Estimates of Loss Function Concentration in Noisy Parametrized Quantum Circuits

Giulio Crognaletti,[1,2,3,*] Michele Grossi,[3,†] and Angelo Bassi[1,2]

[1] Department of Physics, University of Trieste, Strada Costiera 11, 34151 Trieste, Italy
[2] Istituto Nazionale di Fisica Nucleare, Trieste Section, Via Valerio 2, 34127 Trieste, Italy
[3] European Organization for Nuclear Research (CERN), 1211 Geneva, Switzerland

(Received 21 May 2025; revised 29 December 2025; accepted 16 March 2026; published 26 May 2026)

Variational quantum computing offers a powerful framework with applications across diverse fields such as quantum chemistry, machine learning, and optimization. However, its scalability is hindered by the exponential concentration of the loss function, known as the barren plateau problem. While significant progress has been made and prior work has separately analyzed barren plateaus in unitary and noisy settings, their combined impact remains poorly understood, largely due to limitations in conventional Lie-algebraic approaches. In this work, we introduce an analytical framework based on non-negative matrix theory that enables the description of the variance in layered noisy quantum circuits with arbitrary noise channels. This approach enables the derivation of exact expressions in the deep-circuit regime, uncovering the complex interplay between unitary layers and noise. Notably, we identify a noise-induced absorption mechanism—a phenomenon absent in purely unitary dynamics—which provides new insight into how noise shapes circuit behavior. We further present a controlled convergence analysis, establishing general lower bounds on the variance of both deep and shallow circuits. This leads to a principled connection between noise resilience and the expressive capacity of parameterized quantum circuits, particularly under smart initialization strategies. Our theoretical results are supported by numerical simulations and illustrative applications.

DOI: 10.1103/dw2h-ll6r

I. INTRODUCTION

Quantum computers hold the potential for substantial speed increases across various computational tasks, with a notable example being their capacity to transform our understanding of nature through the simulation of quantum systems [1–3]. In this context, variational quantum computing offers a versatile tool, which combines quantum and classical computational resources. The flexibility and relative simplicity of this approach, coupled with its potential for noise resilience, make it a compelling candidate for near-term quantum devices [4,5].

In the past years, a substantial effort has been put in by the community to unlock the potential of variational algorithms. A major hurdle in this direction is the barren plateau (BP) effect, which implies an exponential flattening of the loss landscape, making the optimization step unfeasible with a polynomial amount of quantum resources [6]. More precisely, in the presence of BP, we have an exponentially vanishing probability of being able to efficiently find a loss-minimizing direction after parameter initialization. In the absence of noise, this phenomenon has been linked to several factors, ranging from the expressive power and entangling capability of the circuit [7–12] to the locality of H [13,14] and the entanglement of the initial state ρ [15]. Recently, the contributions coming from all such factors were unified in a Lie-algebraic framework [16–18], showing how barren plateaus ultimately arise as a *curse of dimensionality*. Several BP mitigation strategies have been proposed to circumvent the issue. Among the most popular, there are small-angle initializations [19–23], which leverage the restriction of the domain Θ of θ to limit the expressive power of the circuit, and consequently avoid concentration.

In addition to that, the presence of noise is also deemed detrimental, as it often gives rise to noise-induced barren plateau (NIBP) [24,25] and symmetry breaking [26,27]. The former exacerbates the BP effect, producing a deterministic concentration. In this case, an efficient loss-minimizing direction is exponentially hard to estimate,

*Contact author: giu.crognaletti@gmail.com
†Contact author: michele.grossi@cern.ch

regardless of the parameter's initialization choice. These phenomena are mostly linked to decoherence. Reference [28] offers a review of the subject.

Indeed, the action of noise can be quite destructive for quantum computation. Along these lines, recent research [29–31] puts stringent fundamental bounds on capabilities of quantum error mitigation strategies, emphasizing the limits of scalability of near-term devices, especially in the presence of global depolarizing noise. However, such analyses are mostly limited to unital noise, and besides, they typically rely on the properties of circuit ensembles. Indeed, studies have demonstrated that nonunital noise can significantly influence the emergence of NIBP [32–34], underscoring that the interplay between general noise and unitary circuit layers in variational quantum computing remains a critical and unresolved area of research.

In this work, we approach the problem through a novel formulation grounded in non-negative matrix theory, which enables a unified and general description of these dynamics. This framework not only addresses the previously uncharted aspects of the problem but also seamlessly encompasses both unital and nonunital noise settings. We quantify the deep circuit variance $\mathbb{V}^{\infty}_{\rho,H}$ for layered circuits composed of local 2-designs interleaved with general noise maps, providing a comprehensive picture. More specifically, we are able to recover known results for unitary circuits and strictly contractive noise maps as limiting cases, while unveiling the emergence of a more complex phenomenon, namely, *absorption*, in the intermediate case. Crucially, this can occur only with noise maps that are not strictly contractive—that is, maps that are insufficiently strong to fully erase all quantum information from the system. It is worth emphasizing that this class encompasses both unital and nonunital noise channels, illustrating that unitality of the noise alone is not necessarily detrimental for parameterized quantum circuits (PQCs). The proposed framework also facilitates a systematic analysis of the noise resilience properties of PQCs, providing a foundation for principled BP mitigation strategies. Notably, we demonstrate a direct correlation between a circuit's resilience to noise and its capacity for enhanced expressivity under small-angle initialization-like schemes.

II. MOTIVATING EXAMPLES

In the following, we consider n-qubit quantum systems with a Hilbert space $\mathcal{H} = \bigotimes_{m=1}^{M} \mathcal{H}_m$ of dimension d, divided into M subsystems, each of dimension d_m, representing either single qubits or groups of qubits. More specifically, we study the problem where a quantum state ρ is evolved using a parameterized quantum channel Φ_θ, whose parameters $\theta = (\theta_1, \theta_2, \ldots) \in \Theta$ are optimized by minimizing the loss function

$$\mathcal{L}_{\rho,H} = \mathrm{Tr}\left[\Phi_\theta(\rho) H\right], \tag{1}$$

where H is an observable of the system. Here, the presence of BP is diagnosed by studying the variance $\mathbb{V}_{\rho,H}$ of $\mathcal{L}_{\rho,H}$ for varying parameters θ. In particular, we say that $\mathcal{L}_{\rho,H}$ suffers from BP if $\mathbb{V}_{\rho,H} \in O(e^{-\beta n})$, $\beta > 0$, as in this case, the loss function exponentially concentrates around its mean value in the number n of qubits. Within this framework, we focus on the case of *layered* quantum channels, namely,

$$\Phi_\theta = \mathcal{U}_{\theta_{L+1}} \circ \mathcal{E}_L \circ \mathcal{U}_{\theta_{L-1}} \cdots \circ \mathcal{E}_1 \circ \mathcal{U}_{\theta_1}, \tag{2}$$

where each layer is composed of a unitary part $\mathcal{U}_{\theta_l}$ and an arbitrary quantum channel $\mathcal{E}_l$. If we denote by $\mathcal{B}$ the space of bounded operators acting on $\mathcal{H}$, both such components can be regarded as linear, completely positive functions mapping $\mathcal{B}$ to itself. Furthermore, the intermediate channels $\mathcal{E}_l$ will be deemed strictly contractive if the restriction to the Hermiatian, traceless hyperplane $\mathbb{H}_0 \subset \mathcal{B}$ of its adjoint action $\mathcal{E}_l^\dagger$ has an induced Schatten 1-norm $\|\mathcal{E}_l^\dagger|_{\mathbb{H}_0}\|_1$ strictly less than unity, $\|\mathcal{E}_l^\dagger|_{\mathbb{H}_0}\|_1 < 1$, and contractive otherwise, i.e., $\|\mathcal{E}_l^\dagger|_{\mathbb{H}_0}\|_1 = 1$. According to this definition, contractive channels preserve the 1-norm of certain operators in $\mathcal{B}$, while possibly being strictly contractive for others. Moreover, unitary channels, which preserve all norms across their entire domain, are a special case of contractive channels that are never strictly contractive. We refer to Appendix E for more details.

Since $\mathbb{V}_{\rho,H}$ is ultimately upper-bounded by the norm of H, it is intuitively clear that strictly contractive channels lead to vanishing variance in the deep circuit limit. However, it is important to clarify that this behavior is not inherently linked to the unitality of each individual $\mathcal{E}_l$. Although certain results in the literature may be misconstrued as implying that unital, nonunitary channels inevitably induce deterministic concentration—i.e., NIBP—in deep circuits [24,34], this interpretation does not hold in general.

A simple yet compelling counterexample involves commuting noise and unitary layers, i.e., when $\mathcal{E}_l \circ \mathcal{U}_{l'}(\rho) = \mathcal{U}_{l'} \circ \mathcal{E}_l(\rho)$ for all ρ and all l, l'. In such cases, one can commute all noisy operations to the beginning of the computation, effectively reducing the noisy circuit to a fully unitary one, provided the initial state is redefined as $\tilde{\rho} = \mathcal{E}_L \circ \cdots \circ \mathcal{E}_1(\rho)$. If ρ is a steady state of all intermediate channels $\mathcal{E}_l$, then $\tilde{\rho} = \rho$, and the entire circuit becomes noise-free. Consequently, it cannot exhibit NIBP, although BP may still occur. An explicit example of this is provided in Appendix G.

Similarly, even in the noncommuting case, the emergence of concentration is not guaranteed. In fact, the variance can even increase compared to the purely unitary circuit. This counterintuitive behavior can result from noise-induced symmetry breaking [26,27], a phenomenon typically studied in nonunital channels but also possible in unital ones. In both cases, this effect can lead to an enhanced variance. An example illustrating this mechanism is also presented in Appendix G.

Altogether, the above discussion highlights that, beyond the nontrivial effects observed in nonunital channels, the broader class of contractive, but not *strictly* contractive channels gives rise to a rich variety of largely unexplored variance behaviors. In the following sections we will characterize these scenarios for general channels, encompassing the previously discussed cases as specific instances.

III. GENERAL QUANTUM CHANNELS

To accurately characterize the effects of general noise maps, it is essential to first analyze the structural transformations induced on the computational space by the unitary layers. In particular, it is worth noticing that the subdivision of $\mathcal{H}$ into local subsystems induces also a partition of $\mathcal{B}$. More specifically, if we denote by $\kappa \in \{0,1\}^M$ a binary string of length M, we can split $\mathcal{B}$ into *local* subspaces $\mathcal{B}_\kappa \subset \mathcal{B}$, each spanned by the *traceless* operators acting nontrivially on $\mathcal{H}_m$ if and only if $\kappa_m = 1$. Indeed, we can partition the whole space as

$$\mathcal{B} = \bigoplus_{\kappa\in\{0,1\}^M} \mathcal{B}_\kappa, \tag{3}$$

with $d_\kappa = \dim(\mathcal{B}_\kappa) = \prod_m (d_m^2 - 1)^{\kappa_m}$. Clearly, if $\kappa = 0$, $\mathcal{B}_0 = \operatorname{span}\mathbb{1}$. In this work, such decomposition is delineated as *locality* from $\mathcal{H}$ to $\mathcal{B}$. Specifically, we say that $A \in \mathcal{B}$ is κ-local if $A \in \mathcal{B}_\kappa$, and we associate with A a locality vector $\ell_A \in \mathbb{R}^{2^M}$ defined elementwise by

$$(\ell_A)_\kappa = \sum_{j=1}^{d_\kappa} \operatorname{Tr}\left[B_j A\right]^2, \tag{4}$$

for some Hermitian, orthonormal basis $\{B_j\}_{j=1}^{d_\kappa}$ of $\mathcal{B}_\kappa$. Clearly, from the definition, a κ-local operator A has locality vector $(\ell_A)_\lambda = \delta_{\kappa,\lambda}\|A\|_2^2$, where $\|\cdot\|_2$ is the Hilbert-Schmidt norm. We remark that similar quantities are known in the context of PQCs, and in fact the vector ℓ_A is analogous to the purity measures defined in [16,18]. Based on this, a derived notion of *locality preservation* can be introduced for linear maps acting on $\mathcal{B}$. In particular, given a map $\Lambda : \mathcal{B} \to \mathcal{B}$ and two subspaces $\mathcal{B}_\kappa$ and $\mathcal{B}_\lambda$, we can measure the degree to which Λ is able to put them in communication. The more subspaces are connected, the less locality preserving the map Λ will be. This idea is captured formally in the following definition of a locality transfer matrix (LTM).

Definition III.1 (Locality transfer matrix). Given a linear map $\Lambda : \mathcal{B} \to \mathcal{B}$, its locality transfer matrix T is defined elementwise as

$$T_{\kappa,\lambda} = \frac{1}{d_\kappa}\sum_{j=1}^{d_\kappa} (\ell_{\Lambda(B_j)})_\lambda, \tag{5}$$

for some Hermitian, orthonormal basis $\{B_j\}_{j=1}^{d_\kappa}$ of $\mathcal{B}_\kappa$.

In this formalism, *locality preserving* transformations reflect all maps whose LTM coincides with the identity, i.e., $T_{\kappa,\lambda} = \delta_{\kappa,\lambda}$. Trivially, unitary maps, separable with respect to the partition, namely, $\mathcal{U} : \rho \mapsto U\rho U^\dagger$, where $U = \bigotimes_m U_m$, are locality preserving.

With this description in mind, we assume that $\mathcal{U}_\theta$ of Eq. (2) is locality preserving and hence describes an ideal operation limited to the local subsystems $\mathcal{H}_m$, while each channel $\mathcal{E}_l$ encodes both the operations which entangle the subsystems as well as any interaction between the system and the environment. This formally captures the idea of a variational quantum algorithm running on a real, noisy, device, where quantum computation can be realized very precisely within each subsystem, but is still inaccurate when dealing with more complex entangling operations. Furthermore, we assume that, within each subsystem and for all layers l, $\mathcal{U}_{\theta_l}$ is deep enough to form an approximate 2-design over the local unitary groups $U(d_m)$. This assumption is justified by the fact that such local operations are relatively inexpensive and that the necessary depth can be very small even for large systems, since in general it depends on d_m rather than d. Putting everything together, we can give a formal definition of the family of PQCs covered by our analysis, which we call *layered, locally random channels*.

Definition III.2 (Layered, Locally Random Channel). Given a n-qubit quantum system described by the Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, and a decomposition into M local subsystems $\mathcal{H}_m$ such that $\mathcal{H} = \bigotimes_{m=1}^M \mathcal{H}_m$, a parameterized quantum channel $\Phi_\theta : \mathcal{H} \to \mathcal{H}$ is said to be *layered* if

$$\Phi_\theta = \mathcal{U}_{\theta_{L+1}} \circ \mathcal{E}_L \circ \mathcal{U}_{\theta_{L-1}} \cdots \circ \mathcal{E}_1 \circ \mathcal{U}_{\theta_1}, \tag{6}$$

where $\mathcal{E}_l$ are arbitrary channels and $\mathcal{U}_{\theta_l}$ are unitary parametrized channels. Furthermore, such channel is also said to be *locally random* if $\mathcal{U}_{\theta_l}$ is separable w.r.t. the partition, and each local operation $U_{\theta_{lm}}$ form a unitary 2-design over the local unitary group $U(d_m)$.

A crucial property in the following analysis is the relation between the LTMs of the channel $\mathcal{E}_l$ and its Hermitian adjoint $\mathcal{E}_l^\dagger$ with respect to the Hilbert-Schmidt scalar product, namely, T and $T^\dagger$. Such relation can be characterized for generic linear maps $\Lambda : \mathcal{B} \to \mathcal{B}$ as a direct consequence of Definition III.1 and, in particular, we have $TD = (T^\dagger D)^t$, with $D_{\kappa,\lambda} = d_\kappa \delta_{\kappa,\lambda}$. For the sake of readability, here we introduce a shorthand notation for the scalar product $(\cdot,\cdot)$ in $\mathbb{R}^{2^M}$ such that T and $T^\dagger$ are *also* Hermitian adjoint of one another, i.e.,

$$(a,b) = a^t D^{-1} b = \sum_\kappa \frac{a_\kappa b_\kappa}{d_\kappa}. \tag{7}$$

We refer the reader to Appendix B for a more detailed discussion. Having introduced the main tools, we now characterize the scaling of the variance $\mathbb{V}^L_{\rho,H}$ of $\mathcal{L}_{\rho,H}$ for layered, locally random channels.

IV. RESULTS

The overarching goal is to characterize the properties of the variance $\mathbb{V}^L_{\rho,H}$ as a function of the number L of layers. To this end, we will provide a formal expression for the general case; we will then use it to explicitly compute the variance of $\mathcal{L}_{\rho,H}$ in the deep circuit limit, as well as to set lower bounds for shallow circuits. We refer to Fig. 1 for a schematic summary of the main results of this work.

A. Loss variance calculation

The first task is to derive a formal expression for the variance $\mathbb{V}^L_{\rho,H}$ in terms of the properties of ρ, H and the intermediate channels, showing its relevance and main domains of applicability. Our first result, given in the following proposition, serves as the foundation of the subsequent arguments.

Proposition IV.1 (General formula). Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition III.2. Then, we have

$$\mathbb{E}_\theta \left\{ \mathrm{Tr}\left[\Phi_\theta(\rho)H\right]^2 \right\} = \left(\ell_\rho, \prod_{l=1}^{L} T_l \, \ell_H \right), \tag{8}$$

where $(\cdot,\cdot)$ is the scalar product defined in Eq. (7), and each T_l is the LTM associated with the respective $\mathcal{E}_l^\dagger$.

Exploiting the fact that, according to Definition III.2, each layer $\mathcal{U}_{\theta_l}$ forms a global unitary 1-design, Proposition IV.1 allows a direct calculation of $\mathbb{V}^L_{\rho,H}$ as a function of the LTMs of the intermediate channels as

$$\mathbb{V}^L_{\rho,H} = \left(\ell_\rho, \prod_{l=1}^{L} T_l \, \ell_H \right) - \frac{\mathrm{Tr}\,[H]^2}{d^2}. \tag{9}$$

A proof of the aforementioned statements is given in Appendix D.

Note that this formulation provides an exact formula, which in principle gives access to the study of loss function concentration for any L. Indeed, this is the case when there are few subsystems, since in this case one can easily estimate and manipulate the matrices T_l (see Appendix I for an example). This approach becomes rather cumbersome in a general setting, for large systems, since the resources needed to represent the LTM can grow exponentially. Nevertheless, as shown in the following sections,

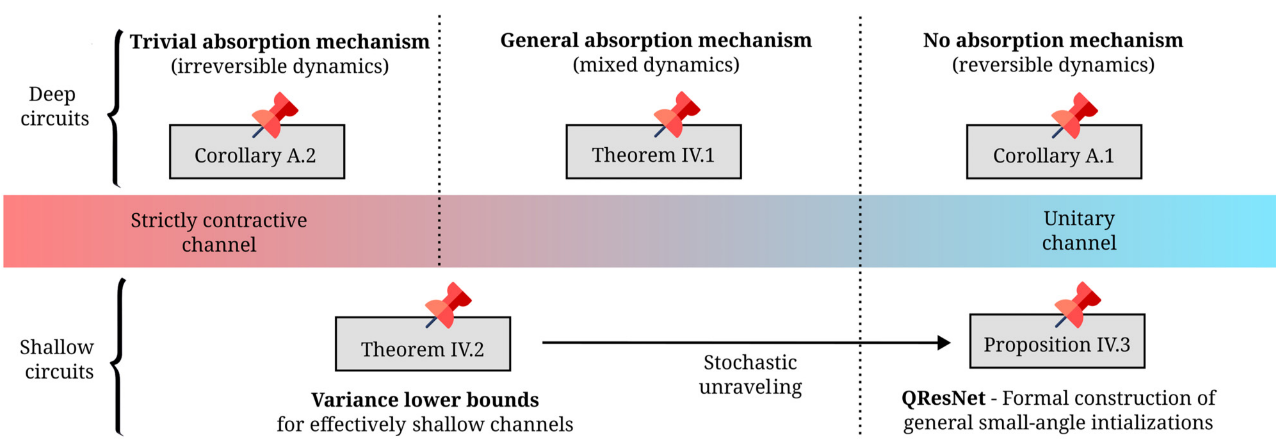

FIG. 1. Loss concentration in variational quantum computing. The analytical framework introduced here uses non-negative matrix theory to capture the interplay between unitary dynamics and noise through a noise-induced absorption mechanism. This enables precise calculation of loss variances for generic noise channels—from strictly contractive to unitary—as illustrated by the colored band in the figure. The upper panel considers deep circuits, where the loss variance $\mathbb{V}^L_{\rho,H}$ saturates to its asymptotic limit. While loss concentration is well understood for strictly contractive [24,25] (irreversible) and unitary [16–18] (reversible) channels, Theorem IV.1 provides an analytical solution for the intermediate mixed-dynamics regime. This reveals a general absorption mechanism unique to this case, explicitly quantified in Appendix C, and shown to reduce to known results in the appropriate limits (Appendix A). The lower panel addresses shallow circuits, establishing a lower bound on $\mathbb{V}^L_{\rho,H}$ via Theorem IV.2. This generalizes previous results on brickwork circuits [13,32] and formally constructs barren-plateau-free ansätze—such as small-angle initializations [21,22]—as stochastic unravelings of weak-noise channels (Sec. IV D). Moreover, it broadens their scope, showing that both unitary and nonunitary QResNets arise naturally from Proposition IV.3.

Proposition IV.1 can still be profitably used as a theoretical tool, as it allows characterizing $\mathbb{V}^L_{\rho,H}$ in both the deep and shallow circuit limit.

B. Deep circuit limit

As the circuits become deeper, it is natural to expect the contribution of the leading eigenvectors of T_l to become dominant, as repeatedly multiplying them gives rise to a process analogous to power iteration methods. The structure of such eigenvectors captures several interesting properties of the interaction between the unitary and nonunitary parts of Φ_θ, particularly *absorption*, which can only arise in this picture.

For simplicity, let us focus on *homogeneous* channels, i.e., we fix $\mathcal{E}_l = \mathcal{E}\,\forall l$, and consider, without loss of generality, a traceless observable H, $\mathrm{Tr}\,[H] = 0$. In this case, Eq. (9) becomes $\mathbb{V}^L_{\rho,H} = \left(\ell_\rho, T^L \ell_H\right)$. Note that, by construction, T is a non-negative matrix, namely, $T_{\kappa,\lambda} \geq 0\,\forall \kappa, \lambda$, and thus it can always be expressed in canonical, block-upper triangular form, where each of the diagonal blocks is irreducible [35]. Notably, this can always be achieved via conjugation with a permutation matrix, implying that, up to a relabeling of the subspaces $\mathcal{B}_k$, T is always of this form. Throughout this work, irreducible components of T will be regarded as *essential* if they cannot lead outside the block and will be denoted by T_z. Otherwise, an irreducible block will be deemed *inessential*, and their collection will be denoted by Q. A useful pictorial representation of these possibilities is shown in Fig. 2, together with the matrix canonical form. A foundational reference for this definition can be found in Ref. [35], while a more detailed discussion is proposed in Appendix E.

By simple matrix multiplication, we can in this way decompose $\mathbb{V}^L_{\rho,H}$ into a sum of terms, coming from each irreducible component T_z and Q, namely,

$$\begin{aligned}\mathbb{V}^L_{\rho,H} &= (\ell_\rho, T^L \ell_H)\\ &= \sum_z \left[(\ell_\rho, T_z^L \ell_H) + (\ell_\rho, A_z^{(L)} \ell_H)\right] + (\ell_\rho, Q^L, \ell_H),\end{aligned} \tag{10}$$

where $A_z^{(L)} = \sum_{l=0}^{L-1} T_z^l R Q^{L-1-l}$ (see Lemma E3 for a formal proof). Given this structure, it will be particularly useful to denote by $\mathcal{B}_z = \bigoplus_{\kappa \in T_z} \mathcal{B}_\kappa$ the union of subspaces put in communication within T_z, by d_z their total dimension, and given $A \in \mathcal{B}$, by $(\ell_A)_z = \sum_{\kappa \in T_z} (\ell_A)_\kappa$ the corresponding locality.

Regardless of the specific channel $\mathcal{E}$, several general properties of the blocks T_z can be identified. In particular, trace preservation implies that the trivial subspace $\mathcal{B}_0$ always gives rise to an essential block of T, denoted by T_0. Since T_z blocks are essential, it follows that $\mathcal{E}^\dagger(\mathcal{B}_z) \subset \mathcal{B}_z$. Moreover, complete positivity guarantees that all blocks T_z are contractive in spectral radius, i.e., $\rho(T_z) \leq 1$. It can further be shown that $\rho(Q) < 1$ for any channel. These observations allow us to split the sum in Eq. (10) according to the spectral radius of the blocks

$$\begin{aligned}\mathbb{V}^L_{\rho,H} = &\sum_{z\,|\,\rho(T_z)=1} \left[(\ell_\rho, T_z^L \ell_H) + (\ell_\rho, A_z^{(L)} \ell_H)\right] +\\ &+ \sum_{z\,|\,\rho(T_z)<1} \left[(\ell_\rho, T_z^L \ell_H) + (\ell_\rho, A_z^{(L)} \ell_H)\right] + (\ell_\rho, Q^L \ell_H).\end{aligned} \tag{11}$$

This decomposition is useful because the two groups of terms exhibit markedly different asymptotic behaviors. As $L \to \infty$, repeated applications of Q and of all blocks T_z with $\rho(T_z) < 1$ yield exponentially decaying contributions, which therefore vanish in the limit, as follows from

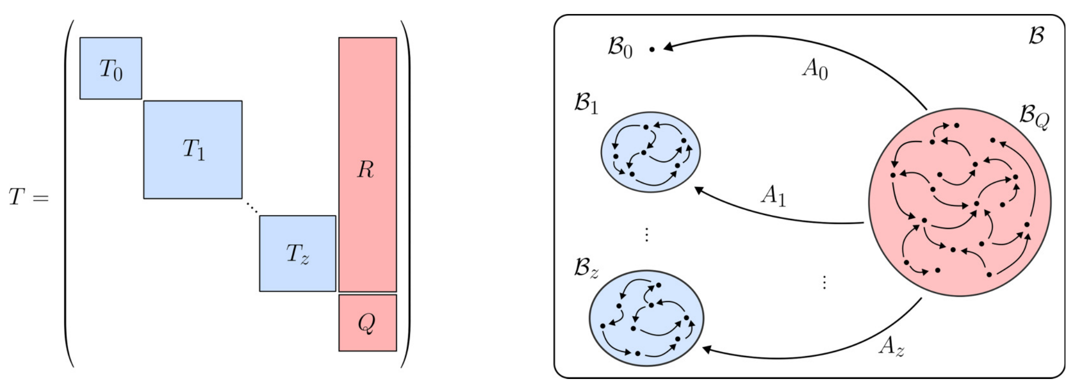

FIG. 2. Graphical representations of the stochastic process describing $\mathbb{V}^L_{\rho,H}$. On the left, we show the structure of the general locality transfer matrix, highlighting the decomposition into irreducible components [35]. Light blue blocks represent irreducible, essential components of T, while red blocks are related to inessential ones. In particular, Q represents the collection of all irreducible, inessential components of T and R and their relation with the essential components T_z. On the right, the same process is represented graphically, in terms of the local subspaces $\mathcal{B}_\kappa$. Here, each dot represents a single subspace, while the arrows represent the adjoint action of the channel $\mathcal{E}$. Essential and inessential components share the same color code of T.

Gelfand's formula and is rigorously shown in Appendix E. By contrast, when $\rho(T_z) = 1$, the iterated action of T_z converges to the projector onto its dominant eigenvalue $r_z = 1$. Since T_z is non-negative and irreducible, Perron's theorem implies that this projector has rank one and can be written as $w_z v_z^t$, where w_z and v_z denote the corresponding left and right dominant eigenvectors. In this case, v_z can be computed explicitly and satisfies $(v_z)_\kappa = 1$ for all κ. Moreover, as shown in Appendix E, the associated terms $A_z^{(L)}$ also remain nonvanishing and converge to a fixed matrix A_z, explicitly derived in the proof of Theorem IV.1.

