

where we used the orthogonality property of Pauli terms with different supports (Lemma 1) and the inequality follows from the definition of mean squared contraction coefficient (Definition 8). \Box

The above Lemma 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 7, it is natural to assign to it an effective depolarizing rate.

Definition 9 (Effective depolarizing rate of noisy random circuits). Let \mathcal{D}_{circ} be the circuit distribution introduced in Definition 7. We define the associated effective depolarizing rate of \mathcal{D}_{circ} as

$$p := 1 - \chi_{\mathcal{D}_{\text{single}}}(\mathcal{N}). \tag{B53}$$

Appendix C: Classical simulation of noisy random circuits

We will provide some additional notation to facilitate the proof of our main result. For convenience, we also rewrite the noisy circuit as

$$\mathcal{C} := \mathcal{V}^{\text{single}} \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_L \circ \cdots \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}_1 \tag{C1}$$

$$= \mathcal{V}^{\text{single}} \circ \mathcal{N}_{n}^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_{L} \circ \cdots \circ \mathcal{N}_{n}^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_{1}$$
(C2)

$$=\mathcal{C}_{L}\circ\mathcal{C}_{L-1}\circ\cdots\circ\mathcal{C}_{1},\tag{C3}$$

where we the layers C_j are given by

$$C_{j} := \begin{cases} \mathcal{V}^{\text{single}} \circ \mathcal{N}_{p}^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_{L} & \text{if } j = L, \\ \mathcal{N}_{p}^{(\text{depo}) \otimes n} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_{j} & \text{if } j \neq L. \end{cases}$$
(C4)

We also consider the (non-physical) linear map obtained by "taking out" the depolarizing channels:

$$\tilde{\mathcal{C}} := \mathcal{V}^{\text{single}} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_L \circ \cdots \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_1 \tag{C5}$$

$$=\tilde{\mathcal{C}}_L \circ \tilde{\mathcal{C}}_{L-1} \circ \cdots \circ \tilde{\mathcal{C}}_1, \tag{C6}$$

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

$$\tilde{\mathcal{C}}_{j} := \begin{cases} \mathcal{V}^{\text{single}} \circ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_{L} & \text{if } j = L, \\ \tilde{\mathcal{N}}^{\otimes n} \circ \mathcal{U}_{j} & \text{if } j \neq L. \end{cases}$$
(C7)

Given an observable O and an initial state ρ , we aim to approximate the expectation value

$$\operatorname{Tr}[O\mathcal{C}(\rho)] = \operatorname{Tr}[\mathcal{C}^{\dagger}(O)\rho].$$
 (C8)

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

$$C_k^{\dagger}(O) = \sum_{\substack{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}_n^{L+1}: \\ |\gamma| < k}} \langle \langle O|P_L \rangle \rangle \Phi_{\gamma}(\mathcal{C}) P_0, \tag{C9}$$

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

$$\Phi_{\gamma}(\mathcal{C}) = \langle \langle P_L | \widehat{\mathcal{C}}_L | P_{L-1} \rangle \rangle \langle \langle P_{L-1} | \widehat{\mathcal{C}}_{L-1} | P_{L-2} \rangle \dots \langle \langle P_1 | \widehat{\mathcal{C}}_1 | P_0 \rangle. \tag{C10}$$

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

$$\tilde{\mathcal{C}}^{\dagger}(O) = \sum_{\gamma = (P_0, P_1, \dots, P_L) \in \mathcal{P}^{L+1}} \langle \langle O|P_L \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_0, \tag{C11}$$

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

$$\Phi_{\gamma}(\tilde{\mathcal{C}}) = \langle \langle P_L | \widehat{\tilde{\mathcal{C}}}_L | P_{L-1} \rangle \rangle \langle \langle P_{L-1} | \widehat{\tilde{\mathcal{C}}}_{L-1} | P_{L-2} \rangle \dots \langle \langle P_1 | \widehat{\tilde{\mathcal{C}}}_1 | P_0 \rangle \rangle. \tag{C12}$$

1. Average contraction implies efficient classical simulation

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

Lemma 12. 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

$$\Phi_{\gamma}(\mathcal{C}) = (1-p)^{|\gamma|} \Phi_{\gamma}(\tilde{\mathcal{C}}). \tag{C13}$$

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

$$\langle \langle P_j | \widehat{\mathcal{C}}_j | P_{j-1} \rangle \rangle = (1-p)^{|P_j|} \langle \langle P_j | \widehat{\widetilde{\mathcal{C}}}_j | P_{j-1} \rangle \rangle. \tag{C14}$$

Iterating over al $i \in [L]$ we obtain the desired result:

$$\Phi_{\gamma}(\mathcal{C}) = (1-p)^{|P_L| + |P_{L-1}| + \dots + |P_1|} \Phi_{\gamma}(\tilde{\mathcal{C}}) := (1-p)^{|\gamma|} \Phi_{\gamma}(\tilde{\mathcal{C}}). \tag{C15}$$

The core result follows by combining the above lemma with the orthogonality properties of different Pauli paths.

