

Estimates of loss function concentration in noisy parametrized quantum circuits

Giulio Croгнаletti,^{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*

(Dated: October 4, 2024)

Variational quantum computing provides a versatile computational approach, applicable to a wide range of fields such as quantum chemistry, machine learning, and optimization problems. However, scaling up the optimization of quantum circuits encounters a significant hurdle due to the exponential concentration of the loss function, often dubbed the barren plateau (BP) phenomenon. Although rigorous results exist on the extent of barren plateaus in unitary or in noisy circuits, little is known about the interaction between these two effects, mainly because the loss concentration in noisy parameterized quantum circuits (PQCs) cannot be adequately described using the standard Lie algebraic formalism used in the unitary case. In this work, we introduce a new analytical formulation based on non-negative matrix theory that enables precise calculation of the variance in deep PQCs, which allows investigating the complex and rich interplay between unitary dynamics and noise. In particular, we show the emergence of a noise-induced absorption mechanism, a phenomenon that cannot arise in the purely reversible context of unitary quantum computing. Despite the challenges, general lower bounds on the variance of deep PQCs can still be established by appropriately slowing down speed of convergence to the deep circuit limit, effectively mimicking the behaviour of shallow circuits. Our framework applies to both unitary and non-unitary dynamics, allowing us to establish a deeper connection between the noise resilience of PQCs and the potential to enhance their expressive power through smart initialization strategies. Theoretical developments are supported by numerical examples and related applications.

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 to solve the given tasks faster. More specifically, this method requires to iteratively optimize the parameters θ of a parameterized quantum circuit (PQC) using a classical optimizer. The goal is to minimize a loss function, typically expressed as the expectation value of an observable H given ρ , the initial quantum state. 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. In the absence of noise, this phenomenon has been linked to sev-

eral factors, ranging from the expressive power and entangling capability of the circuit [7–11], to the locality of H [12, 13] and the entanglement of the initial state ρ [14]. Recently, the contributions coming from all such factors were unified in a Lie-algebraic framework [15–17], 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 [18–22], 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) [23, 24] and symmetry breaking [25, 26]. The former exacerbates the BP effect, producing a deterministic concentration. In this case, an efficient loss-minimizing direction is exponentially hard to estimate, *regardless* of the parameter’s choice. These phenomena are mostly linked to decoherence. Ref. [27] offers a review of the subject.

Indeed, the action of noise can be quite destructive for quantum computation, and overcoming their effect challenging. Along these lines, recent research [28–30] 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. Such analyses are mostly limited to unital noise. In fact, it has been shown how non-unital noise can have non-trivial effects on the appearance of NIBP [31–33], highlighting how the interaction between general noise and unitary circuit layers in

* giulio.croгнаletti@phd.units.it

† michele.grossi@cern.ch

the context of variational quantum computing is still an important open area of research.

In this work, we propose a formulation based on non-negative matrix theory, that allows a general description of these dynamics and naturally fills this gap. In particular, we are able to quantify the deep circuit variance $\mathbb{V}_{\rho,H}^\infty$ for layered circuit interleaved with general noise maps, thus surpassing Lie-algebraic limitations, as well as any restriction on the 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 happen only for noise maps that are not strictly contractive, and thus are not strong enough to completely wipe out all quantum information contained in the system. Within the same framework, also noise resilience properties of PQCs can be studied, as well as BP mitigation strategies. In particular, we show a direct connection between the noise resilience properties of a PQC and the potential for their expressive power to be enhanced by a small-angle initialization-like strategies.

II. SETTING AND NOTATION

In what follows we will consider the framework of variational quantum computing, focusing on finite-dimensional quantum systems whose Hilbert space $\mathcal{H} = \otimes_{m=1}^M \mathcal{H}_m$ is naturally divided into M subsystems, each of dimension d_m , so that $d = \dim(\mathcal{H}) = \prod_{m=1}^M d_m$. More specifically, we will consider n -qubit systems, where \mathcal{H}_m represent either single qubits or groups of qubits. In this context, we study the problem where a quantum state ρ is evolved using a parameterized quantum channel Φ_θ , whose parameters $\theta = (\theta_1, \theta_2, \dots) \in \Theta$ are optimized by minimizing the loss function

$$\mathcal{L}_{\rho,H} = \text{Tr} [\Phi_\theta(\rho)H], \quad (1)$$

where H is an observable of the system. In this context, 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 concentrate around its mean value in the number n of qubits¹.