Collecting these results, we find that in the limit $L \to \infty$ all terms in the second row of Eq. (11) vanish. Substituting the limiting eigenprojectors then yields the general expression

$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z \,|\, \rho(T_z)=1} (\ell_\rho, w_z)(\ell_H)z + (\ell_\rho, w_z)(A\ell_H)_z, \quad (12)$$

where A is a matrix with the same structure as R in Fig. 2, collecting the contributions from all matrices A_z.

Such reasoning is formally encapsulated in the following theorem.

Theorem IV.1 (Deep circuit variance). Let $\rho, H \in \mathcal{B}$ and let Φ_θ a be layered, locally random quantum channel as in Definition III.2. Then the Cesàro average of $\mathbb{V}^L_{\rho,H}$ converges to Eq. (12), and we have

$$\left| \frac{1}{L} \sum_{l=0}^{L} \mathbb{V}^{l}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H} \right| \in O\left(e^{-\beta L} \|H\|_2^2\right), \quad (13)$$

for some constant $\beta > 0$. Additionally, if all essential blocks are aperiodic (i.e., with period $p = 1$), then $\mathbb{V}^L_{\rho,H}$ is convergent, and we have

$$\left| \mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H} \right| \in O\left(e^{-\beta L} \|H\|_2^2\right), \quad (14)$$

where the absorption matrix is given by $A = R(\mathbb{1} - Q)^{-1}$.

As a special case, if the intermediate channel can be reduced unitarily to a tensor product of single-qubit channels, then also w_z of Eq. (12) can be explicitly computed, yielding the following corollary.

Corollary IV.1. If the intermediate channel takes the form $\mathcal{E} = \mathcal{N} \circ \mathcal{W}$, where the noise channel $\mathcal{N} = \bigotimes_m \mathcal{N}_m$ is the composition of single-qubit channels and $\mathcal{W}$ is unitary, then

$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z \,|\, \rho(T_z)=1} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z (A\ell_H)_z}{d_z}. \quad (15)$$

The proof of Theorem IV.1 and Corollary IV.1 can be found in Appendix E, where we complete our analysis computing w_z also for global Pauli noise.

It is worth noting that the loss variance of the circuit need not converge. In fact, while it is bounded (i.e., $\mathbb{V}^L_{\rho,H} \leq \|H\|_\infty^2 \,\forall L$), different circuit subsequences can in general lead to different limits. This is connected to the presence of *cycles* of period $p > 1$ in each T_z and can arise, for instance, when the entangling operation is not chosen carefully with respect to the partition of $\mathcal{H}$. In those cases, the value of $\mathbb{V}^L_{\rho,H}$ will depend on which equivalence class of integers modulo p the depth L belongs to. This is put into an effective example and further discussed Appendix I. For simplicity, in the following, we will assume *aperiodicity* for each irreducible block.

We start the analysis of Theorem IV.1 by unpacking Eq. (15). In this equation, $\mathbb{V}^{\infty}_{\rho,H}$ is shown to depend on four important quantities, namely, the invariant subspaces $\mathcal{B}_z$ of $\Phi^\dagger_\theta$, the respective locality of ρ and H, together with the matrix A. As a reminder, in order for this structure to arise, the invariant subspaces of $\mathcal{U}^\dagger_\theta$ and $\mathcal{E}^\dagger$ have to be *well-aligned*, so that nontrivial subspaces $\mathcal{B}_z \subset \mathcal{B}$ are present. Failure to achieve this will cause the emergence of noise-induced concentration in the sense of Ref. [24], as shown in Appendix A. This extends the notion of alignment already introduced in the literature [28] for ρ, H, and unitary circuits. Moreover, we observe that such subspaces act as *attractors* for the variance, as each summand not only depends on the local components of ρ and H, but also on the components of H belonging to $\mathcal{E}(\mathcal{B}_z) \cap \mathcal{B}_Q$. In this sense, A can be interpreted as an absorption matrix, which quantifies the extent of the contribution of such terms. We remark that this contribution is always non-negative, and it is intimately related to the contractivity properties of $\mathcal{E}^\dagger$. We provide estimates of the absorption terms in the setting of Corollary IV.1 in Appendix C. Overall, this suggests that appropriate nonunitary layers will alleviate the concentration typical of unitary circuits by a mechanism that allows to bring the contribution of the components of H belonging to strictly contractive, high dimensional subspaces to noncontractive, smaller dimensional ones.

As shown in Appendix A, Eq. (15) can be used to recover deep circuit limit variance in the unitary and strictly contractive setting, consistent with existing literature as summarized in Table I. In terms of the LTM structure, these results can be understood as two extreme cases. In the unitary case, all irreducible blocks have unit spectral radius, so that T contains no Q or R blocks, while in the strictly contractive case, all irreducible blocks different than T_0 have spectral radius strictly smaller than unity. Beyond extremes, the effect of nonunitary intermediate channels on the LTM can be viewed as twofold. On the one hand, contractivity can reduce the spectral radius of (some) irreducible blocks similarly to the latter case, thereby decreasing the number of terms in the upper row of Eq. (11) and reducing the limiting variance. On the other hand, noise allows for the appearance of Q and R blocks,

TABLE I. Comparison between literature results on $\mathbb{V}^{\infty}_{\rho,H}$ and this work.

	Properties of Φ_θ				
	Unitary	Single-qubit unital noise	Single-qubit nonunital noise	Strictly contractive noise	Generic noise
Previous works	[13,16–18]	[24,25,34]	[25,32–34]	[25,34]	/
This work	Corollary A1	Corollary A2	Propositions C1 and C2	Corollary A2	Theorem IV.1

which can contribute positively through the absorption terms A_z, a mechanism specific to nonreversible noisy dynamics. The balance between these competing effects for a given noisy channel $\mathcal{E}$ governs the emergence, or absence, of noise-induced concentration, as captured by the asymptotic result of Theorem IV.1 and derived in Appendix E.

Regarding the specific class of nonunital noise, in Appendix C we show how previous results can be expressed as the contribution to the absorption term of $\mathcal{B}_0$, which always forms a norm-preserving subspace due to trace preservation of Φ_θ. However, it is crucial to remark that this specific contribution is not due to the retention of any computational power to the PQC, since the dependence on the initial state is completely lost, but rather to the competing effects between the drive of $\mathcal{U}_\theta$ and $\mathcal{E}$ toward the respective, different fixed points. An example of this phenomenon is provided in Appendix H, where we study how extensive entanglement can further exacerbate the issue. This is in stark contrast with the general case discussed above, i.e., the absorption terms of $\mathcal{B}_z$, $z > 0$, as they keep a nontrivial dependence on the initial state, and as such it can be said to genuinely avoid concentration if d_z scales appropriately. Indeed, this regime can be realized within the single-qubit noise model of Corollary IV.1 by employing a structured entangling gate and a mixture of clean and noisy qubits.

As an illustrative example, consider the 4-qubit circuit with intermediate channel $\mathcal{E} = \mathcal{N} \circ \mathcal{W}$ shown in Fig. 3. As discussed in Appendix C, this setup features two essential blocks, T_0 and T_1, both with unit spectral radius. For a suitable choice of H, Eq. (10) reduces to

$$\mathbb{V}^{L}_{\rho,H} = (\ell_\rho, A_1^{(L)} \ell_H) + (\ell_\rho, Q^L \ell_H), \tag{16}$$

where the first term is associated with absorption into $\mathcal{B}_1$, as characterized by Theorem IV.1, while the second term decays exponentially with L. Notably, both contributions depend nontrivially on ρ, and each can therefore be made to vanish by an appropriate choice of the initial state alone. As shown in Fig. 3, this leads to qualitatively different variance scalings depending on whether the first term is suppressed or not. This behavior already goes beyond the current description in the literature, allowing us to observe and understand new absorption-like dynamics. Further details about this example are provided in Appendix C.

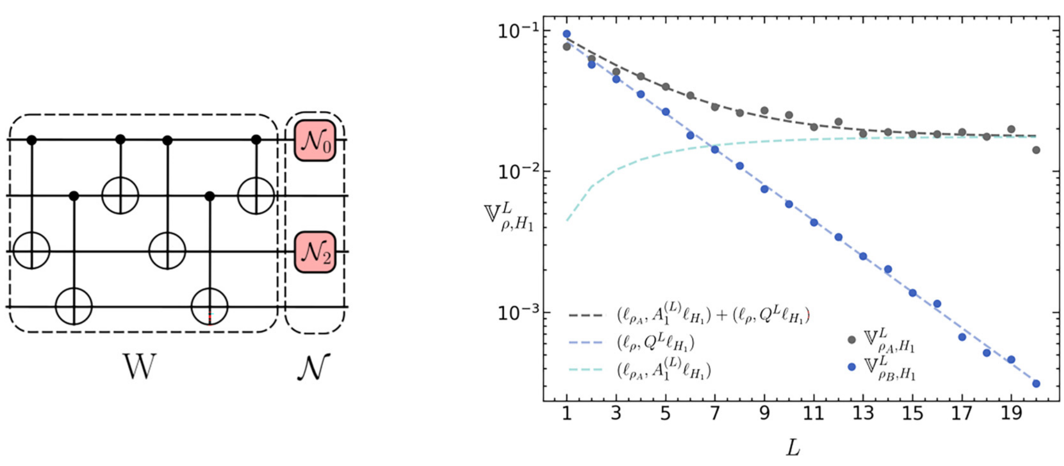

FIG. 3. Example of the absorption effect in a system comprising a mixture of clean and noisy qubits. On the left, we depict the structure of the intermediate channel used, composed of CNOT gates, a depolarizing channel $\mathcal{N}_0$, and an amplitude damping channel $\mathcal{N}_2$. On the right, we show the scaling of the corresponding variance $\mathbb{V}^L_{\rho,H}$ as a function of L, for different initial states and $H = H_1 = \mathbb{1}XX\mathbb{1}$. Dots represent numerically estimated variances from 200 samples using either $\rho_A = |0000\rangle\langle 0000|$ (gray) or $\rho_B \propto |0000\rangle\langle 0000| + |0110\rangle\langle 0110|$ (blue) as the initial state. Dashed lines indicate theoretical predictions for each contribution to the variance. By virtue of the initial states choice, $(\ell_{\rho_A}, Q^L\ell_{H_1}) = (\ell_{\rho_B}, Q^L\ell_{H_1})$, the subscript of ρ is dropped in the legend without ambiguities. Further details are provided in Appendix C.

The results showed so far were based on the analysis of the dominant eigenvectors of T. When L is not deep enough to reach convergence to its asymptotic limit, we need to characterize better the behavior of $\mathbb{V}^L_{\rho,H}$. This is the subject of the following section.

C. Lower bounds on "shallow" circuits

While for shallow circuits we cannot rely on the spectral properties of T to determine $\mathbb{V}^L_{\rho,H}$, we can still use knowledge of the convergence speed to $\mathbb{V}^\infty_{\rho,H}$ to set general lower bounds. Intuitively, such bounds can be obtained by preventing the variance to reach its stationary state, which can be done only if the exponential upper bounds appearing in Theorem IV.1 are sufficiently loose, namely, $\beta L \in O(\log n)$. This implies either that the circuit is *shallow*, i.e., there are not enough layers to reach the asymptotic value $\mathbb{V}^\infty_{\rho,H}$, or *effectively shallow*, i.e., the mixing speed β of T is slowed down according to $\beta \in O(\log n/L)$ so that $\mathbb{V}^\infty_{\rho,H}$ is never reached, regardless of L. As this speed is related to the amount of correlations with respect to the partition introduced by the channel, we study this scenario in the limit of $\mathcal{E}^\dagger$ being close to locality preserving. This idea is formally captured in the following theorem, which extends its application to generally nonhomogeneous layered, locally random channels.

Theorem IV.2 (General lower bound). Let $\rho, H \in \mathcal{B}$ and consider a sequence of quantum channels $\{\mathcal{E}_l\}_{l=1}^L$, and let $\{T_l\}_{l=1}^L$ be the respective LTMs. Finally let $K \subset \{0,1\}^M$ denote a subset of indices and by $\alpha_l = \min_{\kappa\in K}(T_l)_{\kappa,\kappa}$. Then

$$\mathbb{V}^L_{\rho,H} \geq \alpha^L(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{17}$$

where $\mathcal{K} : \mathcal{B} \to \mathcal{B}_K$ is the projector onto $\mathcal{B}_K = \bigoplus_{\kappa\in K} \mathcal{B}_\kappa$ and α is the geometric mean of α_l.

Depending on the scaling of α and the dimensions d_m of the subsystems, Eq. (17) can provide a meaningful lower bound. For instance, focusing on the case of subsystems with constant dimension, we have the following corollary.

Corollary IV.2 (Lower bound examples). Let $\mathcal{H} = \bigotimes_{m=1}^M \mathcal{H}_m$, $d_m \in \Theta(1)$. If either of the conditions

(a) $\alpha > 0, \alpha \in \Omega(1)$ and $L \in O(\log n)$,
(b) $\alpha = 1 - f(n, L)$, $f \in O(\log n/L)$ and $L \in \Omega(\log^{1+\epsilon} n)$ for some arbitrary $\epsilon > 0$

is satisfied, then

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{18}$$

where $F(n) \in \Omega(1/\mathrm{poly}(n))$.

The conditions of Corollary IV.2 reflect the aforementioned scenarios; in particular, condition (a) ensures the absence of concentration for *shallow* circuits, both unitary and noisy. Specifically, this holds true whenever $0 < \alpha \in \Omega(1)$, indicating that the intermediate channel does not become increasingly rapidly entangling, as the problem size grows. As a notable example, this condition is satisfied for brickwork circuits, equipped with local noise and a local observable [32]. Similarly, condition (b) reflects the absence of concentration for *effectively shallow* quantum channels. A significant example is that of deep finite local-depth circuits (FLDCs) [36]. Examples of both are provided in Appendix F, while a comparison with existing literature is provided in Table II. We can interpret condition (b) as a limit to the mixing speed of T by noting that, in the homogeneous case, it is equivalent to the more explicit relation $|1 - \lambda| \in O(\log n/L)$ for all the eigenvalues λ of T by Gershgorin circles theorem [37], which directly implies $\beta \in O(\log n/L)$.

Interestingly, since these results have been obtained by imposing $T \approx \mathbb{1}$, they have a strong resemblance with small-angle initialization strategies [21,22], which similarly hinge on identity manipulation. In fact, while the primary concern of Theorem IV.2 is noise, it can be seen as a theoretical foundation of such initialization strategies. In the following section we discuss in detail the strong relation between smart initializations for noiseless circuits and the properties of noisy layers. In doing so, we analytically estimate the scaling of α for a broad class of circuits, hence providing rigorous lower bounds on $\mathbb{V}^L_{\rho,H}$.

D. Connection with small-angle initializations

Here we expand on the concept of small-angle initialization introduced in Refs. [21,22]. In particular, we establish a general relationship between the insights gained from controlling loss concentration in noisy circuits (as presented in Theorem IV.2) and BP mitigation strategies that typically only apply to ideal, unitary circuits.

The main idea behind small-angle initialization strategies considers a layered quantum circuit $U_\theta = \prod_l U_{\theta_l}$

TABLE II. Comparison between literature results on polynomial lower-bounds on $\mathbb{V}^L_{\rho,H}$ and this work.

	Properties of Φ_θ			
	Shallow brickwork	Finite local depth	Small angles	(Nonunitary) QResNet
Previous works	[13,32]	[36]	[21,22]	/
This work	Corollary IV.2(a)	Corollary IV.2(b)	Proposition IV.2	Propositions IV.2 and IV.3

where the absence of concentration for a given initialization distribution is shown. This is typically very peaked around 0, with variance scaling inversely to the number of layers: $\sigma^2 \in O(1/L)$. The core idea of all such strategies relies on *identity manipulation*, i.e., on choosing initialization distributions such that $U_\theta \approx \mathbb{1}$ with high probability. This introduces a sizable variance to the circuit, at the price of having a large bias toward the identity in the quantum model.

A similar structure can be defined in our framework by considering quantum channels $\Phi_{\theta,\phi}$, as in Definition III.2, where the intermediate channels $\mathcal{E}_{\phi_l}$ are now parameterized. This differs from the main idea of small-angle initialization, as the local components $\mathcal{U}_{\theta_l}$ of the channel remain Haar random, and instead it is the allowed channels $\mathcal{E}_\phi$ that get restricted. Intuitively, this will lead to a different model bias for small angles, i.e., $\Phi_{\theta,\phi} \approx \mathcal{U}_\theta$. We name this model Quantum Residual Network (QResNet), as we interpret the large identity component of $\mathcal{E}_\phi$ as a *skip-connection*, in analogy to classical Residual Networks [38]. Indeed, this structure is enough to avoid concentration when a small-angle initialization strategy is used. To see this, we remark that while a constant channel was used to derive Proposition IV.1, it can be readily generalized to parameterized channels $\mathcal{E}_\phi$, as long as the parameters ϕ_l are independent. In that case it is sufficient to use $\mathbb{E}_\phi\{T_\phi\}$ in place of T, where T_ϕ is the LTM of $\mathcal{E}_\phi^\dagger$ (see Appendix F for a proof). Exploiting this, we can derive the following proposition.

Proposition IV.2 (QResNet). Let $\mathcal{E}_\phi(\cdot) = e^{i\phi G} \cdot e^{-i\phi G}$ be a unitary entangling gate with normalized generator, i.e., $\|G\|_2^2/d = 1$. Furthermore, let μ, σ^2 be the mean and variance of the initialization distribution of ϕ. Then, if $\mu = 0$, $\sigma^2 \in O(\log n/L)$, and $L \in \Omega(\log^{1+\epsilon} n)$ we have

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_H), \tag{19}$$

where $F(n) \in \Omega(1/\mathrm{poly}(n))$.

Eq. (19) represents an application of Corollary IV.2(b) in the case of unitary, parameterized intermediate channels. Note that Theorem IV.2, which is the backbone of Proposition IV.2, is not limited to unitary circuits, but is applicable to generic quantum channels. Indeed, if we take $\mathcal{E}$ to be a noise model, Corollary IV.2 (b) may be analogously interpreted as a condition on the noise rates to avoid concentration. This showcases the connection between BP-free QResNets and noise models escaping NIBP, which can be interpreted as follows: if a noise map satisfies Corollary IV.2, then there exists a QResNet associated with it that is able to avoid BP. As an example, assume that the channel $\mathcal{E} = e^{\Delta t\mathcal{L}}$ is obtained as a solution of the Lindblad equation at time $\Delta t \ll 1$, where $\mathcal{L}(\rho) = \sum_i L_i\rho L_i - 1/2\{L_i^2, \rho\}$, $L_i = L_i^\dagger\ \forall i$, i.e., we consider weak Lindbladian noise [39]. In this case, there always exists a unitary, stochastic unraveling $\{U_\phi\}_\phi$, such that $\mathcal{E}(\rho) = \mathbb{E}_\phi\{U_\phi\rho U_\phi^\dagger\}$ [40]. Then, the ensemble corresponding to the choice $\mathcal{E}_\phi(\rho) = U_\phi\rho U_\phi^\dagger$ will avoid BP if Δt is small enough. Formally, we have the following result.

Proposition IV.3 (General QResNets from a quantum channel). Let $\mathcal{E}(\rho) = e^{t\mathcal{L}}(\rho)$ be a quantum channel generated by the Lindbladian

$$\mathcal{L}(\rho) = \sum_k \gamma_k \left(L_k\rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\} \right), \tag{20}$$

with $\|L_k\|_2^2/d = 1$. Then, when $L \in \Omega(\log^{1+\epsilon} n)$, the ensemble generated by $\mathcal{E}_\phi = E_\phi \cdot E_\phi^\dagger$,

$$E_\phi = \prod_k e^{i\phi_k L_k + \sigma_k^2(L_k^2 - L_k^\dagger L_k)/2} \tag{21}$$

and ϕ_k normally distributed with $\mu = 0$, $\sigma_k^2 \in O\left((\gamma_k/\sum_{k'}\gamma_{k'})(\log n/L)\right)\ \forall k$ satisfies

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_H), \tag{22}$$

where $F(n) \in \Omega(1/\mathrm{poly}(n))$.

If $L_k = L_k^\dagger$, we get a unitary QResNet as in the previous discussion. Interestingly, Proposition IV.3 can be equally applied if the ensemble is not unitary, which can happen if the noise channel $\mathcal{E}$ is nonunital (e.g., $L_k \neq L_k^\dagger$). This effectively extends the framework of small-angle initialization to nonunitary quantum models, e.g., those based on linear combination of unitaries (LCUs) or analogous techniques [41]. Finally, it is important to note that, while a *specific* unraveling was used to derive Proposition IV.3, the stochastic unraveling of a given channel is not unique, allowing multiple QResNets (unitary or nonunitary) to be derived from the same channel. Additionally, different choices among such models may lead to distinct variances [42], all bounded from below by the variance of $\mathcal{E}$. Indeed, any stochastic unraveling is connected to the physical channel solely through their first moment, while higher-order moments are relevant only in the context of the associated QResNet, as they do not represent otherwise measurable quantities.

V. DISCUSSION

The study of loss function concentration is a central topic in variational quantum computing. While the description of this effect in the absence of noise has been recently formulated using Lie-algebraic theory [16–18], this approach inevitably fails in the general setting of nonunitary circuits, where the group structure description is lost. In this work, we employ non-negative matrix theory to derive a general expression for $\mathbb{V}^L_{\rho,H}$ in quantum circuits composed of local 2-designs interleaved with arbitrary quantum channels. This circuit architecture is chosen for its combination of analytical accessibility and physical relevance: it is rich enough to capture qualitatively novel variance behavior, while structured enough to address fundamental questions about the interplay between noise and unitary layers in the emergence of loss concentration.

In particular, within this framework, the variance is formulated in terms of the channel's LTM, which in turn enables the precise computation of $\mathbb{V}^\infty_{\rho,H}$ in the deep-circuit limit. Specifically, the observed structure of $\mathbb{V}^\infty_{\rho,H}$ brings out a new mechanism, which we call *absorption*, whereby components of H pertaining to strictly contractive subspaces of $\mathcal{B}_Q$ can augment the variance of the model by coupling with noncontractive ones through the nonreversible action of the quantum noise channel. Such result is supported by numerical simulations, provided in Appendixes C, H, and G.

This indicates that a mixed configuration of ideal and noisy qubits, as well as a partial realization of error correction, could potentially outperform purely ideal or purely noisy systems in terms of variance and highlights novel properties of $\mathbb{V}^\infty_{\rho,H}$ in this setting, e.g., the initial state dependence of noise-induced concentration, which is crucial for algorithm engineering. This is especially important in the early stages of fault-tolerant quantum computing, where a large quantity of noisy qubits can be utilized, but the availability of ideal qubits is still constrained [43,44]. This may also be beneficial in characterizing variational quantum algorithms executed on hardware, where different qubits may experience varying error rates [39]. In this context, $\mathbb{V}^\infty_{\rho,H}$ can be used to approximate $\mathbb{V}^L_{\rho,H}$ in regimes where L is sufficiently large to significantly impact some qubits but not yet others.

Furthermore, this approach clarifies the role of unitality in NIBPs, shifting the focus on the more relevant aspect of *contractivity* and *alignment* between invariant subspaces of the unitary layers $\mathcal{U}_\theta$ and strictly contractive subspaces of the noise map $\mathcal{E}$. A comparison with existing literature on estimates of $\mathbb{V}^\infty_{\rho,H}$ is provided in Table I.

Additionally, we introduced a general lower bound on the variance of noisy circuits. This bound is derived by restricting the mixing speed of $\mathbb{V}^L_{\rho,H}$ to prevent it from reaching its asymptotic limit. We subsequently employ this approach to introduce QResNets, an initialization strategy analogous to small-angle initializations [19–22] that can effectively mitigate the occurrence of barren plateaus. Moreover, we demonstrate that an analogous procedure can be applied to noise maps, establishing a formal connection between weak noise and QResNets. This enables us to derive BP-free QResNets as stochastic unravelings of sufficiently weak noise maps. Notably, since these unravelings are not necessarily unitary, this approach can yield BP-free architectures beyond unitary circuits, encompassing more complex models such as those employing LCUs [41]. A comparison with existing literature on $\mathbb{V}^L_{\rho,H}$ lower bounds is provided in Table II.

Future research may extend our findings by relaxing the local 2-design property of the unitary layer. Such a generalization would broaden the domain of applicability of our results beyond hardware-specific designs, but it would primarily offer a technical refinement without altering the conceptual insights gained in this work. Furthermore, recent studies have established a connection between the absence of concentration and classical simulability for both ideal [45] and noisy [32] quantum circuits. While these results often focus on strictly contractive noise models, our work suggests a potential avenue for combining these concepts, extending their validity to more complex noise environments. A representation of our contributions is depicted in Fig. 1: the non-negative matrix approach offers new and timely answers to several open questions related to concentration phenomena in quantum circuits and provides valuable insights into the optimal utilization of near-term and early fault-tolerant quantum devices, thus guiding the community toward the effective application of the variational quantum computation framework.

ACKNOWLEDGMENTS

G.C. thanks F. Benatti, G. Nichele, M. Vischi, G. Di Bartolomeo, A. Candido, S. Y. Chang, M. Sahebi, and M. Robbiati for useful discussions, and the CERN Openlab agreement with the University of Trieste. G.C. acknowledges financial support from University of Trieste, INFN and EU Erasmus+ Traineeship Programme. M.G. is supported by CERN through the CERN Quantum Technology Initiative. A.B. acknowledges support from the University of Trieste, INFN, the PNRR MUR project PE0000023 - NQSTI and the EU EIC Pathfinder project QuCoM (GA 101046973).

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

APPENDIX A: LIMITING CASES OF THE DEEP CIRCUIT FORMULA

Starting from the analysis of Sec. IV, we show how we can use Eq. (15) to recover previously known results by considering specific classes of quantum channels $\mathcal{E}$. Formal proofs of these statements can be found in Appendix E.

First we point out that, being an absorption term, the last term in Eq. (15) vanishes for unitary dynamics, which is reversible by definition. Formally, we have the following corollary of Theorem IV.1.

Corollary A1 (Deep, unitary circuits). Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition III.2, where $\mathcal{E}(\cdot) = W \cdot W^\dagger$, $W \in U(d)$ is an arbitrary unitary transformation. Then we have

$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z>0} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z}, \tag{A1}$$

where $\mathcal{B}_z$ are invariant subspaces of $\mathcal{E}$ which can be expressed as the direct sum of $\mathcal{B}_\kappa$.

This corollary contains many interesting properties of the variance in the deep circuit limit. First, it captures the necessity of *alignment* between ρ, H, and Φ_θ in order to achieve a substantial variance, i.e., both ρ and H need to have non-negligible components on the same invariant subspace $\mathcal{B}_z$. Due to the structure of the channel, this idea is extended to the components $\mathcal{U}_\theta$ and $\mathcal{E}$, whose invariant subspaces need to align in order to keep the dimension d_z of $\mathcal{B}_z$ from being exponentially large. Indeed, while $\mathcal{B}_0$ is always an invariant subspace satisfying Corollary A1, misalignment of the local and entangling parts of the circuit could result in the entirety of the remaining space falling under a single, irreducible component of dimension $d^2 - 1$. In such cases we get

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{\left(\|\rho\|_2^2 - 1/d\right)\left(\|H\|_2^2 - \mathrm{Tr}\left[H\right]^2/d\right)}{d^2 - 1}, \tag{A2}$$

which implies the presence of BP regardless of ρ and H as long as $\|H\|_2 < O(d)$. Indeed, one can interpret the misalignment of $\mathcal{U}_\theta$ and $\mathcal{E}$ as introducing an excess of expressibility, which is known to lead to exponential concentration [10].