Theorem 5 (Mean squared error). Let \mathcal{D}_{circ} be an L-layered locally unbiased distribution over noisy circuits, and let p be the effective depolarizing rate of \mathcal{D}_{circ} . We have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left\| \mathcal{C}_k^{\dagger}(O) - \mathcal{C}^{\dagger}(O) \right\|_{\mathcal{F}}^2 \leqslant (1 - p)^{2k} \|O\|_{\mathcal{F}}^2. \tag{C16}$$

Moreover, for any initial state ρ we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \operatorname{Tr} \left[(\mathcal{C}_k^{\dagger}(O) - \mathcal{C}^{\dagger}(O)) \rho \right]^2 \leqslant (1 - p)^{2k} \|O\|_{\mathcal{F}}^2. \tag{C17}$$

Proof. We start by proving the following technical claim.

Claim 2. For all subsets of Pauli paths $S \subseteq \mathcal{P}_n^{L+1}$, we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left(\sum_{\gamma \in \mathcal{S}} \langle \langle O | P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_0 \right)^{\otimes 2} = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \sum_{w=1}^{n} \left(\sum_{\substack{\gamma \in \mathcal{S}: \\ |\gamma| = w}} \langle \langle O | P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_0 \right)^{\otimes 2}.$$
 (C18)

Proof of the Claim. We can prove the result by exploiting the fact that paths with different support are orthogonal, i.e., their Fourier coefficients are uncorrelated. As such,

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left(\sum_{\gamma \in \mathcal{S}} \langle \langle O|P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{0} \right)^{\otimes 2} = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left(\sum_{w=1}^{n} \sum_{\substack{\gamma \in \mathcal{S}: \\ |\gamma| = w}} \langle \langle O|P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{0} \right)^{\otimes 2} \\
= \sum_{w,w'=1}^{n} \sum_{\gamma:|\gamma| = w} \sum_{\gamma':|\gamma'| = w'} \langle \langle O|P_{L+1} \rangle \rangle \langle \langle O|P'_{L+1} \rangle \rangle \underbrace{\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \Phi_{\gamma}(\tilde{\mathcal{C}}) \Phi_{\gamma'}(\tilde{\mathcal{C}})}_{=0 \text{ if } w \neq w'} P_{0} \otimes P'_{0} \\
= \sum_{w=1}^{n} \sum_{\gamma:|\gamma| = w} \sum_{\gamma':|\gamma'| = w} \langle \langle O|P_{L+1} \rangle \rangle \langle \langle O|P'_{L+1} \rangle \rangle \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \Phi_{\gamma}(\tilde{\mathcal{C}}) \Phi_{\gamma'}(\tilde{\mathcal{C}}) P_{0} \otimes P'_{0} \quad (C21) \\
= \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \sum_{w=1}^{n} \left(\sum_{\gamma \in \mathcal{S}:} \langle \langle O|P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{0} \right)^{\otimes 2}, \quad (C22)$$

where in the third step we invoked Lemma 6.

The desired results can be proven by evaluating separately the contribution of paths with different weight and using the fact that the Fourier coefficients of C and those of \tilde{C} coincide, up to a proportionality factor which is exponential in

the path weight. As for the first statement, we have

$$\mathbb{E}_{\mathcal{C}} \left\| \mathcal{C}_{k}^{\dagger}(O) - \mathcal{C}^{\dagger}(O) \right\|_{F}^{2} = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left\| \sum_{\gamma: |\gamma| \geqslant k} \langle \langle O|P_{L+1} \rangle \rangle \Phi_{\gamma}(\mathcal{C}) P_{0} \right\|_{F}^{2} = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}} \left\| \sum_{\gamma: |\gamma| \geqslant k} (1 - p)^{|\gamma|} \langle \langle O|P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{0} \right\|_{F}^{2}$$
(C23)

$$= \frac{1}{2^n} \operatorname{Tr} \left[\mathbb{F} \left(\sum_{\gamma: |\gamma| \ge k} (1 - p)^{|\gamma|} \langle \langle O|P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{\gamma} \right)^{\otimes 2} \right]$$
 (C24)