Within this framework, we focus on the case of *layered* quantum channels, namely

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

where each layer is composed of a unitary part \mathcal{U}_{θ_l} and an arbitrary quantum channel \mathcal{E}_l . Such a channel will be deemed strictly-contractive if its induced Schatten 1-norm $\|\mathcal{E}_l\|$ is strictly less than unity, $\|\mathcal{E}_l\| < 1$, and contractive otherwise, i.e. $\|\mathcal{E}_l\| = 1$. In particular, a unitary channel is a special case of contractive channel, preserving all Schatten norms in its whole domain. We refer to Appendix C 1 b for more details.

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. It is worth noting 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 non trivially on \mathcal{H}_m if and only if $\kappa_m = 1$. Indeed, we can partition the whole space as

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

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

$$(\ell_A)_\kappa = \sum_{j=1}^{d_\kappa} \text{Tr} [B_j A]^2 \quad (4)$$

for some Hermitian, orthonormal basis $\{B_j\}_{j=1}^{d_\kappa}$ of \mathcal{B}_κ . 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 not new in the context of PQCs, and in fact the vector ℓ_A is analogous to the purity measures defined in [15, 17]. Based on this, a derived notion of *locality preservation* can be introduced for linear maps acting on \mathcal{B} . In particular, given a map $\Lambda : \mathcal{B} \rightarrow \mathcal{B}$ and two subspaces \mathcal{B}_κ and \mathcal{B}_λ , 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 II.1 (Locality transfer matrix). Given a linear map $\Lambda : \mathcal{B} \rightarrow \mathcal{B}$, its locality transfer matrix T is defined elementwise as

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

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

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

¹ It is worth mentioning that the term barren plateau can also be used in the literature referring to the exponential vanishing of the variance of $\partial_\theta \mathcal{L}_{\rho,H}$ instead of $\mathcal{L}_{\rho,H}$. However, since it has been shown that these two conditions are equivalent [34], we choose the latter for simplicity.