As a complementary remark, we point out that, conversely to the above, the first term in Eq. (15) pertains to noncontractive subspaces, and as such, vanishes if $\mathcal{E}^\dagger$ is strictly contractive in, at least, one direction in each $\mathcal{B}_z$. This is formalized in the following corollary of Theorem IV.1.

Corollary A2 (Deep, noisy circuits). Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition III.2, where $\mathcal{E}$ is such that $\|\mathcal{E}(A)\|_2 < \|A\|_2$, for at least one $A \in \mathcal{B}_\kappa \subset \mathcal{B}_z \, \forall z > 0$. Then, we have

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{(A\ell_H)_0}{d}. \tag{A3}$$

In particular, if the channel is unital, $\mathbb{V}^{\infty}_{\rho,H} = 0$.

Note that, even if Eq. (A3) is inversely proportional to $d = 2^n$, $\mathbb{V}^{\infty}_{\rho,H}$ is not necessarily exponentially suppressed, as in general the contribution of the observable increases with the same speed, i.e., $\|H\|_2^2 \sim d$. As before, this Corollary captures the main features of NIBP. In fact, it is clear that strictly contractive channels with a unique fixed point fall into the assumptions of Corollary A2 and therefore exhibit some form of concentration. However, Corollary A2 is not limited to them, extending the noise-induced concentration to a wider class of noise maps, which crucially depend on the structure of the unitary part of the channel $\mathcal{U}_\theta$. According to the method introduced here, this can clearly be interpreted as a consequence of the interaction between noise and unitary layers, which effectively "spread" the contractive effect of $\mathcal{E}$ to the whole irreducible component. Unital channels will suffer most severely from NIBP, since in that case the absorption term in Eq. (A3) vanishes. Contrarily, as pointed out in the literature [32,34], nonunital channels may avoid the exponential concentration. Here, we have shown how these results obtained in the literature can be seen as the contribution to the absorption term of $\mathcal{B}_0$, which cannot be strictly contractive due to trace preservation of Φ_θ.

APPENDIX B: LOCALITY AND LOCALITY TRANSFER MATRIX PROPERTIES

In order to show the main properties of locality vectors and LTMs, it is convenient, for each subsystem $\mathcal{H}_m$, to fix an orthonormal basis $\{P_j^{(m)}\}_{j=1}^{d_m-1}$ of $\mathcal{B}_m$ composed of traceless, Hermitian operators together with the identity operator, each normalized with respect to the Hilbert-Schmidt norm, namely, $P_0^{(m)} = \mathbb{1}/\sqrt{d_m}$, $P_{j_m}^{(m)} = P_{j_m}^{(m)\dagger} \, \forall j_m$,$\mathrm{Tr}\left[P_{j_m}^{(m)} P_{k_m}^{(m)}\right] = \delta_{j_m k_m}$. Starting from these, we can build an orthonormal basis $\{P_j\}_j$ for the whole space by means of tensor products. Each basis element will be labeled by the multi-index $j = (j_1, \ldots, j_M)$, where the entries j_m refer to an element of the local bases, and hence $j_m \in \{0, \ldots, d_m^2 - 1\}$. Such a basis will be dubbed a *local* basis. As an example, if each $\mathcal{H}_m = \mathbb{C}^2$ is a qubit, the normalized Pauli strings form a local basis for $\mathcal{B}$. Given a local basis $\{P_j\}_j$, it is possible to group the elements in disjoint sets. In particular, for any given binary string $\kappa = \{0,1\}^M$, we can collect in the set S_κ all basis elements acting nontrivially on $\mathcal{H}_m$ if and only if $\kappa_m = 1$. For practical reasons, we introduce the indicator function $\tilde{\delta}_{i,\kappa}$ for the set S_κ, defined by

$$\tilde{\delta}_{i,\kappa} = \prod_{m=0}^{M} \tilde{\delta}_{i_m,\kappa_m} = \begin{cases} 1 & \text{if} \, \kappa_m = 0 \Leftrightarrow i_m = 0 \, \forall m, \\ 0 & \text{otherwise.} \end{cases} \tag{B1}$$

From the definition, we can derive some simple properties of this function.

Lemma B1. The indicator function $\tilde{\delta}_{i,b}$ has the following properties:

$$\sum_i \tilde{\delta}_{i,\kappa} = d_\kappa, \quad \sum_\kappa \tilde{\delta}_{i,\kappa} = 1, \quad \sum_{\kappa\in K} \tilde{\delta}_{i,\kappa}\tilde{\delta}_{i,\lambda} = \tilde{\delta}_{i,\lambda}\sum_{\kappa\in K}\delta_{\kappa,\lambda}, \tag{B2}$$

where $d_\kappa = \prod_{m=0}^{M}(d_m^2-1)^{\kappa_m}$, $K \subset \{0,1\}^M$, and $\delta_{\kappa,\lambda}$ is the usual Kronecker delta.

Proof. All the results follow directly from the definition. ■

Using this notation, we can express the *locality* ℓ_A of some operator $A \in \mathcal{B}$ and *locality transfer matrix* T of a linear map $\Lambda : \mathcal{B} \to \mathcal{B}$ defined as

$$(\ell_A)_\kappa = \sum_j \mathrm{Tr}\left[P_j A\right]^2 \tilde{\delta}_{j,\kappa} \tag{B3}$$

and

$$T_{\kappa,\lambda} = \frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\left[P_i\Lambda(P_j)\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda}, \tag{B4}$$

respectively. It is immediately realized that both these quantities are basis-independent.

Lemma B2. Given a bounded operator $A \in \mathcal{B}$ and a partition into subsystems, the locality vector ℓ_A is uniquely defined, i.e., it does not depend on the choice of local basis. Similarly, given a quantum channel $\mathcal{E} : \mathcal{B} \to \mathcal{B}$, the corresponding locality transfer matrix T is uniquely defined.

Proof. Let $\{P_j\}$ and $\{B_i\}$ be two local bases for a given subsystem. Then

$$\begin{aligned}
&\sum_j \mathrm{Tr}\left[P_j A\right]^2\tilde{\delta}_{j,\kappa} = \sum_j \mathrm{Tr}\left[\sum_i \mathrm{Tr}\left[B_i P_j\right] B_i A\right]^2 \tilde{\delta}_{j,\kappa}\\
&= \sum_j\sum_{i,i'}\mathrm{Tr}\left[B_iP_j\right]\mathrm{Tr}\left[B_{i'}P_j\right]\mathrm{Tr}\left[B_iA\right]\mathrm{Tr}\left[B_{i'}A\right]\tilde{\delta}_{j,\kappa}\\
&= \sum_{i,i'}\left(\sum_j \mathrm{Tr}\left[B_iP_j\right]\mathrm{Tr}\left[B_{i'}P_j\right]\tilde{\delta}_{j,\kappa}\right)\mathrm{Tr}\left[B_iA\right]\mathrm{Tr}\left[B_{i'}A\right]\\
&= \sum_i \mathrm{Tr}\left[B_iA\right]^2\tilde{\delta}_{i,\kappa},
\end{aligned} \tag{B5}$$

where the last equality is due to

$$\begin{aligned}
&\sum_j \mathrm{Tr}\left[B_iP_j\right]\mathrm{Tr}\left[B_{i'}P_j\right]\tilde{\delta}_{j,\kappa} =\\
&= \prod_{m=1}^{M}\left(\sum_{j_m=1}^{d_m}\mathrm{Tr}\left[B_{i'_m}P_{j_m}\right]\mathrm{Tr}\left[B_{i_m}P_{j_m}\right]\tilde{\delta}_{j_m,\kappa_m}\right)\\
&= \prod_{m=1}^{M}\left(\mathrm{Tr}\left[B_{i'_m}B_{i_m}\right]\tilde{\delta}_{i_m,\kappa_m}\right) = \delta_{i,i'}\tilde{\delta}_{i,\kappa},
\end{aligned} \tag{B6}$$

which holds since both $\{P_{j_m}\}$ and $\{B_{i_m}\}$ are orthonormal bases of $\mathcal{B}_m$ by definition of local basis. A totally analogous calculation yields the same result for the locality transfer matrix T. ■

Thanks to the formulation of Eq. (B4), the relation between the LTM of a map Λ and the Hermitian adjoint $\Lambda^\dagger$ with respect to the Hilbert-Schmidt scalar product can be seen. In particular, we have

$$\begin{aligned}
d_\lambda T_{\kappa,\lambda} &= \sum_{i,j}\mathrm{Tr}\left[P_i\Lambda(P_j)\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda}\\
&= \sum_{i,j}\mathrm{Tr}\left[\Lambda^\dagger(P_i)P_j\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda} = d_\kappa T^\dagger_{\lambda,\kappa},
\end{aligned} \tag{B7}$$

which can be compactly written in matrix form as $TD = (T^\dagger D)^t$, with $D_{\kappa,\lambda} = d_\kappa\delta_{\kappa,\lambda}$. For the sake of readability, here we introduce a shorthand notation for the scalar product $(\cdot,\cdot)$ in $\mathbb{R}^{2^M}$ such that T and $T^\dagger$ are *also* Hermitian adjoint of one another, i.e.,

$$(a,b) = a^tD^{-1}b = \sum_\kappa \frac{a_\kappa b_\kappa}{d_\kappa}. \tag{B8}$$

This trivially follows from the chain $(a,Tb) = a^tD^{-1}Tb = a^tD^{-1}D(T^\dagger)^tD^{-1}b = (T^\dagger a)^tD^{-1}b = (T^\dagger a, b)$.

APPENDIX C: ESTIMATES OF THE ABSORPTION COEFFICIENTS

In this appendix, we estimate the absorption coefficients in the general setting of Corollary IV.1, where the entangling channel admits a decomposition of the form

$$\mathcal{E}(\rho) = \mathcal{N}(W\rho W^\dagger),\ \mathcal{N} = \bigotimes_m \mathcal{N}_m, \tag{C1}$$

that is, a generic unitary entangling gate followed by local single-qubit noise. In the first part, we provide theoretical lower bounds for all absorption coefficients, along with scaling guarantees for the absorption into the trivial subspace, thereby demonstrating consistency with prior literature [32,34]. In the second part, we present numerical evidence of absorption into nontrivial subspaces, revealing qualitatively distinct variance behaviors depending on the choice of initial state. This effect goes beyond the scope of existing analyses and suggests the emergence of a previously uncharacterized phenomenon.

1. Theoretical estimates

We first prove a general lower bound for the absorption coefficients in the setting of Eq. (C4).

Proposition C1. Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition III.2, and let the intermediate channel take the form of Eq. (C4). Then, for all norm-preserving, strongly connected components T_z of the LTM T, we have

$$\frac{(A\ell_H)_z}{d} \geq \sum_{j \mid P_j \in \mathcal{B}_z} (\Lambda_j, \ell_H), \tag{C2}$$

where the operators Λ_j are defined by the action of the noise channels as

$$\Lambda_j = \mathcal{Q}(\tilde{\Lambda}_j),\ \tilde{\Lambda}_j = \bigotimes_m \mathcal{N}_m\left(\frac{P_{jm}}{\sqrt{2}}\right), \tag{C3}$$

with P_j are normalized Pauli strings and $\mathcal{Q}: \mathcal{B} \to \mathcal{B}_{\mathcal{Q}}$ is a projection onto $\mathcal{B}_{\mathcal{Q}}$.

Note that the factor $1/\sqrt{2}$ is introduced in the definition of $\tilde{\Lambda}_j$ to ensure $\mathrm{Tr}\left[\tilde{\Lambda}_j\right] \leq 1$, following the normalization constraints for quantum states. In this way, if we restrict to the trivial subspace $\mathcal{B}_0$, $\tilde{\Lambda}_j = \tau$ becomes a quantum state and, in particular, it becomes the action of the noise channel on the maximally mixed state $\tau = \mathcal{N}(\rho_m)$, $\rho_m = \mathbb{1}/d$.

Proof. By Theorem IV.1, we can express the absorption matrix explicitly in terms of the relevant portions of the LTM. In particular, we have

$$(A\ell_H)_z = \sum_{\zeta,\kappa,\lambda} R_{\zeta,\kappa}(\mathbb{1} - \mathcal{Q})^{-1}_{\kappa,\lambda}(\ell_H)_\lambda \delta_{\zeta,z}\delta_{\kappa,\mathcal{Q}}, \tag{C4}$$

where we used the shortcut notation $\delta_{\zeta,z} = \sum_{\nu\in T_z} \delta_{\zeta,\nu}$ and $\delta_{\kappa,\mathcal{Q}} = \sum_{\nu\in\mathcal{Q}} \delta_{\kappa,\nu}$. Such terms are necessary to avoid counting portions of T_z and $\mathcal{Q}$ into the block R.

As a first step, we compute the term $\sum_\zeta R_{\zeta,\kappa}\delta_{\zeta,z}$ in this noise model. For the sake of readability, let us further denote by $\tilde{\delta}_{i,z}$ the quantity $\sum_\zeta \tilde{\delta}_{i,\zeta}\delta_{\zeta,z}$,

$$\begin{aligned}
\sum_\zeta R_{\zeta,\kappa}\delta_{\zeta,z} &= \frac{1}{d_\kappa}\sum_{i,j} \mathrm{Tr}\left[P_i W^\dagger \mathcal{N}^\dagger(P_j) W\right]^2 \tilde{\delta}_{i,z}\tilde{\delta}_{j,\kappa}\delta_{\kappa,\mathcal{Q}} \\
&= \frac{1}{d_\kappa}\sum_{i,j} \mathrm{Tr}\left[\sum_\mu w_{i,\mu} P_\mu \mathcal{N}^\dagger(P_j)\right]^2 \tilde{\delta}_{i,z}\tilde{\delta}_{j,\kappa}\delta_{\kappa,\mathcal{Q}} \\
&= \frac{1}{d_\kappa}\sum_{i,j} \mathrm{Tr}\left[\sum_\mu w_{i,\mu} \mathcal{N}(P_\mu) P_j\right]^2 \tilde{\delta}_{i,z}\tilde{\delta}_{j,\kappa}\delta_{\kappa,\mathcal{Q}} \\
&= \frac{1}{d_\kappa}\sum_j\sum_{\mu,\nu}\left(\sum_i w_{i,\mu}w_{i,\nu}\tilde{\delta}_{i,z}\right)\mathrm{Tr}\left[\mathcal{N}(P_\mu)P_j\right] \\
&\quad \times \mathrm{Tr}\left[\mathcal{N}(P_\nu)P_j\right]\tilde{\delta}_{j,\kappa}\delta_{\kappa,\mathcal{Q}},
\end{aligned} \tag{C5}$$

where $w_{i,\mu} = \mathrm{Tr}\left[WP_iW^\dagger P_\mu\right]$. Recall that, by definition of LTM, if $P_i \in \mathcal{B}_z$, then $WP_iW^\dagger \in \mathcal{B}_z$ as well. For this reason, $w_{i,\mu} = 0\ \forall\mu$ s.t.$P_\mu \notin \mathcal{B}_z$, i.e., $w_{i,\mu} = w_{i,\mu}\tilde{\delta}_{\mu,z}\tilde{\delta}_{i,z}$. Furthermore, we have that

$$\begin{aligned}
&\sum_i w_{i,\mu}w_{i,\nu}\tilde{\delta}_{i,z}\tilde{\delta}_{\mu,z}\tilde{\delta}_{\nu,z} \\
&= \sum_i w_{i,\mu}w_{i,\nu} \\
&= \sum_i \mathrm{Tr}\left[WP_iW^\dagger P_\mu\right]\mathrm{Tr}\left[WP_iW^\dagger P_\nu\right] \\
&= \mathrm{Tr}\left[\left(\sum_i \mathrm{Tr}\left[P_iW^\dagger P_\nu W\right]\right)P_iW^\dagger P_\mu W\right] \\
&= \mathrm{Tr}\left[W^\dagger P_\nu WW^\dagger P_\mu W\right] \\
&= \mathrm{Tr}\left[P_\nu P_\mu\right] = \delta_{\mu,\nu} = \delta_{\mu,\nu}\tilde{\delta}_{\mu,z}\tilde{\delta}_{\nu,z}.
\end{aligned} \tag{C6}$$

Consequently, we have

$$\begin{aligned}
\sum_\zeta R_{\zeta,\kappa}\delta_{\zeta,z} &= \frac{1}{d_\kappa}\sum_j\sum_{\mu,\nu}\delta_{\mu,\nu}\tilde{\delta}_{\mu,z}\mathrm{Tr}\left[\mathcal{N}(P_\mu)P_j\right] \\
&\quad \mathrm{Tr}\left[\mathcal{N}(P_\nu)P_j\right]\tilde{\delta}_{i,\kappa}\delta_{\kappa,\mathcal{Q}} \\
&= \frac{1}{d_\kappa}\sum_{j,\mu}\mathrm{Tr}\left[\mathcal{N}(P_\mu)P_j\right]^2\tilde{\delta}_{j,\kappa}\tilde{\delta}_{\mu,z}\delta_{\kappa,\mathcal{Q}} \\
&= \frac{1}{d_\kappa}\sum_{j,\mu}\mathrm{Tr}\left[\sqrt{d}\tilde{\Lambda}_\mu P_j\right]^2\tilde{\delta}_{j,\kappa}\tilde{\delta}_{\mu,z}\delta_{\kappa,\mathcal{Q}} \\
&= d\sum_\mu \frac{(\ell_{\tilde{\Lambda}_\mu})_\kappa}{d_\kappa}\tilde{\delta}_{\mu,z}\delta_{\kappa,\mathcal{Q}} = d\sum_\mu \frac{(\ell_{\Lambda_\mu})_\kappa}{d_\kappa}\tilde{\delta}_{\mu,z}.
\end{aligned} \tag{C7}$$

Putting everything into Eq. (C4), we get

$$\begin{aligned}\frac{(A\ell_H)_z}{d} &= \frac{1}{d}\sum_{\zeta,\kappa,\lambda} R_{\zeta,\kappa}(\mathbb{1}-Q)^{-1}_{\kappa,\lambda}(\ell_H)_\lambda \delta_{\zeta,z}\\ &= \sum_\mu \sum_{\kappa,\lambda} \frac{(\ell_{\Lambda_\mu})_\kappa}{d_\kappa}(\mathbb{1}-Q)^{-1}_{\kappa,\lambda}(\ell_H)_\lambda \delta_{\zeta,z}\tilde{\delta}_{\mu,z}\\ &= \sum_\mu (\ell_{\Lambda_\mu},(\mathbb{1}-Q)^{-1}\ell_H)\tilde{\delta}_{\mu,z}\\ &= \sum_\mu (\ell_{\Lambda_\mu},\ell_H)\tilde{\delta}_{\mu,z}+\\ &\quad+\sum_\mu (\ell_{\Lambda_\mu},(\mathbb{1}-Q)^{-1}Q\ell_H)\tilde{\delta}_{\mu,z},\end{aligned} \tag{C8}$$

where we used the identity $(\mathbb{1}-Q)^{-1}=\mathbb{1}-(\mathbb{1}-Q)^{-1}Q$. Expanding $(\mathbb{1}-Q)^{-1}=\sum_{l=0}^{\infty}Q^l$, we get $(\mathbb{1}-Q)^{-1}Q=\sum_{l=1}^{\infty}Q^l$, which is entrywise non-negative. This implies that $(\ell_{\Lambda_\mu},(\mathbb{1}-Q)^{-1}Q\ell_H)\geq 0\,\forall\mu$, proving the proposition. ■

As a special case of Proposition C1, we have that

$$\frac{(A\ell_H)_0}{d}\geq(\ell_{Q(\tau)},\ell_H). \tag{C9}$$

Furthermore, if the map is unital, $\tau=\mathcal{N}(\mathbb{1}/d)=\mathbb{1}/d$ and $\ell_{Q(\tau)}=0$, so the bound becomes vacous in agreement with Corollary A2.

As a second theoretical point, we show how the general scaling for the variance obtained in the literature (e.g., Theorem 6 in Ref. [32]) can also be recovered in this formalism, by considering absorption to the trivial subspace only.

Proposition C2. If the noise channel in Eq. (C4) is further simplified to $\mathcal{N}=\bar{\mathcal{N}}^{\otimes n}$, where $\bar{\mathcal{N}}$ is a single-qubit, nonunital channel, then the trivial subspace $\mathcal{B}_0$ is the only norm-preserving, strictly connected component, its corresponding absorption coefficient scales as

$$\frac{(A\ell_H)_0}{d}\in\sum_i h_i^2 e^{\Theta(|P_i|)}, \tag{C10}$$

where $H=\sum_i h_iP_i$ denotes the Pauli decomposition of H and $|P_i|$ denotes the number of nontrivial terms in P_i.

Proof. First, we show that in this case, all subspaces $\mathcal{B}_\kappa$ pertain to strictly contractive components, except the trivial one. In particular, if $\sum_\kappa Q_{\kappa,\lambda}<1,\forall\lambda$, this is the case by the subinvariance theorem (Theorem E2),

$$\begin{aligned}\sum_\kappa Q_{\kappa,\lambda} &= \frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\left[P_iW^\dagger\mathcal{N}^\dagger(P_j)W\right]^2\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\left[B_i\mathcal{N}^\dagger(P_j)\right]^2\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_{j}\mathrm{Tr}\left[\mathcal{N}^\dagger(P_j)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_{j}\prod_m\mathrm{Tr}\left[\mathcal{N}_m^\dagger(P_{j_m})^2\right]\tilde{\delta}_{j,\lambda}\\ &= \prod_m\frac{1}{d_m}\sum_{j_m=1}^{d_m}\mathrm{Tr}\left[\mathcal{N}_m^\dagger(P_{j_m})^2\right]\\ &= \prod_m\frac{1}{d_m}\sum_{j_m=1}^{d_m}\|\mathcal{N}_m^\dagger(P_{j_m})\|_2^2,\end{aligned} \tag{C11}$$

where the product is carried over all partitions corresponding to the local 2-designs. By Lemma E2, $\|\mathcal{N}_m^\dagger(P_{j_m})\|_2^2\leq\lambda_{j_m}^2+t_{j_m}^2\leq 1$. We claim that, when $\lambda_m=1$, then at least one of such terms in the inner sum in Eq. (C11) is strictly less than one. Indeed, if the equality holds for all j_m, since we are free to choose the local basis P_{j_m}, this would imply that the channel is norm-preserving, i.e., unitary, which is a contradiction since $\mathcal{N}_m$ are assumed to be nonunitary channels. In turn, this implies that the average is also strictly less than one. Contrarily, when $\lambda_m=0$, $\|\mathcal{N}_m^\dagger(P_{0_m})\|_2^2=1$ by trace preservation.

Let c_m denote the average

$$c_m=\frac{1}{d_m}\sum_{j_m=1}^{d_m}\|\mathcal{N}_m^\dagger(P_{j_m})\|_2^2<1, \tag{C12}$$

and let $c=\max_{m|\lambda_m=1}c_m$. Then

$$\sum_\kappa Q_{\kappa,\lambda}<\prod_m c=c^{|\lambda|}, \tag{C13}$$

where $|\lambda|$ denotes the number of nonzero entries of λ, which proves the first part.

For the second part, recall that, by definition of locality vector, if $H=\sum_i h_iP_i$, then $\ell_H=\sum_i h_i^2\ell_{P_i}$, so the absorption coefficients can be computed as

$$\frac{(A\ell_H)_0}{d}=\sum_i h_i^2\frac{(A\ell_{P_i})_0}{d}. \tag{C14}$$

Equation (C4) already gives us a lower bound,

$$\frac{(A\ell_{P_i})_0}{d}\geq(\ell_\tau,\ell_{P_i})=\frac{(\ell_\tau)_\kappa}{d_k}\in\Omega\left(\frac{1}{d_\kappa}\right)=e^{\Omega(|\kappa|)}. \tag{C15}$$

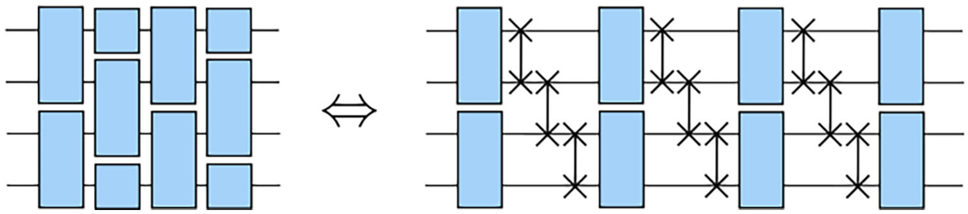

FIG. 4. Relation between the circuit structure used in [32,33] and the present work. The interleaved local two-design architecture is a special case of Eq. (2), with a SWAP cascade as entangling layer.

For the upper bound, we estimate the scaling of $(\ell_\tau, (\mathbb{1} - Q)^{-1} Q \ell_{P_i})$. Let $\tilde{P}_i$ be an operator such that $\ell_{\tilde{P}_i} = Q\ell_{P_i}$. Then

$$\begin{aligned}(\ell_\tau, (\mathbb{1} - Q)^{-1} Q \ell_{P_i}) &\leq (\ell_\tau, (\mathbb{1} - Q)^{-1} \ell_{\tilde{P}_i}) \\ &\leq \mathbb{V}^\infty_{\tau,\tilde{P}_i} \leq \|\tau\|_2^2 \|\tilde{P}_i\|_2^2,\end{aligned} \tag{C16}$$

by construction, since single-qubit channels are contractive under the 2-norm [46]. By construction, we have

$$\|\tilde{P}_i\|_2^2 = \sum_\kappa (Q\ell_{P_i})_\kappa = \sum_{\kappa,\lambda} Q_{\kappa,\lambda} (\ell_{P_i})_\lambda \leq \|P_i\|_2^2 c^{|P_i|}, \tag{C17}$$

since $(\ell_{P_i})_\lambda = \tilde{\delta}_{i,\lambda}$, which shows the upper bound completing the proof. ∎

This includes as a special case Theorem 6 in Ref. [32], as their model can be expressed in our formalism by considering two-qubit local subsystems $\mathcal{H}_m = (\mathbb{C}^2)^{\otimes 2}$ in the definition of Eq. (2), and a SWAP-cascade entangling layer as illustrated in Fig. 4.

2. Numerical validation

While, as shown in Proposition C2, for noise models composed of repeated applications of the same single-qubit nonunital channel, $\mathcal{B}_0$ remains the only relevant subspace in the deep circuit limit, our analysis indicates that when unital and nonunital channels are mixed, additional subspaces can also contribute nontrivially. This suggests the emergence of a novel absorption phenomenon not captured by prior models. We numerically validate this claim below with a pedagogical example.