$$= \frac{1}{2^n} \sum_{w \geqslant k} (1 - p)^{2w} \operatorname{Tr} \left[\mathbb{F} \left(\sum_{\gamma : |\gamma| = w} \langle \langle O|P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{\gamma} \right)^{\otimes 2} \right]$$
 (C25)

$$\leq (1-p)^{2k} \frac{1}{2^n} \sum_{w \geq k} \operatorname{Tr} \left[\mathbb{F} \left(\sum_{\gamma: |\gamma| = w} \langle \langle O| P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{\gamma} \right)^{\otimes 2} \right]$$
 (C26)

$$\leq (1-p)^{2k} \frac{1}{2^n} \sum_{w \geq 0} \operatorname{Tr} \left[\mathbb{F} \left(\sum_{\gamma: |\gamma| = w} \langle \langle O| P_{L+1} \rangle \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{\gamma} \right)^{\otimes 2} \right]$$
 (C27)

$$= (1-p)^{2k} \mathbb{E}_{\mathcal{C}} \left\| \sum_{\gamma: |\gamma| \in \mathcal{P}_n^{L+2}} \langle \! \langle O | P_{L+1} \rangle \! \rangle \Phi_{\gamma}(\tilde{\mathcal{C}}) P_{\gamma} \right\|_{\mathcal{F}}^2 = (1-p)^{2k} \mathbb{E}_{\mathcal{C}} \left\| \tilde{\mathcal{C}}^{\dagger}(O) \right\|_{\mathcal{F}}^2 \leqslant (1-p)^{2k} \|O\|_{\mathcal{F}}^2. \quad (C28)$$

In the second step, we applied Lemma 12, while in the fourth step, we invoked Claim 2. The first inequality follows from the observation that the sum is taken over indices $w \ge k$, implying $(1-p)^w \le (1-p)^k$. The second inequality holds because all the additional terms included in the sum are positive. Finally, the last inequality is a direct consequence of Lemma 11.

The second statement be shown invoking Lemma 3

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \operatorname{Tr} \left[(\mathcal{C}_k^{\dagger}(O) - \mathcal{C}^{\dagger}(O)) \rho \right]^2 = \mathbb{E}_{\mathcal{U}_1 \sim \mathcal{D}_1, \dots, \mathcal{U}_L \sim \mathcal{D}_L} \operatorname{Tr} \left[(\mathcal{C}_k^{\dagger}(O) - \mathcal{C}^{\dagger}(O)) \mathcal{U}_1^{\dagger} \circ \mathcal{U}_1(\rho) \right]^2$$
(C29)

$$\leq \mathbb{E}_{\mathcal{U}_2 \sim \mathcal{D}_2, \dots, \mathcal{U}_L \sim \mathcal{D}_L} \left\| \mathcal{U}_1(\mathcal{C}_k^{\dagger}(O) - \mathcal{C}^{\dagger}(O)) \right\|_{\mathcal{F}}^2 \tag{C30}$$

$$= \mathbb{E}_{\mathcal{U}_2 \sim \mathcal{D}_2, \dots, \mathcal{U}_L \sim \mathcal{D}_L} \left\| \mathcal{C}_k^{\dagger}(O) - \mathcal{C}^{\dagger}(O) \right\|_{\mathcal{F}}^2 \leqslant (1 - p)^{2k} \|O\|_{\mathcal{F}}^2. \tag{C31}$$

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 non-zero contributions.

Definition 10. We say that a Pauli path $\gamma = (P_0, P_1, \dots, P_{L+1})$ is legal if its Fourier coefficient gives a non-zero contribution, i.e., if it satisfies $\langle\!\langle O|P_{L+1}\rangle\!\rangle \mathbb{E}_{\mathcal{C}\sim\mathcal{D}_{\mathrm{circ}}}\left[\Phi_{\gamma}^2(\mathcal{C})\right] \neq 0$.

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

$$P \neq Q \implies \langle\!\langle P | \widehat{\mathcal{N}'} | Q \rangle\!\rangle = 0,$$
 (C32)

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].

Lemma 13 (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 4. Let C be a quantum circuit interspersed by amplitude damping noise and let O be a k-local Hamiltonian such as

$$O = \sum_{P \in \{I,Z\}^{\otimes n}: |P| < k} P. \tag{C33}$$

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.