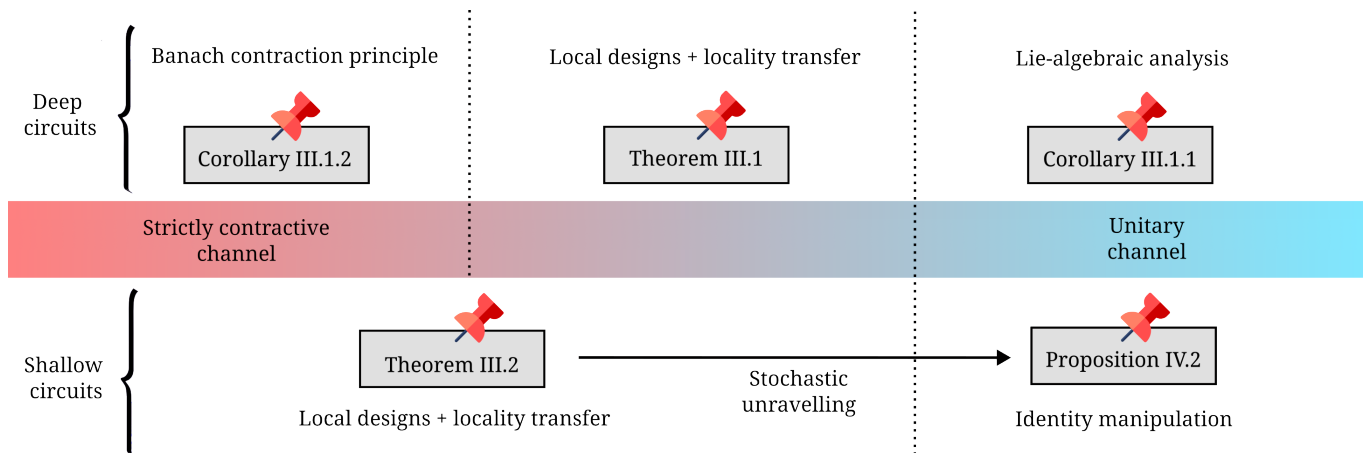


FIG. 1. Loss concentration in variational quantum computing. The analytical formulation proposed here employs non-negative matrix theory to describe the interplay between local 2-designs and noise. This allows for precise calculation of loss variances for generic noise maps, from strictly contractive to unitary channels, as illustrated by the coloured band in the Figure. The upper part considers deep circuits, where the loss variance $\mathbb{V}_{\rho,H}^L$ reaches its asymptotic limit. While loss concentration is well-understood for strictly contractive [23, 24] and unitary [15–17] channels, Theorem III.1 provides an analytical solution for the intermediate case, revealing a noise-induced absorption mechanism unique to this regime. Consistency with known limiting cases is verified in Corollary III.1.2 and Corollary III.1.1. The lower part focuses on “shallow” circuits, where a lower bound on $\mathbb{V}_{\rho,H}^L$ is established using Theorem III.2. This extends previous works on brickwork circuits [12, 31], enabling initialization strategies such as small angle initializations [20, 21] to be represented as stochastic unravellings of noise maps. In Section IV we expand upon their applicability. For instance, both unitary and non-unitary QResNets can be derived thanks to Proposition IV.2.

respect to the partition, namely $\mathcal{U} : \rho \mapsto U\rho U^\dagger$, where $U = \otimes_m U_m$, are locality preserving.

With this description in mind, we assume that \mathcal{U}_θ 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 realised 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}_{θ_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 .

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 $\Lambda : \mathcal{B} \rightarrow \mathcal{B}$ as a direct consequence of Definition II.1, and in particular we have $TD = (T^\dagger D)^t$, with $D_{\kappa,\lambda} = d_\kappa \delta_{\kappa,\lambda}$. For 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.

$$(a, b) = a^t D^{-1} b = \sum_{\kappa} \frac{a_{\kappa} b_{\kappa}}{d_{\kappa}}. \quad (6)$$

We refer the reader to the Appendix A for a formal derivation and more detailed discussion of all the properties of the quantities introduced thus far.

Having introduced the main tools, we now characterize the scaling of the variance $\mathbb{V}_{\rho,H}^L$ of $\mathcal{L}_{\rho,H}$ for quantum channels described by Eq. (2).

III. RESULTS

The overarching goal is to characterise the properties of the variance $\mathbb{V}_{\rho,H}^L$ 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 manuscript.

A. Loss variance calculation

The first task is to derive a formal expression for the variance $\mathbb{V}_{\rho,H}^L$ 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 III.1 (Warm up). *Let $\rho, H \in \mathcal{B}$ and let \mathcal{U}_θ be a locality preserving channel as described in Section II. Then we have*

$$\mathbb{E}_\theta \left\{ \text{Tr} [\mathcal{U}_\theta(\rho)H]^2 \right\} = (\ell_\rho, \ell_H) \quad (7)$$

where (\cdot, \cdot) is the scalar product defined in Eq. (6).

Proposition III.1 already gathers many interesting properties discussed in the literature [15–17]. For instance, it connects the notion of locality introduced in Section II to the second moment of \mathcal{U}_θ , which is composed of terms inversely proportional to the dimension of the corresponding subspace \mathcal{B}_κ . Notably, the terms pertaining to different localities give independent contributions to the variance, without the possibility of mixing. Indeed, this is due to the locality preserving structure of \mathcal{U}_θ , and in general this result can be extended, including intermediate quantum channels \mathcal{E}_l . In principle, they can entangle the subsystems, therefore putting in communication different subspaces \mathcal{B}_κ . To better describe this process, we work in the Heisenberg picture, namely using the adjoint action \mathcal{E}^\dagger of the channel. Doing so, one formally obtains the following Proposition.

Proposition III.2 (General formula). *Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered quantum channel as in described in Section II. Then, we have*

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

where (\cdot, \cdot) is the scalar product defined in Eq. (6), and each T_l is the LTM associated to the respective \mathcal{E}_l^\dagger .