Consider a circuit composed of four qubits, with single-qubit local subsystems $\mathcal{H}_m = \mathbb{C}^2$. Let the entangling gate be of the form Eq. (C4), where W and $\mathcal{N} = \mathcal{N}_0 \otimes \mathcal{N}_2$ are described in Fig. 5. In particular, $\mathcal{N}_0$ is a *depolarizing* channel defined as

$$\mathcal{N}_0(\rho) = \left(1 - \frac{3\gamma_0}{4}\right)\rho + \frac{\gamma_0}{4} X\rho X + \frac{\gamma_0}{4} Y\rho Y + \frac{\gamma_0}{4} Z\rho Z, \tag{C18}$$

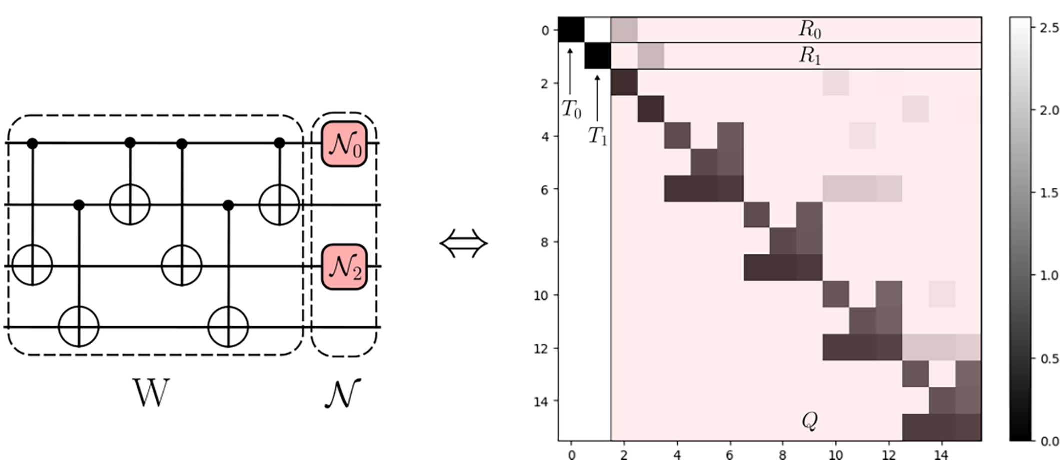

FIG. 5. Intermediate channel for the numerical experiments and corresponding locality transfer matrix. On the left, the circuit used in the numerical study, including the structure of the entangling unitary W, and noise channel $\mathcal{N} = \mathcal{N}_0 \otimes \mathcal{N}_2$, where $\mathcal{N}_0$ is a depolarizing channel and $\mathcal{N}_2$ is an amplitude damping channel. On the right, the heatmap of the locality transfer matrix entries in logarithmic scale, where strictly contractive component Q, and the absorption terms R_0 and R_1 are highlighted in red.

with $\gamma_0 = 5 \times 10^{-2}$ and $\mathcal{N}_2$ is an *amplitude damping* channel, defined as

$$\mathcal{N}_2(\rho) = K_0 \rho K_0^\dagger + K_1 \rho K_1^\dagger,$$
$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma_2} \end{pmatrix} K_0 = \begin{pmatrix} 0 & \sqrt{\gamma_2} \\ 0 & 0 \end{pmatrix}, \tag{C19}$$

with $\gamma_2 = 2 \times 10^{-1}$.

Since the dimension of the system is small, we can block diagonalize T, showing the presence of *two* distinct norm-preserving, strongly connected components, namely, the trivial component T_0 corresponding to $\mathcal{B}_I$ and T_1 corresponding to $\mathcal{B}_1 = \mathcal{B}_{\kappa=(0100)}$. Their respective dimensions are $d_0 = 1$ and $d_1 = d_{\kappa=(0100)} = d_{m=1}^2 - 1 = 3$. We now estimate the magnitude of the absorption effects using Proposition C1. We start by computing the action of $\mathcal{N}_2$ on the Pauli basis. Applying the definition, we have

$$\mathcal{N}_2(\mathbb{1}) = \mathbb{1} + \gamma_2 Z, \mathcal{N}_2(X) = \sqrt{1-\gamma_2}X,$$
$$\mathcal{N}_2(Y) = \sqrt{1-\gamma_2}Y, \mathcal{N}_2(Z) = (1-\gamma_2)Z. \tag{C20}$$

Using these relations, we can now easily compute the operators τ, and $\tilde{\Lambda}_j, j \in \mathcal{B}_{\kappa=(0100)}$. In particular, we have

$$\begin{aligned}
\tau &= \mathcal{N}_0(\mathbb{1}) \otimes \mathbb{1} \otimes \mathcal{N}_2(\mathbb{1}) \otimes \mathbb{1}/d \\
&= \mathbb{1} \otimes \mathbb{1} \otimes (\mathbb{1} + \gamma_2 Z) \otimes \mathbb{1} \\
&= \mathbb{1}/d + \gamma_2 \mathbb{1}\mathbb{1}Z\mathbb{1}/d, \\
\tilde{\Lambda}_j &= \mathcal{N}_0(\mathbb{1}) \otimes P_{j_1} \otimes \mathcal{N}_2(\mathbb{1}) \otimes \mathbb{1}/d \\
&= \mathbb{1} \otimes P_{j_1} \otimes (\mathbb{1} + \gamma_2 Z) \otimes \mathbb{1} \\
&= \mathbb{1}P_{j_1}\mathbb{1}\mathbb{1}/d + \gamma_2 \mathbb{1}P_{j_1}Z\mathbb{1}/d,
\end{aligned} \tag{C21}$$

which yields the operators $\Lambda_0 = \mathcal{Q}(\tau) = \gamma_2 \mathbb{1}\mathbb{1}Z\mathbb{1}/d$ and $\Lambda_j = \mathcal{Q}(\tilde{\Lambda}_j) = \gamma_2 \mathbb{1}P_{j_1}Z\mathbb{1}/d$ to estimate the absorption coefficients. In particular, this leads to the locality vectors

$$(\ell_{\Lambda_0})_\kappa = \gamma_2^2 \delta_{\kappa,(0010)} \text{ and } \sum_j (\ell_{\Lambda_j})_\kappa = 3\gamma_2^2 \delta_{\kappa,(0110)}. \tag{C22}$$

This implies that, by choosing $H_0 \in \mathcal{B}_{\kappa=(0010)}$ and $H_1 \in \mathcal{B}_{\kappa=(0110)}$, we are guaranteed to have a nonvanishing absorption term to $\mathcal{B}_0$ and $\mathcal{B}_1$, respectively. Let's fix them to $H_0 = \mathbb{1}\mathbb{1}X\mathbb{1}$ and $H_1 = \mathbb{1}XX\mathbb{1}$. Then we have

$$(A\ell_{H_0})_0 \geq d\gamma_2^2/3 \text{ and } (A\ell_{H_1})_1 \geq d\gamma_2^2/3, \tag{C23}$$

where the factors of 3 appear in accordance with the dimensions $d_{\kappa=(0010)} = 3$ and $d_{\kappa=(0110)} = 9$. The last step involves putting the lower bounds of Eq. (C23) back into the formula for $\mathbb{V}_{\rho,H}^\infty$ of Eq. (15), namely, estimating the quantity

$$\frac{(\ell_\rho)_z (A\ell_H)_d}{d_z}. \tag{C24}$$

Since all quantum states satisfy $\mathrm{Tr}[\rho] = 1$, $(\ell_\rho)_0 = 1/d \forall \rho$, and the first absorption term is always lower bounded by $\gamma_2^2/3$, regardless of the initial state.

On the contrary, the other absorption term depends on the locality of the initial state on $\mathcal{B}_{z=1}$, which is not guaranteed to be nonzero in general. As an example, the state $\rho_A = |0000\rangle\langle 0000|$ has nontrivial component here, namely, $(\ell_{\rho_0})_{z=1} = 1/d$, but the mixed state $\rho_B \propto |0000\rangle\langle 0000| + |0110\rangle\langle 0110|$ does not and $(\ell_{\rho_B})_{z=1} = 0$. Without resorting to mixed states, also the entangled GHZ state ρ_{GHZ} can be shown to similarly satisfy $(\ell_{\rho_{\mathrm{GHZ}}})_{z=1} = 0$. Indeed, the lower bound established by Proposition C1 only holds for the H-term in $\mathbb{V}_{\rho,H}^\infty$ of Eq. (15), but the contribution could still result in an exponentially decaying variance if the corresponding ρ-term vanishes. Putting this into Eq. (C24), we get the lower bounds

$$\frac{(\ell_\rho)_0 (A\ell_{H_0})_0}{d_0} \geq \frac{\gamma_2^2}{3} \forall \rho \tag{C25}$$

and

$$\frac{(\ell_{\rho_A})_1 (A\ell_{H_1})_1}{d_1} \geq \frac{\gamma_2^2}{9}, \frac{(\ell_{\rho_B})_1 (A\ell_{H_1})_1}{d_1} = 0. \tag{C26}$$

This means that, using the same circuit and observable, we can get a NIBP concentration effect or not, depending on the initial state that we select. Moreover, since H_0 and H_1 are orthogonal Pauli strings, $\ell_{H_0+H_1} = \ell_{H_0} + \ell_{H_1}$, suggesting that, by using the observable $H_{\mathrm{tot}} = H_0 + H_1$, we are guaranteed to observe both contributions when using ρ_A as an initial state.

In the rest of this section, we numerically validate the above claims. In particular, we study the behavior of $\mathbb{V}_{\rho,H}^L$ as for an increasing amount of layers L, changing observable H and initial state ρ to isolate different kinds of contributions. Recalling Lemma E3, we can decompose the variance contributions like

$$\begin{aligned}
\mathbb{V}_{\rho,H}^L &= (\ell_\rho, T^L \ell_H) \\
&= \sum_z \left[(\ell_\rho, T_z^L \ell_H) + (\ell_\rho, A_z^{(L)} \ell_H) \right] + (\ell_\rho, Q^L \ell_H),
\end{aligned} \tag{C27}$$

where the sum includes the nonvanishing terms in the limit of large L and $(\ell_\rho, Q^L \ell_H)$, which instead vanishes in the limit. Thanks to the choice of ρ_A and ρ_B, we have that $(\ell_{\rho_A}, Q^L \ell_{H_1}) = (\ell_{\rho_B}, Q^L \ell_{H_1})$, thus removing the ambiguity when both initial states are present in the same plot.

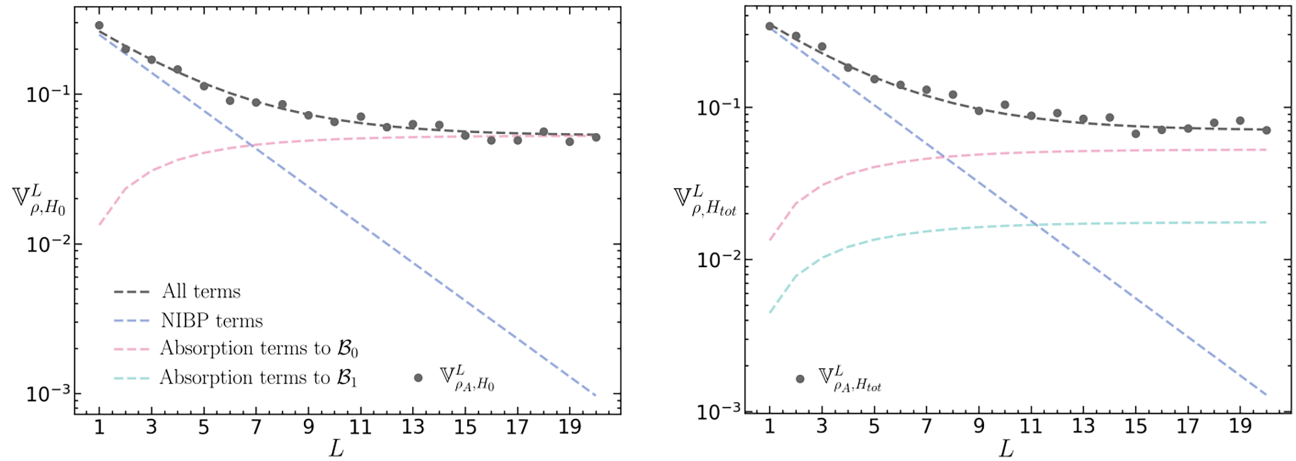

FIG. 6. Example of the absorption effect in a system comprising a mixture of clean and noisy qubits, as described in Fig. 5. On the left, we set $H = H_0$ and $\rho = \rho_A$. For this choice of H, the absorption terms to $\mathcal{B}_{z=1}$ vanish, enabling us to isolate and verify the model's ability to accurately reproduce the nonvanishing variance contributions analogous to those discussed in the literature, but within a more general framework. On the right, we set $H = H_{\text{tot}}$ and $\rho = \rho_A$. In this case, both absorption terms are nonzero, illustrating how, in the more general setting, a combination of the two effects—initial-state-dependent and initial-state-independent—can occur. The total variance matches the sum of these contributions, in agreement with the prediction of Eq. (15).

Furthermore, since for both $H = H_0$ and $H = H_1$, $H \in \mathcal{B}_Q$, the norm-preserving terms vanish, i.e., $T_z^L \ell_H = 0\,\forall L$, leaving only the contributions $(\ell_\rho, A_z^{(L)}\ell_H)$. As rigorously proven in Appendix E, these transient quantities are related to the above analysis by the limit

$$\lim_{L\to\infty} (\ell_\rho, A_z^{(L)}\ell_H) = \frac{(\ell_\rho)_z (A\ell_H)_z}{d_z}, \tag{C28}$$

and in this sense represent the finite-depth equivalent of absorption terms in Theorem IV.1.

After computing each term for the current model, we estimate the variance numerically by randomly sampling $R = 200$ circuit parameters and computing the sample variance of the loss function defined in Eq. (1). Simulations are performed using the Pennylane software framework [47].

The results are presented in both Fig. 3 and Fig. 6. In Fig. 3, we numerically demonstrate the emergence of an absorption term in the variance that depends on the circuit's initial state. Specifically, we show that an NIBP effect can be suppressed by appropriately choosing the initial state ($\rho = \rho_A$) and can reappear otherwise ($\rho = \rho_B$). The agreement with the theoretical predictions indicates that this behavior is explained by absorption into a nontrivial, norm-preserving invariant subspace $\mathcal{B}_{z=1}$. This stands in stark contrast to previous descriptions of loss concentration under both unital and nonunital noise, in which the initial state played no role in determining the limiting variance. Furthermore, as shown in Fig. 6, we also identify an initial-state-independent contribution from absorption into the trivial subspace $\mathcal{B}_0$ (left) as well as a regime combining both effects (right).

Finally, we note that while the numerical simulations in this example are feasible due to the small size of the system, the theoretical tools from Appendix C used to motivate this example do not share this limitation and are, in principle, applicable to much larger systems.

APPENDIX D: PROOF OF PROPOSITION IV.1

In this appendix we provide a proof for the building blocks of the main results of this work. What follows hinge on the structure of the circuit composed of L layers of interleaved unitary and noise channels

$$\Phi_\theta = \mathcal{U}_{\theta_{L+1}} \circ \mathcal{E}_L \circ \mathcal{U}_{\theta_{L-1}} \ldots \circ \mathcal{E}_1 \circ \mathcal{U}_{\theta_1}. \tag{D1}$$

In particular, we assume that $\mathcal{U}_\theta : \rho \mapsto U_\theta \rho U_\theta^\dagger$, where $U_\theta = \bigotimes_m U_{\theta_m}^{(m)}$ is a *local 2-design* for the system. This statement is more precisely captured here.

Definition D1 (Local design). Given a unitary ensemble $\{U_\theta\}_{\theta\in\Theta}$ with a given probability distribution over the parameter space Θ, we say it forms a local t-design for the system if each element is factorized with respect to the partition, i.e., $U_\theta = \bigotimes_m U_{\theta_m}^{(m)}$ each acting solely on $\mathcal{H}_m$, and additionally

$$\int_\Theta d\theta\, U_\theta^{(m)\otimes t} \otimes U_\theta^{(m)*\otimes t} = \int_{V\in U(d_m)} d\mu(V) V^{(m)\otimes t} \otimes V^{(m)*\otimes t}, \tag{D2}$$

where the second integral is performed with respect to the Haar measure.

With this notation in place, we are ready to start. The first Lemma provides a formula for the expectation value of a circuit in the aforementioned class, showing that they form global 1-designs.

Lemma D1 (Global 1-design). Let $A, B \in \mathcal{B}$, and let $\{U_\theta\}_{\theta\in\Theta}$ be a unitary ensemble forming a local 1-design. Then

$$\mathbb{E}_\theta\left\{\mathrm{Tr}\left[AU_\theta^\dagger BU_\theta\right]\right\} = \frac{\mathrm{Tr}\,[A]\,\mathrm{Tr}\,[B]}{d}. \tag{D3}$$

Proof. Let $\{P_j\}_j$ be a local basis for the system, and consider the respective decompositions of A and B, namely, $A = \sum_i a_i P_i$ and $B = \sum_j b_j P_j$. Note that each component a_i is defined as $a_i = \mathrm{Tr}\,[P_i A]$, and consequently $a_0 = (1/\sqrt{d})\mathrm{Tr}\,[A]$ (respectively for B). Then we have the following chain of equalities:

$$\begin{aligned}
&\int_\Theta \prod_{m=1}^M d\theta\,\mathrm{Tr}\left[AU_\theta^\dagger BU_\theta\right] = \\
&= \sum_{i,j} a_i b_j \int_\Theta \prod_{m=1}^M d\theta\,\mathrm{Tr}\left[P_i U_\theta^\dagger P_j U_\theta\right] \\
&= \sum_{i,j} a_i b_j \prod_{m=1}^M \int_{U\in U(d_m)} d\mu(U)\mathrm{Tr}\,[[m]]\, P_{i_m}^{(m)} U^\dagger P_{j_m}^{(m)} U \\
&= \sum_{i,j} a_i b_j \prod_{m=1}^M \frac{1}{d_m}\mathrm{Tr}\,[[m]]\, P_{i_m}^{(m)}\mathrm{Tr}\,[[m]]\, P_{j_m}^{(m)} \\
&= \sum_{i,j} a_i b_j \prod_{m=1}^M \delta_{0,i_m}\delta_{0,j_m} = a_0 b_0,
\end{aligned} \tag{D4}$$

where Tr_m denotes the partial trace over the mth subsystem. ■

Regarding the second moment, it can be computed using the following lemma:

Lemma D2. Let $\{P_j\}$ be a local basis and let $\{U_\theta\}_{\theta\in\Theta}$ be a unitary ensemble forming a local 2-design. Then

$$\begin{aligned}
&\mathbb{E}_\theta\left\{\mathrm{Tr}\left[P_i U_\theta^\dagger P_j U_\theta\right]\mathrm{Tr}\left[P_k U_\theta^\dagger P_l U_\theta\right]\right\} = \\
&= \delta_{ij}\,\delta_{kl}\prod_{m=1}^M\left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0j_m})\frac{1}{d_m^2-1}\right).
\end{aligned} \tag{D5}$$

Proof. To prove this, we make use of the following result of Weingarten Calculus

$$\begin{aligned}
&\int_{U\in U(d)} d\mu(U)\mathrm{Tr}\left[AU^\dagger BU\right]\mathrm{Tr}\left[CU^\dagger DU\right] \\
&= \frac{1}{d^2-1}\left(\mathrm{Tr}\,[A]\,\mathrm{Tr}\,[B]\,\mathrm{Tr}\,[C]\,\mathrm{Tr}\,[D] + \mathrm{Tr}\,[AC]\,\mathrm{Tr}\,[BD]\right) \\
&\quad + \frac{1}{d(d^2-1)}\left(\mathrm{Tr}\,[AC]\,\mathrm{Tr}\,[C]\,\mathrm{Tr}\,[D]\right. \\
&\quad \left.+ \mathrm{Tr}\,[A]\,\mathrm{Tr}\,[C]\,\mathrm{Tr}\,[BD]\right).
\end{aligned} \tag{D6}$$

Based on Eq. (D6), the result follows from direct integration:

$$\begin{aligned}
&\prod_{m=1}^M \int_{U\in U(d_m)} d\mu(U)\mathrm{Tr}\left[P_{i_m}^{(m)} U^\dagger P_{j_m}^{(m)} U\right]\times \\
&\quad \times\,\mathrm{Tr}\left[P_{k_m}^{(m)} U^\dagger P_{l_m}^{(m)} U\right] \\
&= \prod_{m=1}^M \frac{1}{d_m^2-1}\Big(\mathrm{Tr}\left[P_{i_m}^{(m)}\right]\mathrm{Tr}\left[P_{j_m}^{(m)}\right]\mathrm{Tr}\left[P_{k_m}^{(m)}\right]\mathrm{Tr}\left[P_{l_m}^{(m)}\right] \\
&\quad + \mathrm{Tr}\left[P_{i_m}^{(m)}P_{k_m}^{(m)}\right]\mathrm{Tr}\left[P_{j_m}^{(m)}P_{l_m}^{(m)}\right]\Big) \\
&\quad - \frac{1}{d_m(d_m^2-1)}\Big(\mathrm{Tr}\left[P_{i_m}^{(m)}P_{k_m}^{(m)}\right]\mathrm{Tr}\left[P_{j_m}^{(m)}\right]\mathrm{Tr}\left[P_{l_m}^{(m)}\right] \\
&\quad + \mathrm{Tr}\left[P_{i_m}^{(m)}\right]\mathrm{Tr}\left[P_{k_m}^{(m)}\right]\mathrm{Tr}\left[P_{j_m}^{(m)}P_{l_m}^{(m)}\right]\Big) \\
&= \prod_{m=1}^M \frac{1}{d_m^2-1}(d_m^2\,\delta_{0i_m}\delta_{0j_m}\delta_{0k_m}\delta_{0l_m} \\
&\quad + \delta_{i_m k_m}\delta_{j_m l_m} - \delta_{0i_m}\delta_{0k_m}\delta_{j_m l_m} - \delta_{i_m k_m}\delta_{0j_m}\delta_{0l_m}) \\
&= \prod_{m=1}^M \delta_{0i_m}\delta_{0j_m}\delta_{0k_m}\delta_{0l_m} \\
&\quad + \frac{1}{d_m^2-1}(\delta_{i_m k_m} - \delta_{0i_m}\delta_{0k_m})(\delta_{j_m l_m} - \delta_{0j_m}\delta_{0l_m}) \\
&= \delta_{ik}\delta_{jl}\prod_{m=1}^M\left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0j_m})\frac{1}{d_m^2-1}\right).
\end{aligned} \tag{D7}$$

■

In particular, Lemma D2 can be used to compute the variance of loss functions computed as expectation values $\mathcal{L}_{\rho,H} = \mathrm{Tr}\,[\mathcal{U}_\theta(A)B]$, i.e., in the absence of intermediate channels $\mathcal{E}_l$.

Proposition D1. Let $A, B \in \mathcal{B}$, and let $\{U_\theta\}_{\theta\in\Theta}$ be a unitary ensemble forming a local 2-design. Then

$$\mathbb{E}_\theta\left\{\mathrm{Tr}\left[AU_\theta^\dagger BU_\theta\right]^2\right\} = (\ell_A, \ell_B)\,, \tag{D8}$$

where $(\cdot,\cdot)$ is the scalar product defined in Eq. (7).

Proof. Let $\{P_j\}$ be a local basis for the system and consider the respective decompositions of A and B. By Lemma D2 we have

$$\mathbb{E}_\theta\left\{\mathrm{Tr}\left[AU_\theta^\dagger BU_\theta\right]^2\right\} = \sum_{i,j} a_i^2 b_j^2 \prod_{m=1}^{M}\left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0j_m})\frac{1}{d_m^2-1}\right). \tag{D9}$$

In the following, it will be convenient to recast the product on the right-hand side of Eq. (D9) into the equivalent formulation

$$\prod_{m=1}^{M}\left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0i_m})\frac{1}{d_m^2-1}\right) = \\ = \sum_{\kappa\in\{0,1\}^M}\frac{1}{d_\kappa}\prod_{m=1}^{M}(\delta_{0i_m}\delta_{0j_m})^{1-\kappa_m}(1-\delta_{0i_m})^{\kappa_m}(1-\delta_{0j_m})^{\kappa_m}, \tag{D10}$$

where the binary vectors $\kappa \in \{0,1\}^M$ identify all possible sets S_κ introduced in Appendix B and $d_\kappa = \prod_{m=1}^{M}(d_m^2-1)^{\kappa_m}$. Putting it back into Eq. (D9) we get

$$\begin{aligned}
&\sum_{\kappa\in\{0,1\}^M}\frac{1}{d_\kappa}\sum_{i,j}a_i^2b_j^2\prod_{m=1}^{M}(\delta_{0i_m}\delta_{0j_m})^{1-\kappa_m}\times \\
&\quad\times(1-\delta_{0i_m})^{\kappa_m}(1-\delta_{0j_m})^{\kappa_m} \\
&= \sum_{\kappa\in\{0,1\}^M}\frac{1}{d_\kappa}\left(\sum_i a_i^2\prod_{m=1}^{M}\delta_{0i_m}^{1-\kappa_m}(1-\delta_{0i_m})^{\kappa_m}\right)\times \\
&\quad\times\left(\sum_j b_j^2\prod_{m=1}^{M}\delta_{0j_m}^{1-\kappa_m}(1-\delta_{0j_m})^{\kappa_m}\right) \\
&= \sum_{\kappa\in\{0,1\}^M}\frac{1}{d_\kappa}\left(\sum_i a_i^2\tilde{\delta}_{i,\kappa}\right)\left(\sum_j b_j^2\tilde{\delta}_{j,\kappa}\right) \\
&= \sum_{\kappa\in\{0,1\}^M}\frac{(\ell_A)_\kappa(\ell_B)_\kappa}{d_\kappa} = (\ell_A, \ell_B),
\end{aligned} \tag{D11}$$

from which the proposition follows. ∎

This can be extended to the general case introducing the action of the intermediate channels $\mathcal{E}_l$ and, in particular, we get the following proposition.

Proposition D2. Let $A, B \in \mathcal{B}$ and $\Lambda : \mathcal{B} \to \mathcal{B}$ be a linear map. Furthermore, let $\{U_{\theta_1}\}_{\theta_1\in\Theta}$ and $\{V_{\theta_2}\}_{\theta_2\in\Theta}$ be independent, unitary ensembles each forming a local 2-design. Then

$$\mathbb{E}_{\theta_1,\theta_2}\left\{\mathrm{Tr}\left[AU_{\theta_1}^\dagger\Lambda(V_{\theta_2}BV_{\theta_2}^\dagger)U_{\theta_1}\right]^2,\right\} = (\ell_A, T\ell_B), \tag{D12}$$

where T is the locality transfer matrix associated with Λ.

Proof. Let $\tilde{B}_{\theta_2} = \Lambda(V_{\theta_2}BV_{\theta_2})$. By Proposition D2, we have

$$\begin{aligned}
&\mathbb{E}_{\theta_2}\mathbb{E}_{\theta_1}\left\{\mathrm{Tr}\left[AU_{\theta_1}^\dagger B_{\theta_2}U_{\theta_1}\right]^2\right\} \\
&= \mathbb{E}_{\theta_2}\left\{(\ell_A, \ell_{\tilde{B}_{\theta_2}})\right\} \\
&= \sum_{\kappa\in\{0,1\}^M}\frac{1}{d_\kappa}(\ell_A)_\kappa\,\mathbb{E}_{\theta_2}\left\{(\ell_{\tilde{B}_{\theta_2}})_\kappa\right\}.
\end{aligned} \tag{D13}$$

Expanding the definition of the last term with respect to the local basis $\{P_j\}_j$ and applying again Proposition D2 we get

$$\begin{aligned}
&\mathbb{E}_{\theta_2}\left\{(\ell_{\tilde{B}_{\theta_2}})_\kappa\right\} = \sum_i\mathbb{E}_{\theta_2}\left\{\mathrm{Tr}\left[P_i\tilde{B}_{\theta_2}\right]^2\right\}\tilde{\delta}_{i,\kappa} \\
&= \sum_i\mathbb{E}_{\theta_2}\left\{\mathrm{Tr}\left[\Lambda^\dagger(P_i)V_{\theta_2}^\dagger BV_{\theta_2}\right]^2\right\}\tilde{\delta}_{i,\kappa} \\
&= \sum_{\lambda\in\{0,1\}^M}\frac{1}{d_\lambda}\sum_{i,j,k}\mathrm{Tr}\left[P_i\Lambda(P_j)\right]^2\mathrm{Tr}\left[P_kB\right]^2\tilde{\delta}_{j,\lambda}\tilde{\delta}_{k,\lambda}\tilde{\delta}_{i,\kappa} \\
&= \sum_{\lambda\in\{0,1\}^M}\left(\frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\left[P_i\Lambda(P_j)\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda}\right)\times \\
&\quad\times\left(\sum_k\mathrm{Tr}\left[P_kA\right]^2\tilde{\delta}_{k,\lambda}\right) \\
&= \sum_{\lambda\in\{0,1\}^M}T_{\kappa,\lambda}(\ell_B)_\lambda = (T\ell_B)_\kappa,
\end{aligned} \tag{D14}$$

where $\Lambda^\dagger$ is the Hermitian adjoint of Λ with respect to the Hilbert-Schmidt scalar product. ∎

Iterated application of Proposition D2 for an initial state ρ, an observable H, and a general intermediate quantum channel $\mathcal{E}$, yields Eq. (8).