Lemma 14 (Runtime for arbitrary noise). Assume that each unitary layer U_j consists of non-overlapping 1 and 2-qubit gates. For any 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 non-identity terms otherwise their path weights would exceed k (we recall that P_0 does not contribute to the Pauli weight). Moreover, such non-identity terms need to be consecutive, since the adjoint of any quantum channel is unital. Thus, if L > k, then the for legal paths γ with weight smaller than k it holds that

$$P_0 = P_1 = \dots = P_{L-k+1} = I^{\otimes n}.$$
 (C34)

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

The enumeration algorithm corresponds to a Breadth-First Search (BFS) of the sub-tree 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

$$(\star, \star, \dots, \star, P_i, P_{i+1}, \dots, P_L), \tag{C35}$$

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

$$\mathbb{S}_L \leftarrow \{(\star, \star, \dots, \star, P_L)\} : \langle\!\langle P_L | O \rangle\!\rangle \neq 0 \text{ and } |P_L| < k\}$$
 (C36)

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

$$\mathbb{S}_{j} \leftarrow \left\{ (\star, \dots, \star, P_{j}, P_{j+1}, \dots, P_{L}) \right\} : \langle \langle P_{j+1} | \widehat{\mathcal{C}_{j}} | P_{j} \rangle \neq 0 \text{ and } |P_{j}| + |P_{j+1}| + \dots |P_{L}| < k \right\}. \tag{C37}$$

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

$$\mathbb{S}_0 \leftarrow \left\{ (P_0, P_1, \dots, P_L) \right\} : \langle \langle P_1 | \widehat{C_1} | P_0 \rangle \rangle \neq 0 \right\}. \tag{C38}$$

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

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

$$\sum_{w=0}^{k} \left\{ \text{ number of solutions to the equation } \sum_{i=j}^{L} |\ell_i| = w \right\}$$
 (C39)

$$= \sum_{w=0}^{k} {w + L - j \choose L - j} \le (k+1)2^{k+L} \in \exp(\mathcal{O}(k)).$$
 (C40)

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 non-identities) for each Pauli operators $P_j, P_{j+1}, \ldots, P_L$. We recall that P_L satisfies $\langle P_L | O \rangle \neq 0$ 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 ℓ_{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 $\left(c^{\ell_{i+1}}\right)^4 = c^{4\ell_{i+1}}$ different choices. Overall, the number of possible choices for fixed values of $\ell_i, \ell_{j+1}, \ldots, \ell_L$ is at most

$$\prod_{i=j}^{L} c^{4\ell_{i+1}} = c^{4\sum_{i=j}^{L} \ell_{i+1}} \leqslant c^{4k} \in \exp(\mathcal{O}(k)).$$
 (C41)

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

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

$$\prod_{i=j}^{L} 3^{\ell_{i+1}} = 3^{\sum_{i=j}^{L} \ell_{i+1}} \leqslant 3^k \in \exp(\mathcal{O}(k)).$$
 (C42)

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 $\langle P_{j+1} | \widehat{\mathcal{C}}_j | P_j \rangle$ 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

$$M n \exp(\mathcal{O}(k)),$$
 (C43)

Combining Lemma 14 with Theorem 5, we obtain the following upper bound on the time complexity.

Theorem 6 (Time complexity). I've added a few comments. I didn't address them because Let C be a noisy quantum circuit sampled from the distribution \mathcal{D}_{circ} with effective depolarizing rate $p \in \Theta(1)$

- If the noise channel $\mathcal N$ is unital, then there exists a classical algorithm running in time $\operatorname{poly}(n,1/\epsilon,1/\delta)$ that approximates $\operatorname{Tr}[\mathcal O\mathcal C(\rho)]$ within additive error $\epsilon\|\mathcal O\|_{\mathbb F}$ with probability at least $1-\delta$ over the choice of $\mathcal C$.
- 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 poly $(n, 1/\epsilon, 1/\delta)$ that approximates $\text{Tr}[OC(\rho)]$ within additive error $\epsilon ||O||_F$ with probability at least 1δ over the choice of C.

Appendix D: Effective depth of noisy random circuits beyond local 2-designs

We conclude our theoretical investigation with a generalization of the main finding of Ref. [37] to the broad classes of circuits considered in this work.