Exploiting this, it follows that

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

since by construction, each layer \mathcal{U}_{θ_l} forms a global unitary 1-design (see Appendix B). 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 G 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, Proposition III.2 can still be profitably used as a theoretical tool, as it allows characterizing $\mathbb{V}_{\rho, H}^L$ 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 non-unitary 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 , $\text{Tr} [H] = 0$. In this case, Eq. (9) becomes $\mathbb{V}_{\rho, H}^L = (\ell_\rho, T^L \ell_H)$. 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 are irreducible [35]. 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. Foundational reference for this definition can be found in Ref. [35], while a more detailed discussion is proposed in Appendix C. 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} used, some general properties of the blocks T_z can be identified. For instance, due to trace preservation, the trivial subspace \mathcal{B}_0 always forms an essential component of T , which we denote by T_0 . Moreover, since such blocks are essential, it follows that $\mathcal{E}^\dagger(\mathcal{B}_z) \subset \mathcal{B}_z$, and complete positivity ensures that all T_z must be contractive in the sense of the spectral radius, i.e. $\rho(T_z) \leq 1$. Moreover, when the equality holds, the left eigenvector v_z of T_z of the dominant eigenvalue can be explicitly computed, yielding $(v_z)_\kappa = 1 \forall \kappa$. This suggests that, as $L \rightarrow \infty$, the general form for the variance reads

$$\mathbb{V}_{\rho, H}^\infty = \sum_z (\ell_\rho, w_z) (\ell_H)_z + (\ell_\rho, w_z) (A \ell_H)_z, \quad (10)$$

where w_z denotes the right eigenvector of the leading eigenvalue of T_z and A is a matrix of the same shape as R in Fig. 2. As a special case, 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 also w_z can be explicitly computed, yielding

$$\mathbb{V}_{\rho, H}^\infty = \sum_z \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z (A \ell_H)_z}{d_z}. \quad (11)$$

We refer to Appendix C for the proofs of the aforementioned properties. This intuition is formalized in the

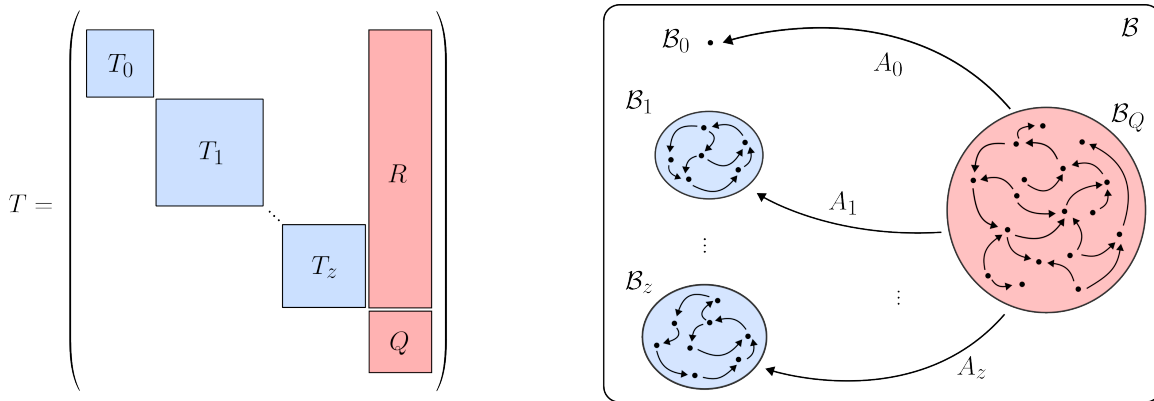


FIG. 2. Graphical representations of the stochastic process describing $\mathbb{V}_{\rho,H}^L$. On the left, we show the structure of the general locality transfer matrix (LTM), highlighting the decomposition into irreducible components [35]. Light blue blocks represent irreducible, essential components 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 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}_κ . 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 colour code of T .

following Theorem.

Theorem III.1 (Deep circuit variance). *Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered quantum channel as in described in Section II. Then the Cesàro average of $\mathbb{V}_{\rho,H}^L$ converges, and we have*

$$\left| \frac{1}{L} \sum_{l=0}^L \mathbb{V}_{\rho,H}^l - \mathbb{V}_{\rho,H}^\infty \right| \in O\left(e^{-\beta L} \|H\|_2^2\right), \quad (12)$$

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

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

where $A = R(\mathbf{1} - Q)^{-1}$ and the sum in Eq. (10) is performed over blocks T_z such that $\rho(T_z) = 1$.