APPENDIX E: PROOF OF THEOREM IV.1

The proof of Theorem IV.1 is based on the characterization of the general LTM for the Hermitian adjoint $\mathcal{E}^\dagger$ of arbitrary quantum channel. To do so, several aspects of non-negative matrix theory, as well as the contractivity properties of $\mathcal{E}^\dagger$ are key. For the sake of clarity and completeness we recall them in the following.

1. Preliminaries

In this appendix, we start with preliminary concepts and definitions involving non-negative matrices, and then we recall some well-known facts and definitions about operator norms and quantum channel contractivity.

a. Elements of non-negative matrix theory

In this appendix, we briefly recap on the main results on non-negative matrix theory useful in the proof of Theorem IV.1. For a complete discussion and proofs of the cited results, we refer the interested reader to Refs. [35,37]. Let's start by the definition of non-negative matrix.

Definition E1 (Non-negative matrix). A $n \times n$ matrix T is said to be non-negative if each entry $(T)_{ij} \geq 0$.

The general behavior of non-negative matrices can vary greatly, but there is a class of matrices, called *irreducible*, which have very informative spectral properties.

Definition E2 (Irreducible matrix). A $n \times n$ non-negative matrix T is said to be *irreducible* if for two arbitrary indices $i,j = 1, \ldots n$, there exist $l = l(i,j) \in \mathbb{N}$ such that $(T^l)_{ij} > 0$. Moreover, we will say that T has period p, where p is the greatest common divisor of all $l(i,i)$ that satisfy $(T^l)_{ii} > 0\ \forall i$.

Equivalently, if we introduce the graph $\mathcal{G}_T$ whose adjacency matrix is T, then it can be shown that T is irreducible if and only if $\mathcal{G}_T$ is strongly connected and that the period p reduces to the great common divisor of the lengths of all closed directed paths in $\mathcal{G}_T$ [35]. Furthermore, it will be useful in the following to distinguish two classes of irreducible matrices, namely, *cyclic* (or *periodic*) and *primitive* (or *aperiodic*), which are characterized as having period $p > 1$ and $p = 1$, respectively.

One of the main results involving irreducible matrices is the celebrated *Perron-Frobenius* theorem, which characterizes the spectral properties of this class. We recall it here for convenience.

Theorem E1 (Perron-Frobenius). Let T be a $n \times n$ non-negative, irreducible matrix. Then there exists an eigenvalue r of T, with corresponding right and left eigenvectors v, w such that

(a) $r \in \mathbb{R}, r > 0$ and is a simple root of the characteristic polynomial,
(b) both w,v are the only eigenvectors that have strictly positive components, i.e., $v_i, w_i > 0\ \forall i = 1, \ldots n$,
(c) v and w are unique up to a scalar multiple, and hence can be taken to be normalized, i.e., $v^t w = 1$,
(d) $r \geq |\lambda|$, for all eigenvalue λ of T,

where r is called the Perron-Frobenius eigenvalue and $P = wv^t$ the Perron projector. Moreover, if T is also aperiodic, then we have the more restrictive

(d') $r > |\lambda|$, for all eigenvalue $\lambda \neq r$ of T.

Another important result is the so-called subinvariance theorem, which is a useful tool to bound the value of r for a given irreducible matrix.

Theorem E2 (Subinvariance Theorem). Let T be a $n \times n$ non-negative, irreducible matrix, $s > 0$ and y be a n-dimensional row vector such that each component $y_i \geq 0$ and satisfying

$$y^t T \leq s y^t, \tag{E1}$$

componentwise. Then $y_i > 0 \forall i$ and $s \geq r$. Moreover, equality holds if and only if $s = r$.

Finally, the last result allows us to exploit the knowledge of the dominant eigenvalue to determine the asymptotic properties of T^L.

Theorem E3 (Asymptotic behavior of irreducible matrices). Let T be a $n \times n$ non-negative, irreducible matrix. Then the Cesàro average of T converges and we have

$$\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} T^l / r^l = P. \tag{E2}$$

Moreover, if T is also aperiodic, then limit of T^L / r^L converges, and we have

$$\lim_{L \to \infty} T^L / r^L = P, \tag{E3}$$

where r and P are the Perron-Frobenius eigenvalue and Perron projector, respectively.

Despite being less structured, it is a well-known fact that general non-negative matrices can be cast to a canonical block upper triangular form, where all blocks in the diagonal are *irreducible* simply by means of a permutation matrix, i.e., by a relabeling of the basis elements. In particular, concerning the diagonal, irreducible blocks appearing in such decomposition, we will use the term *essential* when referring to the blocks such that all $(T)_{i,j} = 0$ for all columns apart from the block itself and *inessential* otherwise. In terms of the graph $\mathcal{G}_T$, this distinction is readily understood. As discussed above, irreducible blocks correspond to strongly connected components, and consequently essential blocks are strongly connected components, which do not have edges connecting vertices in it to vertices pertaining to other components. More simply, we can describe essential components as those whose edges "do not lead outside." In what follows, we name by T_z all essential, irreducible components, and we group into a single block Q all inessential components. The blocks R_z, appearing on top of the block Q, represent the collection of edges coming from inessential components and leading to essential ones. A graphical summary is depicted in Eq. (E4).

$$T = \begin{pmatrix} \boxed{T_0} & & & & \boxed{R_0} \\ & \boxed{T_1} & & & \boxed{R_1} \\ & & \ddots & & \vdots \\ & & & \boxed{T_z} & \boxed{R_z} \\ & & & & \boxed{Q} \end{pmatrix} \tag{E4}$$

This, in particular, allows us to apply the results of this section also to more generic matrices, such as LTMs, which in general are not irreducible.

b. Useful results on quantum channel

In this appendix, we briefly introduce some relevant properties of completely positive (CP) maps. These will be especially useful in the trace preserving case (CPTP), i.e., quantum channels, and the unital case (CPU), i.e., the corresponding adjoint action with respect to the Hilbert-Schmidt scalar product. In what follows, we will denote by $\mathbb{M}_n$ the space of $n \times n$ matrices, which we can endow with a norm as follows.

Definition E3 (Schatten norm). Given $A \in \mathbb{M}_n$ and $p \in [1, \infty]$, we define the Schatten p-norm $\|\cdot\|_p : \mathbb{M}_n \to \mathbb{R}$ as

$$\|A\|_p = \mathrm{Tr}\left[\left(\sqrt{A^\dagger A}\right)^p\right]^{1/p}. \tag{E5}$$

A crucial property of Schatten norms is Hölder inequality, namely, $|\mathrm{Tr}\left[A^\dagger B\right]| \leq \|A\|_p \|B\|_q$, $\forall A, B \in \mathbb{M}_n, \forall p, q$ s.t. $1/p + 1/q = 1$. As a special case, for $p = 2$ we get back the Hilbert-Schmidt norm, and Hölder inequality reduces to the Cauchy-Schwarz inequality.

As a direct consequence of Definition E3, an induced norm on linear operators acting on $\mathbb{M}_n$ can be defined.

Definition E4 (Induced norm). Given a linear operator $\Lambda : \mathbb{M}_n \to \mathbb{M}_n$ we define the induced $p \to q$ norm as

$$\|\Lambda\|_{p\to q} := \sup_{A:\|A\|_p=1} \|\Lambda(A)\|_q. \tag{E6}$$

Depending on the properties of such induced norms, we might refer to the map Λ as contractive or strictly contractive. In particular, we will use the following definitions.

Definition E5 (Contractivity of linear maps). A linear operator $\Lambda : \mathbb{M}_n \to \mathbb{M}_n$ is said to be *contractive* with respect to the p norm if $\|\Lambda\|_{p\to p} \leq 1$, and similarly to be *strictly contractive* if $\|\Lambda\|_{p\to p} < 1$.

Linear maps that are also CPTP are known to always be contractive with respect to the 1-norm [48,49], but are in general not contractive for other p-norms. This property, together with Hölder's inequality, allows putting an upper bound on the value of the variance of an arbitrary, layered quantum circuit.

Lemma E1 (Hölder's inequality for variances). Let $A, B \in \mathbb{M}_n$ be hermitian matrices, and $\Phi_\theta : \mathbb{M}_n \to \mathbb{M}_n$ be a parameterized quantum channel. In particular, consider the L layered map of type Eq. (2). Then

$$\left|\mathrm{Tr}\left[\Phi_\theta^L(A)B\right]\right| \leq \|\mathcal{E}\|_{p\to p}^L \|\|A\|_p \|B\|_q \,\forall\theta \in \Theta, \tag{E7}$$

with $1/p + 1/q = 1$. As a special case, if $A = \rho$ and $B = H$ are a density operator and an observable, respectively, then the variance can be upper bounded by $\mathbb{V}_{\rho,H}^L \leq \|H\|_\infty^2 \forall L$.

Proof. The result is the direct consequence of Hölder's inequality and contractivity of quantum maps. In particular, we have the following chain of inequalities:

$$\begin{aligned} \left|\mathrm{Tr}\left[\Phi_\theta^L(A)B\right]\right| &\leq \|\Phi_\theta^L(A)\|_p \|B\|_q \\ &= \|U_{\theta_L}\mathcal{E}\left(\Phi_\theta^{L-1}(A)\right) U_{\theta_L}^\dagger\|_p \|B\|_q \\ &= \|\mathcal{E}\left(\Phi_\theta^{L-1}(A)\right)\|_p \|B\|_q \\ &\leq \|\mathcal{E}\|_{p\to p} \|\Phi_\theta^{L-1}(A)\|_p \|B\|_q \,\forall\theta \in \Theta. \end{aligned} \tag{E8}$$

By iterative application of this procedure, one can get the result. The final remark holds due to the contractivity of quantum channels, choosing $p = 1$ and $q = \infty$, and noting that $\|\rho\|_1 \leq 1$ for all density matrices. ■

Specific classes of quantum channels can be shown to be contractive with respect to a wider variety of norms. In particular, we have that, for $p \geq 2$, *unitality* of the map is a necessary and sufficient condition for contractivity (see Theorem II.4 in [46]). Within unital channels, unitary transformation $\mathcal{U}$ always saturate the bound, as they have the additional property of being norm-preserving, i.e., $\|\mathcal{U}(A)\|_p = \|U^\dagger A U\|_p = \|A\|_p$. Finally, if we reduce the action of the channel to the subset $\mathbb{H}_0 \subset \mathbb{M}_n$ of Hermitian, traceless matrices, then the unitality property is no longer a necessary condition for contractivity. Indeed, any single-qubit channel $\mathcal{N}$ is contractive in this setting, i.e., $\|\mathcal{N}|_{\mathbb{H}_0}\|_{p\to p} \leq 1 \forall p$ if $n=2$. For single-qubit channels we can even be more explicit, as shown in the following lemma.

Lemma E2 (Single-qubit channel normal form). Let $\{P_i\}$ be the normalized (with respect to the Schatten 2-norm) Pauli basis of $\mathbb{M}_2$, and $\mathcal{N}$ be a single-qubit channel. Then there exist unitary matrices U, V such that $\mathcal{N}'(\cdot) = U^\dagger \mathcal{N}(V^\dagger \cdot V)U$ satisfies

$$\mathcal{N}'(P_0) = P_0 + \sum_{i>0} t_i P_i, \quad \mathcal{N}'(P_i) = \lambda_i P_i \, \forall i > 0, \tag{E9}$$

where $\sum_{i>0}(t_i + \lambda_i \alpha_i)^2 \leq 1, \forall \alpha_i \in \mathbb{R}$ s.t. $\sum_{i>0} \alpha_i^2 \leq 1$.

Proof. It has been shown in [50,51] that any single qubit quantum channel can be cast in the canonical form of Eq. (E9) by means of a change of basis. Furthermore, the constraint on the parameters follows, analogously to [32], by considering that any single-qubit state must have bounded purity, namely, $\mathrm{Tr}\left[\rho^2\right] \leq 1$, and since $\mathcal{N}'$ is a channel, the same must hold for $\mathcal{N}'(\rho)$. Since any qubit state can be decomposed in terms of $\{P_i\}_i$ as $\rho = 1/\sqrt{2}P_0 + 1/\sqrt{2}\sum_{i>0}\alpha_i P_i$, $\alpha_i \in \mathbb{R}$, we get

$$\begin{aligned} \mathrm{Tr}\left[\rho^2\right] &= \frac{1+\alpha_i^2}{2} \leq 1, \\ \mathrm{Tr}\left[\mathcal{N}'(\rho)^2\right] &= \frac{1+\sum_{i>0}(t_i+\lambda_i\alpha_i)^2}{2} \leq 1, \end{aligned} \tag{E10}$$

respectively, which concludes the proof. ■

Considering instead CPU maps, the most relevant result is Kadison-Schwarz inequality, which for our purposes can be stated as follows:

Theorem E4 (Kadison-Schwarz inequality [52]). Let $A, B \in \mathbb{H}_0$, and $\Lambda : \mathbb{M}_n \to \mathbb{M}_n$ be a CPU map. Then

$$\Lambda(A)\Lambda(B) \leq \Lambda(AB). \tag{E11}$$

All these properties will be useful to characterize the spectral properties of interest of the LTM of quantum channels.

2. Further characterizations of LTMs

We now study the structure of the LTM of a general CPU map. Thanks to this analysis, we will be able to compute the limiting value $\mathbb{V}^\infty_{\rho,H}$ by describing the quantum circuit in the Heisenberg picture. Denoting by T the resulting LTM, we start by computing the general form of integer powers T^L of T.

Lemma E3 (Limiting form of T). Let $\Lambda : \mathcal{B} \to \mathcal{B}$ be a CPU map and T be the corresponding LTM. Then T and T^L take the form

$$T = \begin{pmatrix} \boxed{T_0} & & & & \boxed{R_0} \\ & \boxed{T_1} & & & \boxed{R_1} \\ & & \ddots & & \vdots \\ & & & \boxed{T_z} & \boxed{R_z} \\ & & & & \boxed{Q} \end{pmatrix},$$

$$T^L = \begin{pmatrix} \boxed{T_0^L} & & & & \boxed{A_0^{(L)}} \\ & \boxed{T_1^L} & & & \boxed{A_1^{(L)}} \\ & & \ddots & & \vdots \\ & & & \boxed{T_z^L} & \boxed{A_z^{(L)}} \\ & & & & \boxed{Q^L} \end{pmatrix} \tag{E12}$$

up to a basis state index permutation, where each T_z is an irreducible matrix, and $A_z^{(L)} = \sum_{l=0}^{L-1} T_z^l R_z Q^{L-1-l}$.

Note that here, we split the contributions R_z corresponding to each component T_z for convenience. This is related to the convention of Fig. 2 by the relation $R = \sum_z R_z$ with clear meaning of the symbols.

Proof. We start by putting T into the canonical form of Eq. (E4). In this form, the powers of the diagonal blocks T_z^L are trivially the diagonal blocks of T^L. Instead, the result about $A^{(L)}$ follows by induction. In fact, both the base case and the inductive one follow from matrix multiplication rules of $T \cdot T$ and $T^{L-1} \cdot T$, respectively. In particular, we have

$$\begin{aligned} A_z^{(2)} &= T_z R_z + R_z Q, \\ A_z^{(L)} &= T_z^{L-1} R_z + A^{(L-1)} Q \\ &= T_z^{L-1} R_z + \sum_{l=0}^{L-2} T_z^l R_z Q^{L-1-l} \\ &= \sum_{l=0}^{L-1} T_z^l R_z Q^{L-1-l}, \end{aligned} \tag{E13}$$

which gives the proposition. ■

As already shown in Lemma E1, the value of the variance is upper bounded by $\|H\|_\infty^2$. Since by Proposition IV.1, this quantity is linked to $(\ell_\rho, T^L \ell_H)$, it is expected that the spectral radius $\rho(T)$ of T is upper bounded by 1. More specifically, we can prove the following proposition.

Proposition E1 (Spectral radius of T). Let $\Lambda : \mathcal{B} \to \mathcal{B}$ be a CPU map and T be the corresponding LTM. Then T is contractive in the sense of the spectral radius, i.e., $\rho(T) \leq 1$. Moreover, the component Q is strictly contractive, namely, $\rho(Q) < 1$.

Proof. The statement follows as a consequence of the Kadison-Schwarz inequality and the subinvariance theorem.

First note that, for a generic Λ, the structure of Q is not as well-behaved as T_z, as Q is not necessarily *irreducible*. However, as any non-negative matrix, Q can also be cast in canonical block upper triangular form by means of a basis permutation, where each diagonal block is irreducible.

$$Q = \begin{pmatrix} \boxed{Q_1} & \boxed{*} & \cdots & \boxed{*} \\ & \boxed{Q_2} & & \boxed{*} \\ & & \ddots & \vdots \\ & & & \boxed{Q_k} \end{pmatrix} \tag{E14}$$

With this in mind, we can study the spectral radius of T and Q in terms of the spectral radii of each block T_z and Q_k, i.e., the corresponding Perron eigenvalues, since $\rho(T) = \max\{\max_z r_z, \max_k r_{Q_k}\}$ and $\rho(Q) = \max_k r_{Q_k}$.

By Lemma B2, we are free to choose the basis used to express the matrix T. In particular, we choose the normalized Pauli basis $\{P_i\}$, which, besides being a local basis for $\mathcal{B}$, is also unitary up to normalization, namely, $P_i^2 = \mathbb{1}/d\,\forall i$. Exploiting Eq. (B4), we can compute the column sum of T as

$$\begin{aligned} \sum_\kappa T_{\kappa,\lambda} &= \sum_\kappa \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\left[P_i \Lambda(P_j)\right]^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} \\ &= \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\left[P_i \Lambda(P_j)\right]^2 \tilde{\delta}_{j,\lambda} \left(\sum_\kappa \tilde{\delta}_{i,\kappa}\right) \\ &= \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\left[P_i \Lambda(P_j)\right]^2 \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \mathrm{Tr}\left[\Lambda(P_j)^2\right] \tilde{\delta}_{j,\lambda} \\ &\leq \frac{1}{d_\lambda} \sum_j \mathrm{Tr}\left[\Lambda(P_j^2)\right] \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \tilde{\delta}_{j,\lambda} = 1, \end{aligned} \tag{E15}$$

where the third and last equality follow from Lemma B1, the inequality is Kadison-Schwarz and the second to last equality is the unitary property of the basis. If Λ is unitary, the inequality is saturated, and T becomes a *stochastic* matrix. In general, Eq. (E15) shows *substochasticity* of T. Indeed, this condition can be recast in vector form as $v^t T \leq v^t$, where $v_\kappa = 1 \forall \kappa$. In particular, this holds for all irreducible blocks in the diagonal, which by the subinvariance theorem implies $r_z \leq 1$ and $r_{Q_k} \leq 1$, giving $\rho(T) \leq 1$. Focusing on Q, we observe that, by definition, each irreducible block Q_k is *inessential*, i.e., is connected to some other block. In terms of the matrix T, this means that, considering the columns involving Q_k, there is always an index λ in the support of Q_k such that

$$\sum_\kappa (T - Q_k)_{\kappa,\lambda} > 0, \tag{E16}$$

which implies that $\exists \lambda$ s.t. $\sum_\kappa (Q_k)_{\kappa,\lambda} < 1$. Written in matrix form, this reads $v^t Q_k \leq v^t$, $v^t Q_k \neq v^t$, which by the subinvariance theorem, implies $r_{Q_k} \neq 1$. Putting everything together, one gets $\rho(Q) < 1$. ■

When analyzing the single irreducible components T_z, we can be more specific and find an equivalence between the value of the column-sum of the block and the value of the corresponding spectral radius. This is especially useful in the computation of the dominant eigenvectors, which is explicitly stated in the following corollary.

Corollary E1. Let T be a LTM and T_z be an irreducible block, then $\rho(T_z) = 1 \Leftrightarrow \sum_\kappa (T_z)_{\kappa,\lambda} = 1$, or equivalently $v_z^t T_z = v_z^t$, where $(v_z)_\kappa = 1 \forall \kappa$ is the left eigenvector of the dominant eigenvalue.

Proof. The result follows from the same proof strategy as above and is a direct consequence of the subinvariance theorem. ■

Intuitively, the blocks that are out of such hypothesis won't contribute to the large L limit, and indeed the contribution of the Q, T_z, and $A_z^{(L)}$ is bounded to decay exponentially in the number of layers.

Proposition E2. Let T be a LTM, and let T_z be an irreducible block with $r_z < 1$. Then, as $L \to \infty$, $\|T_z^L\| \to 0$ and $\|A_z^{(L)}\| \to 0$ exponentially fast for any matrix norm $\| \cdot \|$. Similarly, also $\|Q^L\| \to 0$.

Proof. The proposition can be proven using Gelfand's formula. In particular since $\lim_{L\to\infty} \|T_z^L\|^{1/L} = r_z$, we can always bound $\|T_z^L\| \le K\tau^L$, for some constant $K > 0$ and $\tau = r_z + \epsilon < 1$ for an arbitrarily small ϵ. In the same way, by Proposition E1 a similar result can be obtained for Q. Finally, the absorption term $A_z^{(L)}$ can also be bounded using Lemma E3. In that case we have

$$\begin{aligned} \|A_z^{(L)}\| &\le \sum_{l=0}^{L-1} \|T_z^L\| \|R_z\| \|Q^{L-1-l}\| \\ &\le \|R_z\| K_T K_Q \tau^l \kappa^{L-1-l} \le K_A \alpha^L, \end{aligned} \tag{E17}$$

with some constant $K_A > 0$ and $\alpha = \max\{\kappa, \tau\} < 1$. This can be obtained again using Gelfand's formula on both T_z and Q, and by subadditivity and submultiplicativity of the matrix norm $\| \cdot \|$. ∎

3. Proof of Theorem IV.1

It is important to note that the limiting value of $\mathbb{V}_{\rho,H}^L$ for layered, locally random channels as in Definition III.2 is obtained by studying the spectral properties of the LTM of the intermediate channel in the Heisenberg picture. In particular, the previous discussion suggests a limiting value for the variance of the form

$$\mathbb{V}_{\rho,H}^\infty = \sum_{z \,|\, \rho(T_z)=1} (\ell_\rho, w_z)(\ell_H)_z + (\ell_\rho, w_z)(A\ell_H)_z, \tag{E18}$$

with some normalized, strictly positive vector w_z, and an absorption matrix A given by the limit [53]

$$A = \lim_{L\to\infty} \sum_z A_z^{(L)}. \tag{E19}$$

Indeed, the following shows that this is the case and provides an analytical form for A.

Theorem E5 (Deep circuit variance). Let $\rho, H \in \mathcal{B}$ and let Φ_θ a be layered, locally random channel as in Definition III.2. Then the Cesàro average of $\mathbb{V}_{\rho,H}^L$ converges to Eq. (E18), and we have

$$\left| \frac{1}{L} \sum_{l=0}^{L} \mathbb{V}_{\rho,H}^l - \mathbb{V}_{\rho,H}^\infty \right| \in O\left(e^{-\beta L} \|H\|_2^2\right), \tag{E20}$$

for some constant $\beta > 0$. Additionally, if all essential blocks are aperiodic, then $\mathbb{V}_{\rho,H}^L$ is convergent, and we have

$$\left| \mathbb{V}_{\rho,H}^L - \mathbb{V}_{\rho,H}^\infty \right| \in O\left(e^{-\beta L} \|H\|_2^2\right), \tag{E21}$$

where the right eigenvector w_z of T_z is a strictly positive vector, i.e., $(w_z)_\kappa > 0 \,\forall \kappa$, and $A = R(\mathbb{1} - Q)^{-1}$ are the absorption coefficients of each essential block.

Proof. Thanks to Proposition E2, only irreducible components with $\rho(T_z) = 1$ will contribute to the limit, so we can restrict our analysis to those alone. Consider then an irreducible block T_z with unit spectral radius and of period d. While the full version of Perron-Frobenius theorem does not directly apply to T_z, it is a well-known result of nonnegative matrix theory [35,37] that the matrix T_z^d can be cast to a block diagonal form by a permutation, with irreducible and *aperiodic* blocks, for which we can apply it. However, it is crucial to notice that while $\lim_{N\to\infty} T_z^{dN} = T_z^{(d\infty)}$ exists, this does not imply that $\lim_{L\to\infty} T_z^L$ does. In fact, different subsequences might have different limiting values and, in particular, $\lim_{N\to\infty} T_z^{dN+m} = T_z^{(d\infty)} T^m$, which is different for all $m = 0, \ldots d-1$. In the periodic scenario then, T_z^L does not have a limit, and the only convergent quantity is the *Cesàro average*, i.e.,

$$P_z = \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} T_z^l = T_z^{(d\infty)} \frac{1}{d} \sum_{m=0}^{d-1} T_z^m, \tag{E22}$$

where $P_z = w_z v_z^t$ can be shown to be the Perron projector associated with T_z [35,37]. Despite being more cumbersome, a totally analogous approach allows determining the limiting values of $A^{(dN+m)}$ as well, as shown in the following proposition. ∎

Proposition E3. Given a LTM with a periodic irreducible block T_z of period d, then

$$\lim_{N\to\infty} A^{(dN+m)} = T^{(d\infty)} A^{(m)} + A^{(d\infty)} Q^m, \tag{E23}$$

where $A^{(d\infty)} = \sum_{m=0}^{d-1} T^{(d\infty)} A^{(d)} (\mathbb{1} - Q^d)^{-1}$ and $A^{(0)} = 0$.

Proof. Starting from the definition of $A^{(l)}$, we can first find the limiting value $A^{(d\infty)}$ of $A^{(dN)}$,

$$\begin{aligned} A^{(dN)} &= \sum_{l=0}^{Nd-1} T_z^l R_z Q^{dN-1-l} \\ &= \sum_{m=0}^{d-1}\sum_{n=0}^{N-1} T_z^{nd+m} R_z Q^{(N-1)d-nd+d-1-m} \\ &= \sum_{m=0}^{d-1}\sum_{n=0}^{N-1} T_z^{nd} T_z^m R_z Q^{d-1-m}(Q^d)^{N-n} \\ &= \sum_{n=0}^{N-1} T_z^{nd} A^{(d)}(Q^d)^{N-n} \\ &= T_z^{(d\infty)} A^{(d)} \sum_{n=0}^{N-1}(Q^d)^{N-n} + \sum_{n=0}^{N-1} \Delta_z^{(n)} A^{(d)}(Q^d),^{N-n}, \end{aligned} \tag{E24}$$

where $\Delta_z^{(N)} = T_z^{dN} - T_z^{(d\infty)}$. Since T_z^d is block diagonal, and each of the d blocks T_{zm} is irreducible and aperiodic, we have that $T_{zm}^N \to P_{zm}$ exponentially fast, i.e., $\|T_{zm}^N - P_{zm}\| = \|\Delta_{zm}^{(N)}\| < K_{zm}\tau_{zm}^N$, for some $K > 0$ and $\tau < 1$. Then, $\|\Delta_z^{(N)}\| \le \sum_{m=0}^d \|\Delta_{zm}^{(N)}\| \le K_z \tau_z^N$, where $\tau = \max_m\{\tau_{zm}\} < 1$. Together with Proposition E1, this implies that the last term in Eq. (E24) approaches zero with the same exponential speed, similarly to what happens in Proposition E2. Putting everything together, and considering that $\sum_{n=0}^{N-1} X^n \to (\mathbb{1} - X)^{-1}\ \forall X$ such that $\rho(X) < 1$, we have

$$A^{(d\infty)} = T^{(d\infty)} A^{(d)}(\mathbb{1} - Q^d)^{-1}. \tag{E25}$$

At this point, by induction similarly to Lemma E3, one can easily show that

$$A^{(dN+m)} = T^{(dN)} A^{(m)} + A^{(dN)} Q^m, \tag{E26}$$

which yields the proposition. ■

In particular, the Cesàro average converges and we have the expression

$$A_z = \lim_{L\to\infty} \frac{1}{L}\sum_{l=1}^{L} A^{(l)} = \frac{1}{d}\sum_{m=0}^{d-1} T^{(d\infty)} A^{(m)} + A^{(d\infty)} Q^m. \tag{E27}$$

Recalling that the right eigenvector v_z can be explicitly calculated when $\rho(T_z) = 1$ (see Corollary E1), this allows to obtain the final form of $\mathbb{V}_{\rho,H}^\infty$ by ordinary matrix vector multiplication. As a special case, if all relevant blocks T_z are aperiodic, then the limits of T_z^L and $A_z^{(L)}$ converge, and we have $P_z = T_z^{(1\infty)}$ and $A_z = A_z^{(1\infty)} = P_z R_z(\mathbb{1} - Q)^{-1}$. Finally, the exponential upper bound in Eqs. (E20) and (E21) is also obtained as a consequence of the preceding analysis. In particular, if we denote by T^∞ the matrix with P_z in place of T_z, A_z in place of R_z and zero otherwise, we have

$$\begin{aligned} &\left|\frac{1}{L}\sum_{l=0}^{L} \mathbb{V}_{\rho,H}^l - \mathbb{V}_{\rho,H}^\infty\right| \\ &\quad = \left|\left(\ell_\rho, \frac{1}{L}\sum_{l=0}^{L} T^l - T^\infty \ell_H\right)\right| \\ &\quad \le \left\|\frac{1}{L}\sum_{l=0}^{L} T^l - T^\infty\right\| \sqrt{(\ell_\rho,\ell_\rho)(\ell_H,\ell_H)}, \end{aligned} \tag{E28}$$

by Cauchy-Schwarz inequality. Since all blocks converge exponentially fast from the above discussion, the matrix norm is also exponentially decaying. Moreover, by construction $(\ell_A,\ell_A) \le \sqrt{\sum_\kappa (\ell_A)_\kappa^2} \le \sum_\kappa (\ell_A)_\kappa = \|A\|_2^2\ \forall A \in \mathcal{B}$, which concludes the proof.