Theorem 7. For all observables O it holds that

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left(\left\| \mathcal{C}^{\dagger}(O) \right\|_{\text{F}}^{2} - 4^{-n} \operatorname{Tr} \left[\mathcal{C}^{\dagger}(O) \right] \right) \leqslant (1 - p)^{2L} \|O\|_{\text{F}}^{2}. \tag{D1}$$

Moreover, given two quantum states ρ, σ , we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \operatorname{Tr}[O\mathcal{C}(\rho - \sigma)]^2 \leqslant 4(1 - p)^{2L} \|O\|_{\mathcal{F}}^2. \tag{D2}$$

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

$$\mathbb{E}_{\mathcal{C}_{j}}\left(\left\|\mathcal{C}_{j}^{\dagger}(O)\right\|_{\mathcal{F}}^{2}-4^{-n}\operatorname{Tr}\left[\mathcal{C}_{j}^{\dagger}(O)\right]\right)=\mathbb{E}_{\mathcal{C}_{j}}\sum_{w=1}^{n}\left\|\tilde{\mathcal{C}}_{j}^{\dagger}\circ\mathcal{N}_{p}^{(\text{depo})\otimes n}\left(\sum_{P:|P|=w}a_{P}P\right)\right\|_{\mathcal{F}}^{2}$$
(D3)

$$= \mathbb{E}_{\mathcal{C}_j} \sum_{w=1}^n (1-p)^{2w} \left\| \tilde{\mathcal{C}}_j^{\dagger} \left(\sum_{P:|P|=w} a_P P \right) \right\|_{\mathcal{D}}^2 \tag{D4}$$

$$\leq (1-p)^2 (\|O\|_F^2 - 4^{-n} \operatorname{Tr}[O]),$$
 (D5)

where the inequality follows from Lemma 11. Iterating over all j we obtain the desired result.

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

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \operatorname{Tr}[O\mathcal{C}(\rho - \sigma)]^2 = \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \operatorname{Tr}\left[\left(O - \operatorname{Tr}[O]\frac{I^{\otimes n}}{2^n}\right)\mathcal{C}(\rho - \sigma)\right]^2$$
(D6)

$$\leq 2\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \operatorname{Tr} \left[\left(O - \operatorname{Tr}[O] \frac{I^{\otimes n}}{2^n} \right) \mathcal{C}(\rho) \right]^2 + 2\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \operatorname{Tr} \left[\left(O - \operatorname{Tr}[O] \frac{I^{\otimes n}}{2^n} \right) \mathcal{C}(\sigma) \right]^2$$
(D7)

$$\leq 4\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \left(\left\| \mathcal{C}^{\dagger}(O) \right\|_{F}^{2} - 4^{-n} \operatorname{Tr} \left[\mathcal{C}^{\dagger}(O) \right] \right) \tag{D8}$$

$$\leq 4(1-p)^2 \left(\|O\|_F^2 - 4^{-n} \operatorname{Tr}[O] \right),$$
 (D9)

where the second inequality follows by Lemma 3

We also obtain the following corollary for the trace distance.

Corollary 1. Given two quantum states ρ, σ , we have

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \| \mathcal{C}(\rho - \sigma) \|_1^2 \leqslant 4^{n+1} (1 - p)^{2L}. \tag{D10}$$

Proof. We will exploit the inequality $\|\cdot\cdot\cdot\|_1^2 \leq 2^n\|\cdot\cdot\cdot\|_2^2$ 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

$$\mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{circ}} \| \mathcal{C}(\rho - \sigma) \|_{1}^{2} \leq 2^{n} \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{circ}} \| \mathcal{C}(\rho - \sigma) \|_{2}^{2}$$
(D11)

$$= \mathbb{E}_{\mathcal{C} \sim \mathcal{D}_{\text{circ}}} \sum_{P \in \mathcal{P}_n} \text{Tr}[P\mathcal{C}(\rho - \sigma)]^2 \leqslant 4^{n+1} (1-p)^{2L}, \tag{D12}$$

where we invoked Theorem 7 in the last step.

Appendix E: Additional numerical results

This section provides additional numerical results that complement those presented in the main text. As discussed in the main text, simulating deep noisy random circuits with non-unital noise (e.g., amplitude damping) is arguably more interesting than simulating circuits in noiseless or depolarizing scenarios, where simply outputting zero is often sufficient [26, 56].

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 to accurately simulate such noisy circuits, even at high depths.

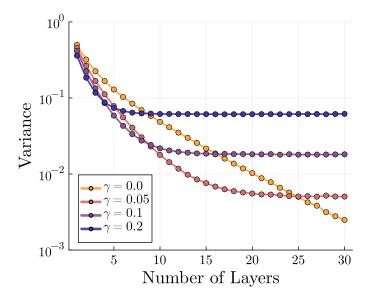


Figure 6. Absence of barren plateaus with amplitude damping noise. The variance, estimated via Theorem 4 with 10^6 randomly sampled Pauli paths, of a Z expectation value in the middle of a 1D lattice is shown for a system of 60 qubits. Here, we used the same ansatz as in Figs. 3 and 4, specifically a 1D bricklayer topology of RX-RZ-RZZ gates. The parameter γ in the plot refers to the amplitude damping rate.