It is worth noting that, the loss variance of the circuit need not converge. In fact, while it is bounded (i.e. $\mathbb{V}_{\rho,H}^L \leq \|H\|_\infty^2 \forall L$), different circuit sub-sequences 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 $\mathbb{V}_{\rho,H}^L$ will depend on which equivalence class of integers modulo p the depth L belongs to². This is put into an effective example in Appendix G. For simplicity, in the following we will assume *aperiodicity* (i.e. $p = 1$) for each irreducible block.

We start the analysis of Theorem III.1 by unpacking Eq. (11). In this equation, $\mathbb{V}_{\rho,H}^\infty$ is shown to depend

on four important quantities, namely the invariant subspaces \mathcal{B}_z of Φ_θ^\dagger , 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 $\mathcal{U}_\theta^\dagger$ and \mathcal{E}^\dagger have to be *well-aligned*, so that non-trivial subspaces $\mathcal{B}_z \subset \mathcal{B}$ are present³. This extends the notion of alignment already introduced in the literature [27] 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 . Overall, this suggests that appropriate non-unitary 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 non-contractive, smaller dimensional ones.

Additionally, we can recover previously known results by considering specific classes of quantum channels \mathcal{E} . First we point out that, being an absorption term, the last term in Eq. (11) vanishes for unitary dynamics, which is reversible by definition. Formally, we have the following Corollary.

Corollary III.1.1 (Deep, unitary circuits). *Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered quantum channel as described in Section II, where $\mathcal{E}(\cdot) = W \cdot W^\dagger$, $W \in U(d)$ is an arbitrary*

² However, this possibility is a rather pathological one.

³ Indeed, \mathcal{B}_z needs to be expressible as a direct sum of both. Failure to achieve this will cause the emergence of noise-induced concentration in the sense of Corollary III.1.2.

unitary transformation. Then we have

$$\mathbb{V}_{\rho,H}^{\infty} = \sum_{z>0} \frac{(\ell_{\rho})_z (\ell_H)_z}{d_z}, \quad (14)$$

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

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. ρ and H need both 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}_{θ} 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 III.1.1, 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

$$\mathbb{V}_{\rho,H}^{\infty} = \frac{(\|\rho\|_2^2 - 1/d) (\|H\|_2^2 - \text{Tr}[H]^2/d)}{d^2 - 1}, \quad (15)$$

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}_{θ} 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. (11) pertains to non-contractive 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.

Corollary III.1.2 (Deep, noisy circuits). *Let $\rho, H \in \mathcal{B}$ and let Φ_{θ} a be layered quantum channel as described in Section II, 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*

$$\mathbb{V}_{\rho,H}^{\infty} = \frac{(A\ell_H)_0}{d}. \quad (16)$$

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

Note that, even if Eq. (16) is inversely proportional to $d = 2^n$, $\mathbb{V}_{\rho,H}^{\infty}$ 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 noise-induced barren plateaus (NIBP). In fact, it is clear that strictly contractive channels with a unique fixed point, fall into the assumptions of Corollary III.1.2, and therefore exhibit some form of concentration. However, Corollary III.1.2 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}_{θ} . 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. (16) vanishes. Contrarily, as pointed out in the literature [31, 33], non-unital channels may avoid the exponential concentration. Here, we show 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} . However, it is crucial to remark that this 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}_{θ} and \mathcal{E} towards the respective, different fixed points. An example of this phenomenon is provided in Section IV. This is in stark contrast with the previously discussed case of \mathcal{B}_z , $z > 0$, as, although more difficult to realize, it keeps a non-trivial dependence on the initial state, and as such it can be said to genuinely avoid concentration if d_z scales appropriately.