While for an explicit calculation of w_z one should in general rely on case-specific analyses, a general result can be derived for a subclass of channels especially useful in the context of quantum computing, namely, single-qubit noise.

Proposition E4 (Single-qubit noise). Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition III.2. Assume moreover that the intermediate channel is of the form $\mathcal{E} = \mathcal{N}(W\rho W^\dagger)$, where W is a unitary transformation and $\mathcal{N} = \mathcal{N}_1 \otimes \cdots \otimes \mathcal{N}_n$ is a composition of single-qubit quantum channels. Then

$$\mathbb{V}_{\rho,H}^\infty = \sum_{z\,|\,\rho(T_z)=1} \frac{(\ell_\rho)_z(\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z(A\ell_H)_z}{d_z}. \tag{E29}$$

Proof. Without loss of generality, we can consider the single qubit channels $\mathcal{N}_m$ to be in their normal form of Lemma E2. In particular, this holds due to the invariance of the LTM with respect to changes of local bases (Lemma B2). In terms of the adjoint maps $\mathcal{N}_m^\dagger$, this condition reads $\mathcal{N}_m^\dagger(P_{j_m}) = t_{j_m}P_0 + \lambda_{j_m}P_j$ and can be used to compute $\mathrm{Tr}\left[\mathcal{N}_m^\dagger(P_{j_m})^2\right] = t_{j_m}^2 + \lambda_{j_m}^2 \le 1$ by Lemma E2. We now show that, when $P_j \in \mathcal{B}_z$ pertains to an irreducible component of spectral radius $\rho(T_z) = 1$, then the inequality must be saturated. In particular, thanks to Corollary E1 we know that $\sum_\kappa (T_z)_{\kappa,\lambda} = 1$. By Eq. (E15) this implies

$$\begin{aligned}\sum_\kappa (T_z)_{\kappa,\lambda} &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[(W^\dagger \mathcal{N}^\dagger(P_j)W)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[\mathcal{N}^\dagger(P_j)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_j \prod_{m=1}^{M} \mathrm{Tr}\left[\mathcal{N}_m^\dagger(P_{j_m})^2\right]\tilde{\delta}_{j,\lambda} = 1. \end{aligned} \tag{E30}$$

Since each term in the product is upper-bounded by 1, Eq. (E30) implies $\mathrm{Tr}\left[\mathcal{N}_m^\dagger(P_{j_m})^2\right] = t_{j_m}^2 + \lambda_{j_m}^2 = 1\,\forall j_m$. This condition is only compatible with Lemma E2 if $t_{j_m}^2 = 0$ and $\lambda_{j_m}^2 = 1$. This result can now be used to show that the adjoint of $T_z^\dagger$ of the LTM T_z must also be column stochastic. Indeed, if we consider the expansion of $WP_jW^\dagger$ with respect to the normalized Pauli basis, we can get

$$\begin{aligned}\sum_\kappa (T_z)^\dagger_{\kappa,\lambda} &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[\mathcal{N}(WP_jW^\dagger)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[\left(\sum_i \mathrm{Tr}\left[P_iWP_jW^\dagger\right]\mathcal{N}(P_i)\right)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[\left(\sum_i \mathrm{Tr}\left[P_iWP_jW^\dagger\right]\prod_m \lambda_{i_m}P_i\right)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_{i,j}\prod_m \lambda_{i_m}^2 \mathrm{Tr}\left[P_iWP_jW^\dagger\right]^2\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[WP_j^2W^\dagger\right]\tilde{\delta}_{j,\lambda} = 1. \end{aligned} \tag{E31}$$

This allows to compute the right eigenvector w_z of the leading eigenvalue of T_z. Using Eq. (B7), we have indeed

$$\sum_\lambda (T_z)_{\kappa,\lambda}d_\lambda = \sum_\lambda (T_z)^\dagger_{\lambda,\kappa}d_\kappa = d_\kappa, \tag{E32}$$

which implies $(w_z)_\lambda = d_\lambda/d_z$, where the normalization factor $d_z = \sum_\lambda d_\lambda$ is necessary to ensure $w_z v_z^t = P_z$ is indeed a projection, thus concluding the proof. ■

As a consequence of this last result, we can compute the variance of generic unitary circuits.

Corollary E2 (Unitary circuits). Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition III.2, with unitary intermediate channel $\mathcal{E}(\rho) = W\rho W^\dagger$. Then we have

$$\mathbb{V}^\infty_{\rho,H} = \sum_{z>0}\frac{(\ell_\rho)_z(\ell_H)_z}{d_z}. \tag{E33}$$

Proof. This form of the variance is a special case of Proposition E4, putting $\mathcal{N}(\rho) = \rho$, and noting that the absorption terms must vanish. In particular, this follows from Eq. (E15), observing that unitary channels saturate Kadison-Schwarz inequality, which combined with Corollary E1 imply $Q = 0$ and consequently $R = 0$ and $A = 0$. ■

A similar computation is also possible for global noise channels, like correlated Pauli noise. In particular, $(w_z)_\kappa$ can be analytically computed the intermediate if channel $\mathcal{E}$ is of the form $\mathcal{E}(\rho) = \mathcal{N}(W\rho W^\dagger)$, with $\mathcal{N}(\rho)$ a *diagonal* Pauli channel. More precisely, we have the following proposition.

Proposition E5 (Global Pauli noise map). Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition III.2. Assume moreover that the intermediate channel is of the form $\mathcal{E}(\rho) = \mathcal{N}(W\rho W^\dagger)$, where $\mathcal{N}(\rho) = \sum_k p_k\bar{P}_k\rho\bar{P}_k$, with $p_k \geq 0$, $\sum_k p_k = 1$ and $\bar{P}_k$ denote non-normalized Pauli string, i.e., satisfying $\bar{P}_k^2 = \mathbb{1}$. Then

$$\mathbb{V}^\infty_{\rho,H} = \sum_{z\,|\,\rho(T_z)=1}\frac{(\ell_\rho)_z(\ell_H)_z}{d_z}. \tag{E34}$$

In particular, $\rho(T_z) = 1$ if and only if $[\bar{P}_k, P_j] = 0\ \forall P_j \in \mathcal{B}_z\ \forall \bar{P}_k | p_k > 0$.

Proof. The proof follows the same path as Proposition E4. In particular, from the definition of the intermediate channel, it follows that $\mathcal{N}$ is unital. This implies that $\mathrm{Tr}\left[\mathcal{N}(P_j)^2\right] \leq \mathrm{Tr}\left[\mathcal{N}(P_j^2)\right] = 1\,\forall j$ by Kadison-Schwarz inequality. More specifically, we can compute the action of $\mathcal{N}$ explicitly on (normalized) Pauli strings as follows:

$$\mathcal{N}(P_j) = \sum_k p_k\bar{P}_kP_j\bar{P}_k = \sum_k p_k s_{k,j}P_j = \mathcal{N}^\dagger(P_j), \tag{E35}$$

where $s_{k,j} = 1$ if $\bar{P}_k$ and P_j commute, and $s_{i,j} = -1$ otherwise. From this it follows that

$$\mathcal{N}(P_j)^2 = \mathcal{N}^\dagger(P_j)^2 = \left(\sum_k p_k s_{k,j}\right)^2\frac{\mathbb{1}}{d}. \tag{E36}$$

Combining it to Corollary E1 and Eq. (E30), we get that $\rho(T_z) = 1 \Leftrightarrow \sum_\kappa (T_z)_{\kappa,\lambda} = 1 \Leftrightarrow \sum_k p_k s_{k,j} = 1$, i.e., $s_{k,j} = 1\forall k|p_k > 0\,\forall j \in T_z$.

Let's denote for convenience by c_j the quantity $c_j = \sum_k p_k s_{k,j}$. The previous discussion implies $|c_j| \leq 1\,\forall j$, with equality holding for all elements of $\mathcal{B}_z$. Then we have

$$\begin{aligned}\sum_\kappa (T_z)^\dagger_{\kappa,\lambda} &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[\mathcal{N}(WP_jW^\dagger)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[\left(\sum_i \mathrm{Tr}\left[P_iWP_jW^\dagger\right]\mathcal{N}(P_i)\right)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[\left(\sum_i \mathrm{Tr}\left[P_iWP_jW^\dagger\right]c_iP_i\right)^2\right]\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_{i,j} c_i^2\mathrm{Tr}\left[P_iWP_jW^\dagger\right]^2\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left[WP_j^2W^\dagger\right]\tilde{\delta}_{j,\lambda} = 1.\end{aligned} \tag{E37}$$

By Eq. (E32), we get again $(w_z)_\lambda = d_\lambda/d_z$.

As a last step, we show how the absorption coefficients must vanish in this case. This is a consequence of the fact that $|c_j| \le 1$. Indeed, by computing a generic element $T_{\kappa,\lambda}$, we get

$$\begin{aligned}T_{\kappa,\lambda} &= \frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\left[P_iW^\dagger\mathcal{N}(P_j)W\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda}\\ &= \frac{1}{d_\lambda}\sum_{i,j}c_j^2\mathrm{Tr}\left[P_iW^\dagger P_jW\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda}\\ &\le \frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\left[P_iW^\dagger P_jW\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda} = T^W_{\kappa,\lambda},\end{aligned} \tag{E38}$$

where T^W is the LTM of the unitary channe defined by the unitary W alone. Since by Corollary E2 Q (and consequently R) vanish for unitary channels, then R must also vanish here, implying $A = 0$. ■

On the opposite limit, if the noise map is strictly contractive in at least one direction in each $\mathcal{B}_z$, then the combination of noise and entanglement is strong enough to kill the variance in each of the absorbing subspaces. As a consequence, only the absorption term to $\mathcal{B}_0$ remains, since no channel can be contractive thereby trace preservation.

Corollary E3 (Noise-induced concentration). Let ρ, $H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition III.2, with intermediate quantum channel $\mathcal{E}$, and let $\{P_j\}$ denote the normalized Pauli basis. If $\|\mathcal{E}^\dagger(P_j)\|_2 < 1$ for some $j \in T_z$, $\forall z$, then

$$\mathbb{V}^\infty_{\rho,H} = \frac{(A\ell_H)_0}{d}. \tag{E39}$$

In particular, if the channel is unital, $\mathbb{V}^\infty_{\rho,H} = 0$.

Proof. This form of the variance is a special case of Eq. (E18), where all absorbing components T_z, $\rho(T_z) = 1$ vanish when $z > 0$. In particular, this follows from Eq. (E15), observing that the above condition implies that for $\mathcal{E}^\dagger$ Kadison-Schwarz inequality is strict, which combined with Corollary E1 imply $T_z = 0\,\forall z > 0$, leaving only T_0. Since terms T_z, $\rho(T_z) < 1$ do not contribute to the variance in the limit, the only remaining term is the absorption to T_0, namely, $(\ell_\rho)_0(A\ell_H)_0$. Finally, the corollary follows from the normalization condition $\mathrm{Tr}\,[\rho] = 1$ on ρ, which ensures $(\ell_\rho)_0 = 1/d$. ■

APPENDIX F: VARIANCE LOWER BOUNDS

In this appendix we prove the general variance lower bounds outlined in Sec. IV.

1. Proof of Theorem IV.2

Here we employ Proposition IV.1 to prove a general lower-bound on slowly entangling circuits. In particular, such result is based on the approximation $T_l \approx \mathbb{1}$, which holds either for shallow circuits, i.e., $L \in O(\log n)$, or deeper circuits, but with weakly entangling intermediate channels. The discussion is based on the following result.

Theorem F1. Let $\rho, H \in \mathcal{B}$ and consider a sequence of quantum channels $\{\mathcal{E}_l\}_{l=1}^L$, and let $\{T_l\}_{l=1}^L$ be the respective LTMs. Finally let $K \subset \{0,1\}^M$ denote a subset of indices, and by $\alpha_l = \min_{\kappa\in K}(T_l)_{\kappa,\kappa}$. Then

$$\mathbb{V}^L_{\rho,H} \ge \alpha^L(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{F1}$$

where $\mathcal{K}(\cdot) = \sum_{\kappa\in K}\sum_j \mathrm{Tr}\left[P_j\ \cdot\right]P_j\tilde{\delta}_{j,\kappa}$ is a projector onto the space spanned by K and $\alpha = \left(\prod_{l=0}^L \alpha_l\right)^{1/L}$ is the geometric mean of α_l.

Proof. Consider a single circuit layer. Then, we can write

$$\begin{aligned}(\ell_\rho, T_l\ell_H) &= \sum_{\kappa,\lambda}\frac{(\ell_\rho)_\kappa T_{\kappa,\lambda}(\ell_H)_\lambda}{d_\kappa}\\ &\ge \sum_{\kappa\in K}\frac{(\ell_\rho)_\kappa T_{\kappa,\kappa}(\ell_H)_\kappa}{d_\kappa}\\ &\ge \alpha_l\sum_{\kappa\in K}\frac{(\ell_\rho)_\kappa(\ell_H)_\kappa}{d_\kappa} = \alpha_l(\ell_\rho, \ell_{\mathcal{K}(H)}),\end{aligned} \tag{F2}$$

where the inequality holds since all terms in the sum are non-negative by construction. The claim follows from repeated application of the latter. ■

Despite its simplicity, Theorem F1 can be used to deduce general bounds on weakly entangling circuits, which are the foundation of small-angle initialization strategies. In particular, we get the following corollary.

Corollary F1. Let $\mathcal{H}=\bigotimes_{m=1}^{M}\mathcal{H}_m, d_m\in\Theta(1)$. If either of the conditions

(a) $\alpha>0, \alpha\in\Omega(1)$ and $L\in O(\log n)$,
(b) $\alpha=1-f(n,L)$, $f\in O(\log n/L)$ and $L\in\Omega(\log^{1+\epsilon} n)$ for some arbitrary $\epsilon>0$

is satisfied, then

$$\mathbb{V}^{L}_{\rho,H}\geq F(n)(\ell_\rho,\ell_{\mathcal{K}(H)}), \tag{F3}$$

where $F(n)\in\Omega(1/\mathrm{poly}(n))$.

Proof. Exploiting Theorem F1, it suffices to show that $F(n)=\alpha^L\in\Omega(1/\mathrm{poly}(n))$. In the first case, this follows directly from the shallow nature of the circuit and, in particular, $F(n)\in\alpha^{O(\log(n))}=\Omega(1/n^{-\log(\alpha)})\subset\Omega(1/\mathrm{poly}(n))$. For the second case instead, it is useful to consider $\log(F(n))$,

$$\begin{aligned}-\log(F(n))&<-L\log\left(1-C\frac{\log(n)}{L}\right)\\&=C\log(n)\left(1+C\frac{\log(n)}{2L}+O\left(\frac{\log^2(n)}{L^2}\right)\right)\in O(\log(n)),\end{aligned} \tag{F4}$$

which in turn implies $F(n)\in e^{-O(\log(n))}=\Omega(1/n^C)\subset\Omega(1/\mathrm{poly}(n))$. ■

2. Two applications of Corollary IV.2

In this appendix, we showcase a general technique that uses Corollary IV.2 to prove the absence of concentration in both an instance of FLDC given by Fig. 7 and shallow brickwork circuits depicted in Fig. 4.

Given a n qubit circuit, consider a partition of the full Hilbert space $\mathcal{H}$ into the tensor product of $M=n/2$ local subspaces $\mathcal{H}_m$, each of dimension $d_m=4$, and comprising a qubit pair. In this setting, consider a two-local observable H, with support over the first two qubits. Since the support of H matches the first local subspace $\mathcal{H}_{m=1}$, the locality vector of H is given by definition,

$$(\ell_H)_\kappa=\delta_{\kappa,(100\cdots0)}\|H\|_2^2, \tag{F5}$$

where the notation $(100\cdots0)$ indicates a M-digit long string, where only the first is set to one. With reference to Theorem IV.2, we set $K=\{(100\cdots0)\}$, as in this way $\mathcal{K}(H)=H$.

Thanks to this choice, we are now ready to explicitly compute α_l, which is given by

$$\alpha_l=\min_{\kappa\in K}(T_l)_{\kappa,\kappa}=(T_l)_{(100\cdots0),(100\cdots0)}. \tag{F6}$$

In the following we will use Eq. (F6) to compute the scaling of the geometric mean α.

a. Finite local depth circuit

As shown in Fig. 7, the chosen class of FLDCs admits a particularly simple decomposition in terms of Eq. (2) involving two-qubit local subspaces and an array of SWAP gates composed of only one single gate per layer, easing calculations.

Let's denote by $\mathcal{S}_l$ the intermediate channel corresponding to the lth layer in Fig. 7 (right). Using Eq. (B4), α_l can be explicitly rewritten as

$$\alpha_l=\frac{1}{15}\sum_{i,j}\mathrm{Tr}\left[P_i\mathcal{S}_l^\dagger(P_j)\right]^2\tilde{\delta}_{i,(100\cdots0)}\tilde{\delta}_{j,(100\cdots0)}, \tag{F7}$$

where the $\tilde{\delta}$ are nonvanishing only when the Pauli strings P_i and P_j have support on the first two qubits.

We now consider the cases $l=1$ and $l=2$. Since the SWAP gate in these cases involves the first and third qubits, whenever P_j acts nontrivially on the first qubit, the operator $\mathcal{S}_l^\dagger(P_j)$ will act nontrivially on the third qubit. As

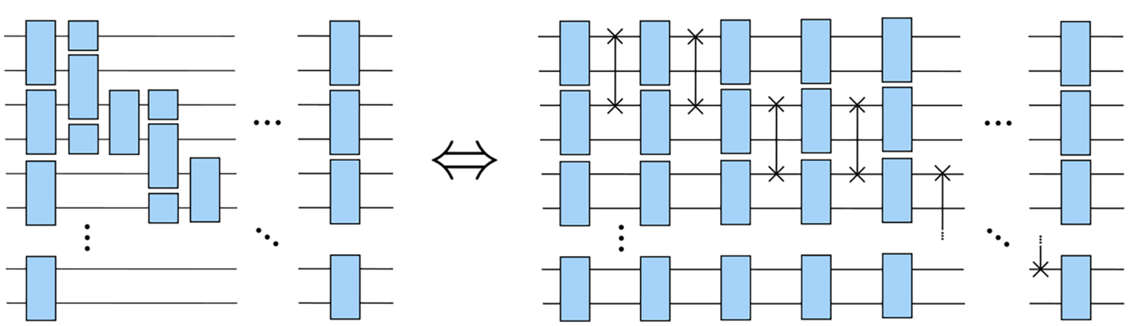

FIG. 7. Example of a deep finite local-depth circuit. Each blue block represents a local unitary two-design over pairs of qubits. This can be reduced to Eq. (2) employing a SWAP entangling layer and by noticing that since they form unitary designs, the composition of two subsequent, equal blocks is equivalent to a single block.

a result, it is orthogonal to all P_i appearing in the sum, and its contribution vanishes. Conversely, if P_j acts trivially on the first qubit, then $\mathcal{S}_l^\dagger(P_j) = P_j$, and therefore $\mathrm{Tr}\left[P_i \mathcal{S}l^\dagger(P_j)\right] = \delta_{i,j}$. There are exactly three such operators, namely, $\mathbb{1}X$, $\mathbb{1}Y$, and $\mathbb{1}Z$. Consequently, the sum yields $\alpha_1 = \alpha_2 = 1/5 \in \Theta(1)$. Finally, for $l > 2$, the map $\mathcal{S}_l^\dagger$ acts trivially on the first two qubits. It follows that $\mathrm{Tr}\left[P_i \mathcal{S}l^\dagger(P_j)\right] = \delta_{i,j}$ for all i,j, and hence $\alpha_l = 1$.

Computing the geometric mean, and expanding in the limit of large L, yields

$$\alpha = \left(\prod_{l=1}^{L} \alpha_l\right)^{\frac{1}{L}} = \left(\frac{1}{5}\right)^{\frac{2}{L}} \approx 1 - \frac{2\log(5)}{L}, \tag{F8}$$

which satisfies the assumptions of Corollary IV.2(b) for deep enough circuits, giving the lower bound

$$\mathbb{V}_{\rho,H}^L \geq F(n)(\ell_\rho, \ell_H),\ F(n) \in \Omega(1/\mathrm{poly}(n)). \tag{F9}$$

If we further specify ρ and H, for instance, $\rho = |0\rangle\langle 0|$ and $H = XX + YZ + ZX$, then $(\ell_\rho, \ell_H) = 1/5$, giving the final bound $\mathbb{V}_{\rho,H}^L \in \Omega(1/\mathrm{poly}(n))$.

Using an analogous proof, the same result can be extended to a wider class of FLDCs, observables, and initial states.

b. Shallow brickwork circuit

Similarly to the previous case, the brickwork circuit admits a particularly simple decomposition, this time involving a cascade of SWAP gates, as explicitly shown in Fig. 4 (right). We denote by $\mathcal{S}$ the intermediate channel depicted in the figure.

For completeness, we now assume that after each application of $\mathcal{S}$, a single layer of unital, single-qubit noise $\mathcal{N} = \bigotimes_{k=1}^n \mathcal{N}_k$ is applied to the system. The resulting intermediate channel therefore reads $\mathcal{E} = \mathcal{N} \circ \mathcal{S}$. Using Eq. (B4), the coefficients α_l can be written explicitly as

$$\alpha_l = \frac{1}{15}\sum_{i,j} \mathrm{Tr}\left[P_i \mathcal{E}^\dagger(P_j)\right]^2 \tilde{\delta}_{i,(100\cdots 0)}\tilde{\delta}_{j,(100\cdots 0)}, \tag{F10}$$

where, as before, the $\tilde{\delta}$ terms are nonvanishing only when the Pauli strings P_i and P_j have support on the first two qubits. By the invariance of the LTM under local basis transformations, we may assume without loss of generality that each single-qubit noise channel $\mathcal{N}_k$ is expressed in its normal form. Moreover, by Lemma E2, the unitality condition $\mathcal{N}_k(\mathbb{1}) = \mathbb{1}$ implies that, for a Pauli string

$$P_j = \bigotimes_{k=1}^n P_j^{(k)}, \tag{F11}$$

one has $\mathcal{N}_k^\dagger(P_j^{(k)}) = \lambda_{i,k} P_j^{(k)}$, with $\lambda_{j,k} \in \Theta(1)$. Combining these observations yields

$$\begin{aligned}\mathrm{Tr}\left[P_i \mathcal{E}^\dagger(P_j)\right] &= \mathrm{Tr}\left[P_i \mathcal{S}^\dagger(\mathcal{N}^\dagger(P_j))\right] \\ &= \prod_{k=1}^n \lambda_{j,k} \mathrm{Tr}\left[P_i \mathcal{S}^\dagger(P_j)\right].\end{aligned} \tag{F12}$$

Since we only consider Pauli strings P_j with support on the first two qubits, and since $\mathcal{N}_k^\dagger(\mathbb{1}) = \mathbb{1}$ by construction, Eq. (F12) simplifies in this case to the factor $\lambda_{j,1}\lambda_{j,2}$.

Analogously to the previous section, it is easy to verify that concerning $\mathcal{S}$, the same three terms of above are nonvanishing in the sum of Eq. (F12), namely, $P_j = \mathbb{1}X, \mathbb{1}Y, \mathbb{1}Z$. Moreover, in these cases only the second Pauli is nontrivial, further simplifying the product in Eq. (F12) to $\lambda_{j,2}$. Putting everything together, we get

$$\alpha_l = \frac{\lambda_{\mathbb{1}X,2}^2 + \lambda_{\mathbb{1}Y,2}^2 + \lambda_{\mathbb{1}Z,2}^2}{15} > 0. \tag{F13}$$

Since the α_l does not depend on l, we have

$$\alpha = \left(\prod_{l=1}^{L} \alpha_l\right)^{\frac{1}{L}} = \alpha_l \in \Omega(1) \tag{F14}$$

and thus α satisfies Corollary IV.2(a), giving the lower bound

$$\mathbb{V}_{\rho,H}^L \geq F(n)(\ell_\rho, \ell_H),\ F(n) \in \Omega(1/\mathrm{poly}(n)), \tag{F15}$$

for shallow enough circuits, i.e., $L \in O(\log n)$.

If we further specify ρ and H, for instance, $\rho = |0\rangle\langle 0|$ and $H = XX + YZ + ZX$ as before, then $(\ell_\rho, \ell_H) = 1/5$, giving again the final bound $\mathbb{V}_{\rho,H}^L \in \Omega(1/\mathrm{poly}(n))$.

Using an analogous proof, the same result can be extended to a wider class of noisy shallow circuits, observables, and initial states.

3. Small-angle initializations lower bounds

In order to prove the general lower bounds on small-angle initializations provided in proposition Eq. (19), it is useful to start from the following proposition, as it is the fundamental building block in this type of proof.

Proposition F1. Let $\{E_\phi\}_\phi$ be an ensemble such that $\mathcal{E}(\rho) = \mathbb{E}_\phi\{E_\phi \rho E_\phi^\dagger\}$ is a quantum channel. Further, denote by T_ϕ the transfer matrix associated with each $\mathcal{E}_\phi^\dagger(\cdot) = E_\phi^\dagger \cdot E_\phi$, and by T the transfer matrix of $\mathcal{E}^\dagger$. Then we have

$$\mathbb{E}_\phi\{T_\phi\} \geq T, \tag{F16}$$

with equality holding if and only if $\mathcal{E}$ is unitary.

Proof. Let A, B be arbitrary bounded operators and consider

$$\begin{aligned}\mathrm{Tr}\left[\mathcal{E}(A)B\right]^2 &= \mathrm{Tr}\left[A\mathcal{E}^\dagger(B)\right]^2 \\ &= \mathbb{E}_\phi\left\{\mathrm{Tr}\left[A\mathcal{E}_\phi^\dagger(B)\right]\right\}^2 \\ &\leq \mathbb{E}_\phi\left\{\mathrm{Tr}\left[A\mathcal{E}_\phi^\dagger(B)\right]^2\right\}\ \forall A, B,\end{aligned} \tag{F17}$$

which follows from the observation that $f(\phi) = \mathrm{Tr}\left[A\mathcal{E}_\phi^\dagger(B)\right] \in \mathbb{R}$ and so $\mathbb{V}_\phi\{f\} \geq 0$. Applying this to the entries of T and T_ϕ gives the general inequality. Finally, the equality follows from $\mathbb{V}_\phi\{f\} = 0$, which means that $f(\phi) = \mathrm{Tr}\left[AK^\dagger BK\right]$ is a constant, where $K = E_\phi\ \forall\phi$. Hence, since $\mathcal{E}(\cdot) = K \cdot K^\dagger$ is CPTP, it must also be unitary. ■

In order to translate this rather abstract formulation into a practical recipe, we need to identify the conditions that allow to treat the contribution of given an ensemble $\{\mathcal{E}_\phi\}$ of parameterized intermediate channels to the variance in terms of the mean LTM $\mathbb{E}_\phi\left\{T_\phi\right\}$. In particular, it is easily verified that, if ϕ is sampled independently of the other parameters, then

$$\begin{aligned}\mathbb{V}^L_{\rho,H} &= \mathbb{E}_\phi\mathbb{E}_\theta\left\{\mathrm{Tr}\left[\Phi_{\theta,\phi}(\rho)H\right]^2\right\} \\ &= \mathbb{E}_\phi\left\{(\ell_\rho, T_\phi\ell_H)\right\} = (\ell_\rho, \mathbb{E}_\phi\{T_\phi\}\ell_H).\end{aligned} \tag{F18}$$

Combining this observation with Corollary F1, we can get the QResNet lower bound of Proposition IV.2. First, a technical Lemma.

Lemma F1. Let $A \in \mathcal{B}$, and P_i be a normalized Pauli string, i.e., $\mathrm{Tr}\left[P_i^2\right] = 1$. Then $\mathrm{Tr}\left[P_iA^\dagger P_iA\right] \in \mathbb{R}$ and

$$|\mathrm{Tr}\left[P_iA^\dagger P_iA\right]| \leq \mathrm{Tr}\left[A^\dagger AP_i^2\right] = \frac{\|A\|_2^2}{d}. \tag{F19}$$

Proof. Consider the decomposition of A into the normalized Pauli basis $A = \sum_j a_jP_j$. The coefficients $\{a_j\}_j$ are related to the 2-norm of A by the following relation:

$$\begin{aligned}\|A\|_2^2 = \mathrm{Tr}\left[A^\dagger A\right] &= \sum_{j,k} a_j^* a_k \mathrm{Tr}\left[P_jP_k\right] \\ &= \sum_{j,k} a_j^* a_k \delta_{j,k} = \sum_j |a_j|^2.\end{aligned} \tag{F20}$$

Using this decomposition, we get

$$\begin{aligned}\mathrm{Tr}\left[P_iA^\dagger P_iA\right] &= \sum_{j,k} a_j^* a_k \mathrm{Tr}\left[P_iP_jP_iP_k\right] \\ &= \sum_{j,k} a_j^* a_k s_{i,j}\frac{\mathrm{Tr}\left[P_jP_k\right]}{d} \\ &= \sum_{j,k} a_j^* a_k s_{i,j}\frac{\delta_{j,k}}{d} = \frac{1}{d}\sum_j |a_j|^2 s_{i,j},\end{aligned} \tag{F21}$$

where $s_{i,j} = 1$ if P_i and P_j commute, and $s_{i,j} = -1$ otherwise. Since $|s_{i,j}| = 1$, we get

$$\begin{aligned}|\mathrm{Tr}\left[P_iA^\dagger P_iA\right]| &= \left|\frac{1}{d}\sum_j |a_j|^2 s_{i,j}\right| \\ &\leq \frac{1}{d}\sum_j |a_j|^2|s_{i,j}| = \frac{\|A\|_2^2}{d}.\end{aligned} \tag{F22}$$

The equality comes directly from the fact that $P_i^2 = \mathbb{1}/d$ by the normalization constraint. ■

Proposition F2 (QResNet). Let $\mathcal{E}_\phi(\cdot) = e^{i\phi G} \cdot e^{-i\phi G}$ be a unitary entangling gate, and let μ, σ^2 be the mean and variance of the initialization distribution $\mathcal{P}$ of ϕ. Then, if $\mu = 0$, $\sigma^2 \in O(\log n/\|G\|_2^2 L)$, and $L \in \Omega(\log^{1+\epsilon} n)$

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_H), \tag{F23}$$

where $F(n) \in \Omega(1/\mathrm{poly}(n))$.

Proof. Let T be the locality transfer matrix of $\mathcal{E} = \mathbb{E}_\phi\{\mathcal{E}_\phi\}$ and consider the diagonal element $T_{\kappa,\kappa}$. Then, by definition, we have

$$\begin{aligned}T_{\kappa,\kappa} &= \frac{1}{d_\kappa}\sum_{i,j}\mathrm{Tr}\left[\mathbb{E}_\phi\{P_i\mathcal{E}_\phi^\dagger(P_j)\}\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\kappa} \\ &\geq \frac{1}{d_\kappa}\sum_i \mathrm{Tr}\left[\mathbb{E}_\phi\{P_i\mathcal{E}_\phi^\dagger(P_i)\}\right]^2\tilde{\delta}_{i,\kappa} \\ &= \frac{1}{d_\kappa}\sum_i \mathrm{Tr}\left[\mathbb{E}_\phi\{P_ie^{-i\phi G}P_ie^{i\phi G}\}\right]^2\tilde{\delta}_{i,\kappa}.\end{aligned} \tag{F24}$$

Since $\sigma^2 \to 0$ as $n \to \infty$, to find the asymptotic behavior of the diagonal elements of T we can expand $e^{i\phi G}$ around μ

and obtain $e^{i\phi G} = \mathbb{1} + i\phi G - \phi^2 G^2/2 + O(\phi^3\|G\|_2^3)$. Substituting this into Eq. (F24), we get

$$\begin{aligned} T_{\kappa,\kappa} &\approx \frac{1}{d_\kappa}\sum_i \mathrm{Tr}\Big[\mathbb{E}_\phi\{P_i(\mathbb{1} + i\phi G - \phi^2G^2/2)P_i(\mathbb{1}+ \\ &\quad + i\phi G - \phi^2 G^2/2)\}\Big]^2 \tilde{\delta}_{i,\kappa} \\ &= \frac{1}{d_\kappa}\sum_i \left(1 - \mathbb{E}_\phi\{\phi^2\}\left(\mathrm{Tr}\left[G^2P_i^2\right] - \mathrm{Tr}\left[P_iGP_iG\right]\right)\right)^2 \tilde{\delta}_{i,\kappa} \\ &\geq 1 - 4\|G\|_2^2\sigma^2 \,\forall\kappa \in \{0,1\}^M. \end{aligned} \tag{F25}$$

Exploiting Corollary F1, we get the proposition by showing $T_{\kappa,\kappa} \geq 1 - f(n,L)$, where $f(n,L) \in O(\log n/L)$. In particular, this follows directly from the scaling of σ^2. The same proof ensures the absence of concentration on a unitary QResNet, provided that $\mathcal{P}$ is chosen as an initialization probability by Proposition F1. ■

Going beyond unitary ansätze, a similar technique—combined with the stochastic representation of solutions to the Lindblad equation [39,40]—can be employed to construct a BP-free QResNet from any noise model defined by a set of jump operators L_k.

Proposition F3 (General QResNets from a quantum channel). Let $\mathcal{E}(\rho) = e^{t\mathcal{L}}(\rho)$ be a quantum channel generated by the Lindbladian

$$\mathcal{L}(\rho) = \sum_k \gamma_k\left(L_k\rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\}\right), \tag{F26}$$

with $\|L_k\|_2^2/d = 1$. Then, when $L \in \Omega(\log^{1+\epsilon} n)$, the ensemble generated by $\mathcal{E}_\phi = E_\phi \cdot E_\phi^\dagger$, $E_\phi = \prod_k e^{i\phi_k L_k + \sigma_k^2(L_k^2 - L_k^\dagger L_k)/2}$, and ϕ_k normally distributed with $\mu = 0, \sigma_k^2 \in O\left((\gamma_k/\sum_{k'}\gamma_{k'})(\log n/L)\right) \forall k$ satisfies

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_H), \tag{F27}$$

where $F(n) \in \Omega(1/\mathrm{poly}(n))$.

Proof. We prove this statement, and we compute the scaling of α in Corollary F1. Again, consider the diagonal elements $T_{\kappa,\kappa}$.

$$\begin{aligned} T_{\kappa,\kappa} &= \frac{1}{d_\kappa}\sum_{i,j}\mathrm{Tr}\left[P_ie^{t\mathcal{L}}(P_j)\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\kappa} \\ &\geq \frac{1}{d_\kappa}\sum_i \mathrm{Tr}\left[P_ie^{t\mathcal{L}}(P_i)\right]^2\tilde{\delta}_{i,\kappa}. \end{aligned} \tag{F28}$$

Expanding for $t \ll 1$, we get $e^{t\mathcal{L}}(P_i) \approx P_i + t\mathcal{L}(P_i)$, which can be used to lower bound $T_{\kappa,\kappa}$,

$$\begin{aligned} \mathrm{Tr}\left[P_ie^{t\mathcal{L}}(P_i)\right] &\approx \mathrm{Tr}\left[P_i^2\right] + t\mathrm{Tr}\left[P_i\mathcal{L}(P_i)\right] \\ &= 1 + \sum_k t\gamma_k\left(\mathrm{Tr}\left[P_iL_kP_iL_k^\dagger\right] - \frac{1}{2}\mathrm{Tr}\left[P_i\{L_k^\dagger L_k, P_i\}\right]\right) \\ &= 1 + \sum_k t\gamma_k\left(\mathrm{Tr}\left[P_iL_kP_iL_k^\dagger\right] - \mathrm{Tr}\left[L_k^\dagger L_kP_i^2\right]\right) \\ &\geq 1 - \sum_k t\gamma_k\left(|\mathrm{Tr}\left[P_iL_kP_iL_k^\dagger\right]| + \mathrm{Tr}\left[L_k^\dagger L_kP_i^2\right]\right) \\ &\geq 1 - 2\sum_k t\gamma_k\|L_k\|_2^2/d = 1 - 2t\sum_k\gamma_k. \end{aligned} \tag{F29}$$

Putting it back into the definition of $T_{\kappa,\kappa}$ we get

$$\begin{aligned} T_{\kappa,\kappa} &\geq \frac{1}{d_\kappa}\sum_i\left(1 - 2t\sum_k\gamma_k\right)^2\tilde{\delta}_{i,\kappa} \\ &\geq 1 - 4t\sum_k\gamma_k. \end{aligned} \tag{F30}$$

This directly implies that $\alpha \geq 1 - 4t\sum_k\gamma_k$ for small t, which satisfies Corollary F1, if $t \in O\left((1/\sum_{k'}\gamma_{k'})(\log n/L)\right)$. Finally, the claim follows by Corollary F1 unraveling the master equation in the Stratonovich formalism. In particular, we can describe the evolution according to $\mathcal{L}$ by the (stochastic) differential equation,

$$\frac{d\left|\psi\right\rangle}{ds} = \sum_k\left[i\sqrt{\gamma_k}L_k\xi_k(s) + \frac{\gamma_k}{2}\left(L_k^2 - L_k^\dagger L_k\right)\right]\left|\psi\right\rangle, \tag{F31}$$

where $\xi_k(s)$ represents a *white noise*. Equation (F31) is formally solved by

$$\left|\psi\right\rangle_t = \mathcal{T}\left\{e^{\int_0^t\sum_k\left[i\sqrt{\gamma_k}L_k\xi_k(s) + \frac{\gamma_k}{2}\left(L_k^2 - L_k^\dagger L_k\right)\right]ds}\right\}\left|\psi\right\rangle_0, \tag{F32}$$

where $\mathcal{T}\{\cdot\}$ denotes the time ordering operator. Since $t \to 0$ as n diverges, we can again expand this solution to first order in t to get the asymptotic scaling. In particular, this allows us to discard the time ordering, and, exploiting the defining relation of the white noise in terms of the Weiner process $\int_0^t \xi_k(s)ds = x_k$, where $x_k \sim \mathcal{N}(0,t)\forall k$ is normally distributed, we get, for each x, $\left|\psi\right\rangle_{x,t} = \bar{E}_x\left|\psi\right\rangle_0$, where

$$\begin{aligned} \bar{E}_x &= e^{\sum_k\left[i\sqrt{\gamma_k}L_kx_k + \frac{t\gamma_k}{2}\left(L_k^2 - L_k^\dagger L_k\right)\right]} \\ &\approx \prod_k e^{i\sqrt{\gamma_k}L_kx_k + \frac{t\gamma_k}{2}\left(L_k^2 - L_k^\dagger L_k\right)}. \end{aligned} \tag{F33}$$

This gives an approximate solution $e^{t\mathcal{L}} \approx \mathbb{E}_x\left[\bar{E}_x \rho \bar{E}_x^\dagger\right]$ Finally, we define $\phi_k = x_k\sqrt{\gamma_k}$, which is also normally distributed but with variance $\sigma_k^2 = t\gamma_k \in O\left(\frac{\gamma_k}{\sum_{k'}\gamma_{k'}}(\log n/L)\right)$. The proof is completed by making the substitution $E_\phi = \bar{E}_{x_k\sqrt{\gamma_k}}$. ■

APPENDIX G: NIBP-FREE LANDSCAPES UNDER UNITAL NOISE

To illustrate the emergence of nontrivial behavior in the loss variance $\mathbb{V}_{\rho,H}$ under unital noise channels, we focus on two representative examples, i.e., commuting noise and unitary maps, and noise-induced symmetry breaking. Numerical validation of both involves the same small quantum circuit composed of $n = 2$ qubits and correlated noise. In these cases, the presence of nonunitary layers is either irrelevant or beneficial to the overall magnitude of the variance.

1. Commuting unitary and noise maps

Here, we provide a numerical example of the commuting unitary and noise case discussed in Sec. II. Consider the circuit depicted in Fig. 8, which features a unitary layer $\mathcal{U}$ as shown, combined with a fixed noise channel $\mathcal{E}_l = \mathcal{E} = \mathcal{E}_{XX} \circ \mathcal{E}_{XY}\, \forall l$. Specifically, $\mathcal{E}$ is defined as the composition of two dephasing-like channels on the XX and XY bases, respectively, namely, $\mathcal{E}_{XX}(\rho) = (1-p)\rho + pXX\rho XX$ and $\mathcal{E}_{XY}(\rho) = (1-p)\rho + pXY\rho XY$.

By analyzing the generators of the parameterized gates R_X and R_{ZZ} (i.e., X and ZZ), one can verify that they commute with both the other rotations appearing in $\mathcal{U}$ (i.e., R_Z and R_{XX}) and with the noise channel for any choice of parameters. Consequently, their action can be moved to the end of the circuit, leaving a mixed channel at the beginning. Notably, the state $\rho = |00\rangle\langle 00| + |11\rangle\langle 11|$ is a fixed point of this mixed channel, as it is singularly as a fixed point of each of the remaining operations. As a result, by choosing ρ as the initial state, the circuit is effectively equivalent to its purely unitary counterpart and thus free of NIBP. This is confirmed by an analytical computation of the variance via Eq. (12), which yields a nonvanishing result that is independent of the number of layers. In particular, Fig. 9 numerically corroborates the prediction of $\mathbb{V}^L_{\rho,H} = 1$ for $H = ZZ + XI + YZ$. Furthermore, by substituting $\mathcal{U}$ with $\mathcal{U}'$, obtained by replacing the first RX gate with an RY gate, Fig. 9 shows how the NIBP phenomenon reappears, highlighting again the crucial role in the interplay between noise and circuit structure.

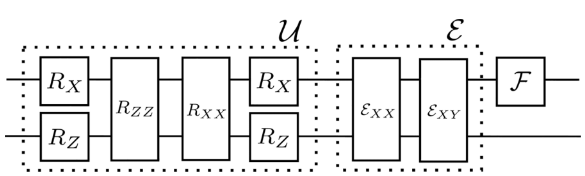

FIG. 8. Single layer structure used in the numerical examples. The circuit acts on a two-qubit register and is divided into three blocks, each representing either unitary gates ($\mathcal{U}$) or noise channels ($\mathcal{E}$ and $\mathcal{F}$).

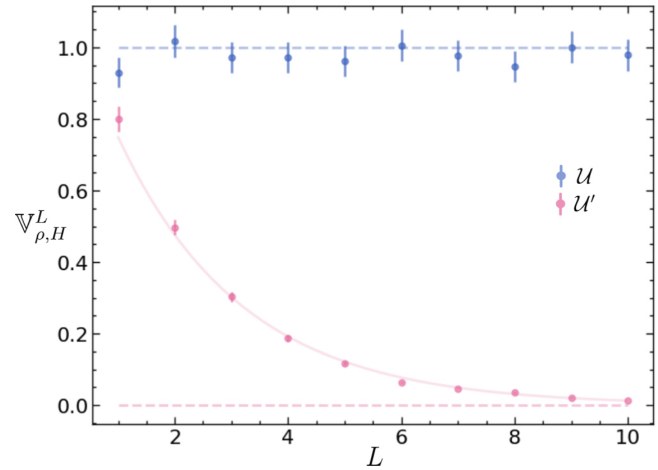

FIG. 9. Example of noise-induced barren plateau-free unital, nonunitary channel. Dotted lines represent the theoretical deep-circuit limiting values for the variance in the case of layers defined by $\mathcal{U} \circ \mathcal{E}$ (blue) and $\mathcal{U}' \circ \mathcal{E}$ (pink), where $\mathcal{U}'$ is obtained from $\mathcal{U}$ in Fig. 8 by substituting the first RX gate with an RY gate. The solid pink line is an exponential fit of the convergence to the limiting value.

2. Symmetry breaking can improve the variance

Even in the noncommuting case, we argue that noisy circuits can outperform purely unitary ones, providing the example of symmetry breaking in equivariant quantum circuits [26,27]. To carry out the analysis, we use the following definition.

Definition G1 (Equivariant unitary transformation). Given a symmetry group $G \subset U(d)$, a unitary transformation $W \in U(d)$ is said to be *G-equivariant* if

$$[W, V] = 0\, \forall V \in G \subset U(d). \tag{G1}$$

For convenience, we will express the commutation relation in Definition G1 as

$$[\mathcal{W}, \mathcal{V}] = \mathcal{W} \circ \mathcal{V} - \mathcal{V} \circ \mathcal{W} = 0, \tag{G2}$$

in terms of the superoperators $\mathcal{W}(\cdot) = W \cdot W^\dagger$ and $\mathcal{V}(\cdot) = V \cdot V^\dagger$, with $V \in G$.

Analogously to Definition G1, a parameterized gate $\mathcal{U}_\theta^{(\mathrm{eq})}$ is said to be G-equivariant if Definition G1 holds for all values of the parameters. Layered equivariant circuits are then constructed by composing such equivariant gate layers, namely,

$$\Phi_\theta^{(\mathrm{eq})} = \mathcal{U}_{\theta_{L+1}}^{(\mathrm{eq})} \circ \mathcal{U}_{\theta_L}^{(\mathrm{eq})} \cdots \circ \mathcal{U}_{\theta_1}^{(\mathrm{eq})}. \tag{G3}$$

When circuit equivariance is combined with a suitable choice of initial state ρ and observable H, strong constraints can be placed on the variance $\mathbb{V}_{\rho,H}^L$ of $\Phi_\theta^{(\mathrm{eq})}$ for any depth L. As an illustrative example, consider a *G-invariant* initial state ρ, i.e., satisfying $\mathcal{V}(\rho) = \rho$ for all $V \in G$. In this case, the loss function can be rewritten as

$$\begin{aligned}\mathcal{L}_{\rho,H} &= \mathrm{Tr}\left[\Phi_\theta^{(\mathrm{eq})}(\rho)H\right] = \mathrm{Tr}\left[\Phi_\theta^{(\mathrm{eq})} \circ \mathcal{V}(\rho)H\right] \\ &= \mathrm{Tr}\left[\mathcal{V} \circ \Phi_\theta^{(\mathrm{eq})}(\rho)H\right] = \mathrm{Tr}\left[\Phi_\theta^{(\mathrm{eq})}(\rho)\mathcal{V}^\dagger(H)\right].\end{aligned} \tag{G4}$$

Averaging the final expression over the Haar measure on G yields

$$\begin{aligned}\mathcal{L}_{\rho,H} &= \int d\mu(V)\mathrm{Tr}\left[\Phi_\theta^{(\mathrm{eq})}(\rho)\mathcal{V}^\dagger(H)\right] \\ &= \mathrm{Tr}\left[\Phi_\theta^{(\mathrm{eq})}(\rho)\mathcal{P}_G(H)\right],\end{aligned} \tag{G5}$$

where $\mathcal{P}_G$ denotes the averaging projector onto the G-invariant subspace of $\mathcal{B}$. If the observable H has no G-invariant component, i.e., $\mathcal{P}_G(H) = 0$, then $\mathcal{L}_{\rho,H} \equiv 0$, implying a vanishing variance for all L.

In the presence of generic noise maps $\mathcal{E}_l$, however, the equivariance of the full circuit

$$\tilde{\Phi}_\theta^{(\mathrm{eq})} = \mathcal{U}_{\theta_{L+1}}^{(\mathrm{eq})} \circ \mathcal{E}_L \circ \mathcal{U}_{\theta_L}^{(\mathrm{eq})} \cdots \circ \mathcal{E}_1 \circ \mathcal{U}_{\theta_1}^{(\mathrm{eq})}, \tag{G6}$$

may be broken. Allowing for a nonvanishing commutator between the noisy circuit and the group action, Eq. (G4) generalizes to

$$\mathcal{L}_{\rho,H} = \mathrm{Tr}\left[\tilde{\Phi}_\theta^{(\mathrm{eq})}(\rho)\mathcal{V}^\dagger(H)\right] + \mathrm{Tr}\left[[\tilde{\Phi}_\theta^{(\mathrm{eq})}, \mathcal{V}](\rho)H\right]. \tag{G7}$$

While the first term, corresponding to the equivariant component of the noisy channel, vanishes upon averaging over the Haar measure of G as in Eq. (G5), the second term encodes explicit *symmetry breaking* and can, in general, be nonzero. In particular, it can be shown to depend on the commutator between the averaging projector $\mathcal{P}_G$ and the noisy map $\tilde{\Phi}_\theta^{(\mathrm{eq})}$. As a consequence, whenever this commutator does not vanish, the variance in the noisy setting is expected to be strictly positive, highlighting a qualitative difference with respect to the purely unitary case.

To better understand this phenomenon, we go back to the circuit of Fig. 8. Now, we change initial state, setting it to $\rho = \mathbb{1} \otimes |0\rangle\langle 0|$, while keeping H. We now show that by virtue of this choice, the variance of the unitary circuit becomes identically zero. Indeed, as already noted in the previous numerical example, R_X and R_{ZZ} commute with both the other rotations appearing in $\mathcal{U}$ (i.e., R_Z and R_{XX}), and in the absence of noise, can thus be moved to the beginning of the circuit. Moreover, the new initial state ρ is a steady state of both these transformations, and as a consequence, their presence can be ignored. Now, by selecting G as the group generated by the Lie closure $\mathfrak{g} = \langle iX, iZZ\rangle_{Lie}$, we immediately get from the previous discussion that the remaining circuit is G-equivariant and ρ is a G-invariant. The averaging projector $\mathcal{P}_G$ can be computed using Weingarten calculus,

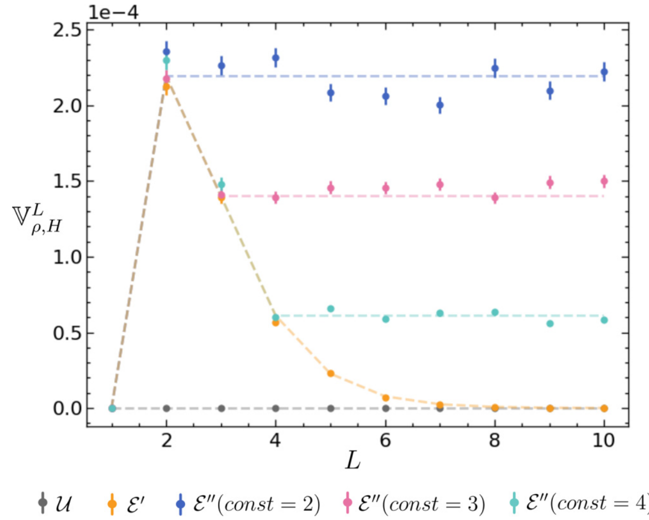

FIG. 10. Variance improvement with unital channels. Dotted lines represent the theoretical prediction given by Eq. (8), while points indicate the numerically estimated circuit variances.

$$\mathcal{P}_G(\cdot) = \sum_{P\in\{\mathbb{1},IZ,XX,XY\}} \mathrm{Tr}[P\cdot]P, \tag{G8}$$

where the sum runs over normalized Pauli strings, yielding the final result $\mathcal{P}_G(H) = 0$. As a consequence, $\mathcal{L}_{\rho,H} \equiv 0$ and $\mathbb{V}_{\rho,H}^L = 0\,\forall L$. This result is numerically confirmed in Fig. 10.

Introducing a symmetry-breaking noise channel, we can still achieve an improved variance, as predicted by Eq. (G7). Here, we consider a modified noise channel $\mathcal{E}' = \mathcal{E} \circ \mathcal{F}$, where $\mathcal{F}$ is a single-qubit channel applied to the first qubit, defined as $\mathcal{F}(\rho) = (1-p)\rho + p\mathcal{A}(\rho)$, and

$$\begin{aligned}\mathcal{A}(\rho) = &\frac{q}{\sqrt{2}}\mathrm{Tr}[X\rho]X + \frac{q}{\sqrt{2}}\mathrm{Tr}[X\rho]Y \\ &+ (1-q)\mathrm{Tr}[Y\rho]Y \\ &+ \frac{q}{\sqrt{2}}\mathrm{Tr}[Z\rho]Z + \frac{q}{\sqrt{2}}\mathrm{Tr}[Z\rho]Y.\end{aligned} \tag{G9}$$

By computing its Choi matrix, one can confirm that $\mathcal{F}$ is completely positive and trace-preserving (CPTP) and therefore constitutes a valid quantum channel whenever $0 \le p \le 1$ and $0 \le q \le 1/2$.

Although $\mathcal{E}'$ is still not strictly contractive, it is *contractive enough* to satisfy Corollary A2 and thus provably induces the NIBP effect, as corroborated numerically in Fig. 10. Interestingly, however, the presence of $\mathcal{A}$, which does not commute with R_X and R_{ZZ}, breaks the equivariant structure of before, allowing transient variance increases. Indeed, the behavior of $\mathbb{V}_{\rho,H}$ under this channel is not monotonic with respect to the circuit depth: it initially increases before eventually decaying to zero in the deep-circuit limit. This observation motivates the consideration of a nonuniform noise profile, where $\mathcal{E}_l \neq \mathcal{E}_{l'}$ for some l, l', defined as

$$\mathcal{E}''_l = \begin{cases} \mathcal{E}' & \text{if } l \le \text{const}, \\ \mathcal{E} & \text{otherwise}. \end{cases} \tag{G10}$$

In such a setup, the circuit benefits from the absence of concentration observed in the first example, as well as the transient variance enhancement from the second, converging to a nonzero value dependent on the constant rather than the number of layers. This hybrid effect is clearly demonstrated in Fig. 10. The mechanism at play here is at the core of the *absorption* effects shown in section IV.B.

Finally, we emphasize that, despite the limited system size considered ($n = 2$ qubits), these examples are still insightful in illustrating the absence of NIBP. This is because the key quantity of interest is the scaling with respect to the circuit depth L, rather than the number of qubits n.

APPENDIX H: NONUNITAL NOISE AND ENTANGLEMENT

In this appendix we explicitly derive the absorption terms pertaining to the trivial component $\mathcal{B}_0$, in the presence of strictly contractive, yet nonunital noise, first in general, and then on a specific system used for the numerical calculations.

1. Numerics

Exploiting the knowledge about the structure of the absorption matrix A, derived from Theorem IV.1, it is possible to study the scaling of the variance $\mathbb{V}^{\infty}_{\rho,H}$ as a function of the noise strength and entangling power of the unitary circuit. To do so, let us consider a nonunital map of the form

$$\mathcal{E}_c(\rho) = (1-p)\mathcal{E}(\rho) + p\tilde{\rho}, \tag{H1}$$

where $\tilde{\rho} \neq \mathbb{1}/d$ is an arbitrary quantum state, $\mathcal{E}$ is a unitary channel representing the entangling operation, and p is the error probability associated with $\mathcal{E}_c$. Intuitively, we can think of the resulting channel Φ_θ as the repetition of L layers, each made up of the composition of $\tilde{\Phi}(\rho) = (1-p)\rho + p\tilde{\rho}$ and of $\mathcal{E} \circ \mathcal{U}_{\theta_l}$. Clearly $\tilde{\rho}$ is the unique fixed-point of $\tilde{\Phi}$, while the local unitaries are assumed to form 2-designs, the only fixed point for the unital part, valid for all parameters, is the maximally mixed state $\mathbb{1}/d$. This causes the emergence of competing effects, which are the ultimate origin of the variance in such models. However, depending on the relative strength of the two effects, the behavior of $\mathbb{V}^{\infty}_{\rho,H}$ as a function of p may vary greatly. Theorem IV.1 allows to quantify the variance scaling in the limit of a rapidly entangling and slowly entangling channel, i.e., $\beta \to \infty$ and $\beta \to 0$, respectively. In particular, for both limits, the LTM approaches a projection, namely, the dominant eigenprojection in the former and the identity in the latter. In these cases, $\mathbb{V}^{\infty}_{\rho,H}$ is given by the following lemma.

Lemma H1. Let $\mathcal{E}_c$ be a quantum channel of type Eq. (H1), with $0 < p \le 1$ and let T be the transition matrix of $\mathcal{E}^\dagger$. Then

$$\mathbb{V}^{\infty}_{\rho,H} = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1}\ell_H). \tag{H2}$$

In particular, if T is a projection, then

$$\mathbb{V}^{\infty}_{\rho,H} = \left(\frac{p}{2-p} - p^2\right)(\ell_{\tilde{\rho}}, T\ell_H) + p^2(\ell_{\tilde{\rho}}, \ell_H). \tag{H3}$$

The first thing to notice is that the dependence on the initial state of $\mathbb{V}^{\infty}_{\rho,H}$ is completely lost: the component associated with it decays exponentially fast in the number of layers when $p \in \Theta(1)$, and as a consequence, vanishes in the limit. The remaining terms, instead, pertain to the fixed point of the noise channel $\tilde{\rho}$ and therefore still appear. In particular, the last term pertains to the very last layer, while the first collects the contribution of all preceding layers. Clearly, since the decay of these contributions is exponential, only layers where $L - l \lesssim -2\log p$ will contribute sensibly to the variance.

Lemma H1 allows to compute the scaling as a function of p in the two opposite limits. Starting from the slowly entangling case, i.e., $T \approx \mathbb{1}$, the scaling is approximately *linear* in p. More precisely, we have

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{p}{2-p}(\ell_{\tilde{\rho}}, \ell_H). \tag{H4}$$

The result shows that the first term in Eq. (H3) dominates, suggesting that $\mathbb{V}^{\infty}_{\rho,H}$ emerges from the contribution of the last $O(-\log p)$ layers. In this sense, for fixed noise rates, only the last portions of the channels are relevant for VQAs [32]. Conversely, in the opposite limit, the dependence on p is more complex, as it now depends on the irreducible components T_z. For simplicity, if we consider the case of a highly expressive ansatz, we may take T to have only one irreducible component T_1. In this case, the last term in Eq. (H3) dominates, and we get a *quadratic* dependence on p up to an exponentially vanishing correction, namely,

$$\mathbb{V}^{\infty}_{\rho,H} = p^2(\ell_{\tilde{\rho}}, \ell_H) + O\left(4^{-n}\right). \tag{H5}$$

This worsens the concentration, suggesting that now only the very last layer is able to produce a sizable effect, hence giving an effectively constant depth circuit.

We provide numerical evidence for the application discussed here, derived from the formalism introduced in this work. We utilize Pennylane [47] to construct and optimize the PQC described in the following.

Regarding the initial state, we use $\rho = (|0\rangle\langle 0|)^{\otimes n}$ for simplicity. As for the channel, we consider maps $\mathcal{E}_c$ of the family $\mathcal{E}_c = \mathcal{N} \circ \mathcal{E}$, $\mathcal{E}_c(\rho) = (1-p)\mathcal{E}(\rho) + p\tilde{\rho}$. In particular, $\mathcal{E}$ is a unitary, entangling channel depicted in Fig. 11 and $p \in (0, 1]$ represents the noise strength of the noise map $\mathcal{N}(\rho) = (1-p)\rho + p\tilde{\rho}$ with fixed point $\tilde{\rho}$. Specifically, we fix $\tilde{\rho}$ to be a highly entangled, pure state, i.e., the GHZ state $\tilde{\rho} = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})(\langle 0|^{\otimes n} + \langle 1|^{\otimes n})/2$.

Concerning H, we fix $H = h\sum_{k=1}^{n} 2^{n/2} Z_k \otimes Z_{k+1}$, as it represents the simplest observable involving all n qubits while having a nonvanishing normalization factor $(\ell_{\tilde{\rho}}, \ell_H)$. In particular, as shown in Appendix H 3, if we fix $h = 9/n$, we have $(\ell_{\tilde{\rho}}, \ell_H) = 1$, which makes the scaling of $\mathbb{V}^{\infty}_{\rho,H}$ especially easy to check. Finally, the entangling part is chosen according to Fig. 11. While the number of layers L needed to reach convergence to $\mathbb{V}^{\infty}_{\rho,H}$ is logarithmic, as shown in Theorem IV.1 and numerically assessed in the inset of Fig. 12, the mixing speed varies depending on the entangling part. For this reason, $L = 8$ layers are sufficient in the rapidly entangling case, but $L = 20$ are necessary in the slowly entangling case. Finally, all simulations are performed using $n = 10$ qubits. The results are shown in Fig. 12. The plots show both the quadratic and

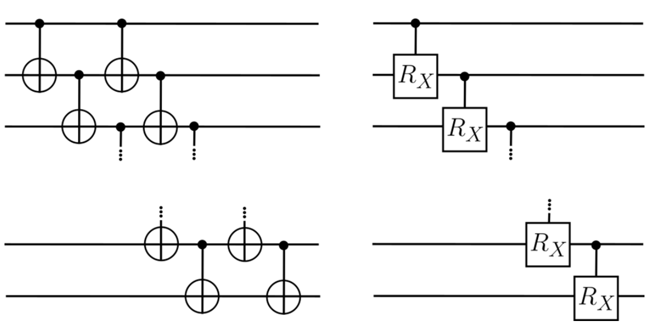

FIG. 11. Entangling unitaries used in the examples. On the left, the rapidly entangling configuration is composed of a double cascade of CNOT gates, while on the right, the slowly entangling one is composed of a single cascade of controlled RX gates, where $RX(\theta) = e^{i\theta X/2}$ and X is the Pauli X gate. In particular, we fixed $\theta = \pi/20$.

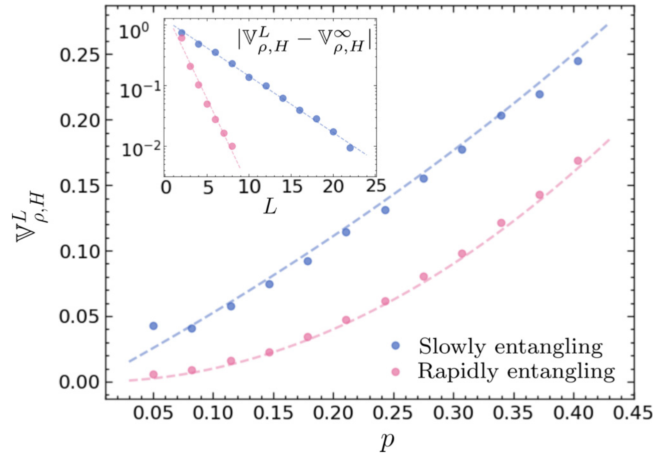

FIG. 12. Scaling of $\mathbb{V}^{L}_{\rho,H}$ as a function of the noise strength and the entangling power of the intermediate channel. The main figure illustrates the scaling of $\mathbb{V}^{\infty}_{\rho,H}$ with noise strength p for both rapidly entangling (pink) and slowly entangling (light blue) channels, using $L = 8$ and $L = 20$, respectively. The dotted lines represent the theoretical predictions of Eqs. (H5) and (H4). The inset verifies the exponential convergence of $\mathbb{V}^{L}_{\rho,H}$ to $\mathbb{V}^{\infty}_{\rho,H}$ at $p = 0.1$, justifying the chosen number of layers. The dotted lines represent an exponential fit to the numerical data. All plots are obtained using a $n = 10$ qubit system.

linear scaling with p predicted by our model, as well as an exponential decay in the difference $|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}|$ for fixed noise rates. A slight deviation from the predicted scaling is observed in the slowly entangling setting for $p \approx 0$. This effect can be explained by the finite amount of layers used: the slow speed of convergence imposed by the condition $T \approx \mathbb{1}$ prevents the variance to reach the asymptotic limit $\mathbb{V}^{\infty}_{\rho,H}$, while the action of the noise is still too weak to erase the contribution of the first layers, hence deviating from the predicted behavior. This same phenomenon is not observed in the rapidly entangling case, as entanglement production already exponentially suppresses those contributions.

2. Calculation of the absorption term

Explicit computation of the absorption matrix $A = R(\mathbb{1} - Q)^{-1}$ is a nontrivial task, which should be tackled on a case-by-case basis. Indeed, in the nonunital case, this term describes the complex phenomenon arising from the interaction of two competing effects, which drive the system toward different states. Analytical summation of A is however can be feasible and still give insight on the interaction of the two. Indeed, using a simplified model, we can perform this calculation and still be able to appreciate the different effects that rapidly entangling and slowly entangling circuits have on a fixed, nonunital noise as a function of its strength. In particular, here we prove Lemma H1, which we repeat for convenience.

Lemma H2. Let $\mathcal{E}_c$ be a quantum channel of type Eq. (H1), with $0 < p \leq 1$ and let T be the transition matrix of $\mathcal{E}^\dagger$. Then

$$\mathbb{V}^{\infty}_{\rho,H} = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1}\ell_H). \tag{H6}$$

In particular, if T is a projection, then

$$\mathbb{V}^{\infty}_{\rho,H} = \left(\frac{p}{2-p} - p^2\right)(\ell_{\tilde{\rho}}, T\ell_H) + p^2(\ell_{\tilde{\rho}}, \ell_H). \tag{H7}$$

Proof. The first result follows directly from Theorem E5, in particular, Corollary E3, by computation of A_0. In particular, we can explicitly compute R_0 elementwise as

$$\begin{aligned}
(R_0)_{0,\kappa} &= \frac{1}{d_\kappa}\sum_j \mathrm{Tr}\left[\frac{\mathbb{1}}{\sqrt{d}}\mathcal{E}^\dagger_c(P_j)\right]^2 \tilde{\delta}_{j,\kappa} \\
&= \frac{1}{d_\kappa}\sum_j \mathrm{Tr}\left[\mathcal{E}_c\left(\frac{\mathbb{1}}{\sqrt{d}}\right)P_j\right]^2 \tilde{\delta}_{j,\kappa} \\
&= \frac{1}{d_\kappa}\sum_j \left((1-p)\mathrm{Tr}\left[\Phi\left(\frac{\mathbb{1}}{\sqrt{d}}\right)P_j\right]\right. \\
&\qquad \left. + p\sqrt{d}\,\mathrm{Tr}\left[\tilde{\rho}P_j\right]\right)^2 \tilde{\delta}_{j,\kappa} \\
&= dp^2\frac{1}{d_\kappa}\sum_j \mathrm{Tr}\left[\tilde{\rho}P_j\right]^2 \tilde{\delta}_{j,\kappa} = dp^2(\ell_{\tilde{\rho}})_\kappa,
\end{aligned} \tag{H8}$$

by unitality of Φ. By an analogous calculation, it can be shown that $Q = (1-p)^2 T$, and by trace preservation $(P_0)_{0,0} = 1$. With these elements, we can compute $A_0 = P_0 R_0 (1-Q)^{-1}$ and we get

$$(A_0)_{0,\lambda} = dp^2(\ell_{\tilde{\rho}})_\kappa(1 - (1-p)^2 T)^{-1}_{\kappa,\lambda}. \tag{H9}$$

In particular, since for any initial state ρ, we have $(\ell_\rho)_0 = 1/d$ by normalization, we get the final result

$$(\ell_\rho, A_0\ell_H) = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1}\ell_H). \tag{H10}$$

Using this, we can explicitly compute the right-hand side in the simplified setting where T is a projection. In that case in particular, we have that

$$\begin{aligned}
&(\mathbb{1} - (1-p)^2 T)^{-1} \\
&\quad = (\mathbb{1} - T) + (1 - (1-p)^2)^{-1} T \\
&\quad = (\mathbb{1} - T) + \frac{1}{p(2-p)} T \\
&\quad = \mathbb{1} + \left(\frac{1}{p(2-p)} - 1\right) T,
\end{aligned} \tag{H11}$$

which gives the result. ■

3. Choice of the system for the numerical example

In this appendix we give an explicit construction of the noise channels used in the numerical example shown in Fig. 3, as well as the choice and normalization procedure of the observable H used. In particular, the analysis focuses on the case of unitary entangling map Φ whose transfer matrix is well approximated by a projector and a highly entangled, pure fixed point, namely, the GHZ state $\tilde{\rho} = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})(\langle 0|^{\otimes n} + \langle 1|^{\otimes n})/2$. Thanks to Lemma H1, we know that the main contribution to $\mathbb{V}^{\infty}_{\rho,H}$ in this setting comes from the term $(\ell_{\tilde{\rho}}, \ell_H)$, so to numerically assess the results, it is useful to choose the observable H in order to normalize this factor. In particular, given the qubit structure, we can express the $\tilde{\rho}$ in terms of the *normalized* Pauli basis $\{\mathbb{1}, X, Y, Z\}^{\otimes n}$, which allows to easily compute the locality vectors. Using the spectral decomposition of the Pauli matrices, it is easy to see that

$$\begin{aligned}
|0\rangle\langle 0| &= \frac{\mathbb{1} + Z}{\sqrt{2}}, \; |0\rangle\langle 1| = \frac{X + iY}{\sqrt{2}}, \\
|1\rangle\langle 1| &= \frac{\mathbb{1} - Z}{\sqrt{2}}, \; |1\rangle\langle 0| = \frac{X - iY}{\sqrt{2}}.
\end{aligned} \tag{H12}$$

Using this decomposition, we can find a formula for $\tilde{\rho}$ exploiting a generalization of the binomial theorem.

Theorem H1. Let $A, B \in \mathbb{M}_d(\mathbb{C})$ be square matrices, and let S be the set of all permutations of n elements. Then

$$(A + \omega B)^{\otimes n} = \sum_{j=0}^{n} \omega^j \sum_{\sigma\in S} \sigma(A^{\otimes(n-j)} \otimes B^{\otimes j}), \tag{H13}$$

where $\omega \in \mathbb{C}$ and the permutation σ is applied to the qubit ordering.

In particular, the following corollary will be the most useful in performing the computations

Corollary H1. Let $A, B \in \mathbb{M}_d(\mathbb{C})$ be square matrices, and let S be the set of all permutations of n elements. Then

$$\begin{aligned}
&\frac{(A + \omega B)^{\otimes n} + (A - \omega B)^{\otimes n}}{2} \\
&\quad = \sum_{j=0}^{\lfloor n/2 \rfloor} \omega^{2j} \sum_{\sigma\in S} \sigma(A^{\otimes(n-2j)} \otimes B^{\otimes 2j}),
\end{aligned} \tag{H14}$$

where $\omega \in \mathbb{C}$ and the permutation σ is applied to the qubit ordering.

Proof. This corollary follows directly from Theorem H1, and noticing that all even-indexed terms in both $(A + \omega B)^{\otimes n}$ and $(A - \omega B)^{\otimes n}$ are equal, while odd-numbered terms have an opposite sign and therefore cancel out. ■

Applying the corollary to the appropriate pairs of projectors, we can get the final expression

$$\tilde{\rho} = \frac{1}{2^{n/2}} \sum_{j=0}^{\lfloor n/2 \rfloor} \sum_{\sigma \in S} \sigma(\mathbb{1}^{\otimes(n-2j)} \otimes Z^{\otimes 2j}) + (-1)^j \sigma(X^{\otimes(n-2j)} \otimes Y^{\otimes 2j}). \tag{H15}$$

As it is clear from Eq. (H15), the fixed point of the channel has a nonvanishing component only on Pauli strings that are either nontrivial on all qubits or nontrivial in *only* in an *even* number of qubits. Then it follows that the simplest observable involving all qubits and with nonvanishing variance is of form $H = h\sum_{k=1}^{n} 2^{n/2} Z_k \otimes Z_{k+1}$, where the factor $2^{n/2}$ accounts for normalization of Z and cancels out with the corresponding factor in Eq. (H15) in the calculations of $(\ell_{\tilde{\rho}}, \ell_H)$. Finally, since each term in the sum is orthogonal, it gives an independent contribution of $1/(d^2-1)^2 = 1/9$, consequently by choosing $h = 9/n$ we have that $(\ell_{\tilde{\rho}}, \ell_H) = 1$ is normalized.

APPENDIX I: EXAMPLE OF A CIRCUIT WITH NONCONVERGENT VARIANCE

As Theorem IV.1 suggests, the generic circuit of Eq. (2) need not have a well-defined value for the deep circuit limit of its variance, namely, the limit $\lim_{L\to\infty} \mathbb{V}^L_{\rho,H}$ no need to exist. As discussed in Appendix E, this property is related to the presence of *cycles* in T, i.e., the presence of periodic irreducible blocks T_z with period $p > 1$. As a specific example, consider the circuit depicted in Fig. 13. The simple structure of this circuit allows to explicitly compute $\mathbb{V}^L_{\rho,H}$ as a function of L. Assuming $\mathrm{Tr}[H] = 0$, we have

$$\mathbb{V}^L_{\rho,H} = \begin{cases} \frac{(\|\rho\|_2^2 - 1/2)\|H\|_2^2}{3} & \text{if } L \text{ is even,} \\ 0 & \text{if } L \text{ is odd,} \end{cases} \tag{I1}$$

from which it is clear that the deep circuit limit does not converge. Moreover, thanks to the contained dimension of the system, it is possible to compute and represent T,

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow T_0 = T_2 = \begin{pmatrix} 1 \end{pmatrix}, \; T_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{I2}$$

Here we can see that T_1 has in fact period 2, which implies the presence of two distinct limiting values whenever both ρ and H have a component belonging to T_1, consistently with what was discussed above.

On the contrary, in accordance with Theorem IV.1, the Cesàro average of the variance is always well-defined, and in this case we get

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} \mathbb{V}^l_{\rho,H} = \frac{(\|\rho\|_2^2 - 1/2)\|H\|_2^2}{6}, \tag{I3}$$

which can be recovered both from direct calculation and by computing the right leading eigenvector $w_1 = (1/2, 1/2)^t$ of T_1.

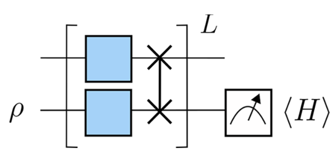

FIG. 13. Simple 2-qubit circuit, designed to have a nonconvergent variance. In this construction, the channel $\mathcal{E}$ is chosen to be unitary and, in particular, to be a SWAP gate. Moreover, both the initial state ρ and the observable H are chosen to be nontrivial only on the second qubit and are therefore represented as single-qubit operators.

[1] R. P. Feynman, Simulating physics with computers, Int. J. Theor. Phys. **21**, 467 (1982).

[2] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[3] A. Di Meglio *et al.*, Quantum computing for high-energy physics: State of the art and challenges, PRX Quantum **5**, 037001 (2024).

[4] E. Fontana, N. Fitzpatrick, D. M. Ramo, R. Duncan, and I. Rungger, Evaluating the noise resilience of variational quantum algorithms, Phys. Rev. A **104**, 022403 (2021).

[5] M. Vischi, G. Di Bartolomeo, M. Proietti, S. Koudia, F. Cerocchi, M. Dispenza, and A. Bassi, Simulating photonic devices with noisy optical elements, Phys. Rev. Res. **6**, 033337 (2024).

[6] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[7] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Diagnosing barren plateaus with tools from quantum optimal control, Quantum **6**, 824 (2022).

[8] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021).

[9] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn, Avoiding barren plateaus using classical shadows, PRX Quantum **3**, 020365 (2022).

[10] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum **3**, 010313 (2022).

[11] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Absence of barren plateaus in quantum

convolutional neural networks, Phys. Rev. X **11**, 041011 (2021).

[12] E. Cervero Martín, K. Plekhanov, and M. Lubasch, Barren plateaus in quantum tensor network optimization, Quantum **7**, 974 (2023).

[13] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[14] A. V. Uvarov and J. D. Biamonte, On barren plateaus and cost function locality in variational quantum algorithms, J. Phys. A: Math. Theor. **54**, 245301 (2021).

[15] S. Thanasilp, S. Wang, N. A. Nghiem, P. Coles, and M. Cerezo, Subtleties in the trainability of quantum machine learning models, Quantum Mach. Intell. **5**, 21 (2023).

[16] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. Ortiz Marrero, M. Larocca, and M. Cerezo, A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits, Nat. Commun. **15**, 7172 (2024).

[17] E. Fontana, D. Herman, S. Chakrabarti, N. Kumar, R. Yalovetzky, J. Heredge, S. H. Sureshbabu, and M. Pistoia, Characterizing barren plateaus in quantum ansätze with the adjoint representation, Nat. Commun. **15**, 7171 (2024).

[18] N. L. Diaz, D. García-Martín, S. Kazi, M. Larocca, and M. Cerezo, Showcasing a barren plateau theory beyond the dynamical lie algebra, arXiv:2310.11505.

[19] C.-Y. Park and N. Killoran, Hamiltonian variational ansatz without barren plateaus, Quantum **8**, 1239 (2024).

[20] C.-Y. Park, M. Kang, and J. Huh, Hardware-efficient ansatz without barren plateaus in any depth, arXiv:2403.04844.

[21] K. Zhang, L. Liu, M.-H. Hsieh, and D. Tao, in *Proceedings of the 36th International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2022).

[22] Y. Wang, B. Qi, C. Ferrie, and D. Dong, Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization, Phys. Rev. Appl. **22**, 054005 (2024).

[23] R. Puig, M. Drudis, S. Thanasilp, and Z. Holmes, Variational quantum simulation: A case study for understanding warm starts, PRX Quantum **6**, 010317 (2025).

[24] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[25] M. Schumann, F. K. Wilhelm, and A. Ciani, Emergence of noise-induced barren plateaus in arbitrary layered noise models, Quantum Sci. Technol. **9**, 045019 (2024).

[26] G. Crognaletti, G. Di Bartolomeo, M. Vischi, and L. Loris Viteritti, Equivariant variational quantum eigensolver to detect phase transitions through energy level crossings, Quantum Sci. Technol. **10**, 015048 (2025).

[27] C. Tüysüz, S. Y. Chang, M. Demidik, K. Jansen, S. Vallecorsa, and M. Grossi, Symmetry breaking in geometric quantum machine learning in the presence of noise, PRX Quantum **5**, 030314 (2024).

[28] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, Barren plateaus in variational quantum computing, Nat. Rev. Phys. **7**, 174 (2025).

[29] Y. Quek, D. Stilck França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, Nat. Phys. **20**, 1648 (2024).

[30] R. Takagi, H. Tajima, and M. Gu, Universal sampling lower bounds for quantum error mitigation, Phys. Rev. Lett. **131**, 210602 (2023).

[31] K. Tsubouchi, T. Sagawa, and N. Yoshioka, Universal cost bound of quantum error mitigation based on quantum estimation theory, Phys. Rev. Lett. **131**, 210601 (2023).

[32] A. A. Mele, A. Angrisani, S. Ghosh, S. Khatri, J. Eisert, D. S. França, and Y. Quek, Noise-induced shallow circuits and absence of barren plateaus, arXiv:2403.13927.

[33] B. Fefferman, S. Ghosh, M. Gullans, K. Kuroiwa, and K. Sharma, Effect of nonunital noise on random-circuit sampling, PRX Quantum **5**, 030317 (2024).

[34] P. Singkanipa and D. A. Lidar, Beyond unital noise in variational quantum algorithms: Noise-induced barren plateaus and limit sets, Quantum **9**, 1617 (2025).

[35] E. Seneta, *Non-Negative Matrices and Markov Chains* (Springer, New York, 2006).

[36] H.-K. Zhang, S. Liu, and S.-X. Zhang, Absence of barren plateaus in finite local-depth circuits with long-range entanglement, Phys. Rev. Lett. **132**, 150603 (2024).

[37] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra* (Society for Industrial and Applied Mathematics, Philadelphia, 2000).

[38] K. He, X. Zhang, S. Ren, and J. Sun, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV, USA, 2016), pp. 770–778.

[39] G. Di Bartolomeo, M. Vischi, F. Cesa, R. Wixinger, M. Grossi, S. Donadi, and A. Bassi, Noisy gates for simulating quantum computers, Phys. Rev. Res. **5**, 043210 (2023).

[40] S. L. Adler and A. Bassi, Collapse models with non-white noises, J. Phys. A: Math. Theor. **40**, 15083 (2007).

[41] J. Heredge, M. West, L. Hollenberg, and M. Sevior, Non-unitary quantum machine learning, Phys. Rev. Appl. **23**, 044046 (2025).

[42] H. M. Wiseman and G. J. Milburn, Quantum theory of field-quadrature measurements, Phys. Rev. A **47**, 642 (1993).

[43] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[44] A. Katabarwa, K. Gratsea, A. Caesura, and P. D. Johnson, Early fault-tolerant quantum computing, PRX Quantum **5**, 020101 (2024).

[45] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing, Nat. Commun. **16**, 7907 (2025).

[46] D. Pérez-García, M. M. Wolf, D. Petz, and M. B. Ruskai, Contractivity of positive and trace-preserving maps under L_p norms, J. Math. Phys. **47**, 083506 (2006).

[47] V. Bergholm *et al.*, Pennylane: Automatic differentiation of hybrid quantum-classical computations, arXiv:1811.04968.

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

[49] A. S. Holevo and R. F. Werner, Evaluating capacities of bosonic Gaussian channels, Phys. Rev. A **63**, 032312 (2001).

[50] C. King and M. B. Ruskai, Minimal entropy of states emerging from noisy quantum channels, IEEE Trans. Inf. Theory **47**, 192 (2001).

[51] M. Beth Ruskai, S. Szarek, and E. Werner, An analysis of completely-positive trace-preserving maps on M_2, Linear Algebra Appl. **347**, 159 (2002).

[52] R. V. Kadison, A generalized Schwarz inequality and algebraic invariants for operator algebras, Ann. Math. **56**, 494 (1952).

[53] This is especially clear in light of Eq. (10) of the main text, and the relation $(A\ell_H)_z = A_z\ell_H$, coming from the definition of locality and Eq. (E19).