So far the results we showed 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 behaviour of $\mathbb{V}_{\rho,H}^L$. 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 $\mathbb{V}_{\rho,H}^L$, we can still use knowledge of the convergence speed to $\mathbb{V}_{\rho,H}^{\infty}$ 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 III.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 $\mathbb{V}_{\rho,H}^{\infty}$, or *effectively shallow*, i.e. the mixing speed β of T is slowed down according to $\beta \in O(\log n/L)$ so that $\mathbb{V}_{\rho,H}^{\infty}$ is never reached, regardless of the scaling of L .

We remark that, when using the above terminology, we never explicitly make reference to the scaling of L with respect to the dimension of the system: depending on the entangling channel \mathcal{E} , even circuits typically considered shallow, e.g. $L \in \Theta(\log n)$, can manifest the behaviour of $\mathbb{V}_{\rho,H}^{\infty}$ if the mixing speed grows fast enough, e.g. $\beta \in \Omega(n/\log n)$. As this speed is related to the amount of non-local 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 non-homogeneous channels.

Theorem III.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 LTM. 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*

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

where $\mathcal{K} : \mathcal{B} \rightarrow \mathcal{B}_K$ 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 III.2.1 (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

$$\mathbb{V}_{\rho, H}^L \geq F(n)(\ell_\rho, \ell_{\mathcal{K}(H)}), \quad (18)$$

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

The conditions of Corollary III.2.1 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 entangling⁴, as the problem size grows. As a notable example, this condition is satisfied for brickwork circuits, equipped with local noise and a local observable [31]. Similarly, condition (b) reflects the absence of concentration for *effectively shallow* quantum channels. A significant example is that of finite local-depth circuits (FLDCs) [36]. 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, which similarly hinge on identity manipulation. In fact, while the primary concern of Theorem III.2 is noise, it could still be regarded as a theoretical foundation of such initialization strategies [20, 21]. In this work we show a deeper relation between these types of smart initializations for noiseless circuits and the properties of noisy layers. This is the subject of the following Section.

⁴ This needs only to apply within the support $\mathcal{B}_H \subset \mathcal{B}$ of the observable H , or, equivalently, the support $\mathcal{B}_\rho \subset \mathcal{B}$ of ρ , in accordance with Eq. (17).

IV. APPLICATIONS

A. Small angle initializations

Supported by the contribution in Section III, here we expand on the concept of small angle initialization introduced in Refs. [20, 21]. In particular, we establish a general relationship between the insights gained from controlling loss concentration in noisy circuits (as presented in Theorem III.2) and BP mitigation strategies that typically only apply to ideal circuits.

First, let us recall the main idea behind small angle initialization strategies. In general, a layered quantum circuit $U_\theta = \prod_l U_{\theta_l}$ is considered, and 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 towards the identity in the quantum model.

A similar structure can be defined in our framework by considering quantum channels $\Phi_{\theta, \phi}$, as in Eq. (2), where the intermediate channels \mathcal{E}_{ϕ_l} are now parameterized. This differs from the main idea of small angle initialization, as the local components \mathcal{U}_{θ_l} of the channel remain Haar random, and instead it is the allowed channels \mathcal{E}_ϕ 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}_ϕ 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 III.2, it can be readily generalized to parameterized channels \mathcal{E}_ϕ , 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}_ϕ^\dagger . Exploiting this, we can derive the following Proposition.

Proposition IV.1 (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 of ϕ . Then, if $\mu = 0$, $\sigma^2 \in O(\log n / \|G\|_2^2 L)$, and $L \in \Omega(\log^{1+\epsilon} n)$ we have*

$$\mathbb{V}_{\rho, H}^L \geq F(n)(\ell_\rho, \ell_H), \quad (19)$$

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

This result represents an application of Corollary III.2.1 (b) in the case of unitary, parameterized intermediate channels. In analogy to Ref. [20], Proposition IV.1 requires σ^2 to decay inversely with L , but it is now independent of the specific distribution. Note that Theorem III.2, which is the backbone of Proposition IV.1, 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 III.2.1 (b) may be analogously interpreted as a condition on the noise rates to avoid concentration. This showcases a connection between QRes-Nets that can avoid BP and noise models, for which the strength is not strong enough to cause NIBP. Formally, this is captured by the following Proposition.

Proposition IV.2 (Noise map and QResNet). *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 to 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*

$$\mathbb{E}_\phi\{T_\phi\} \geq T, \quad (20)$$

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

This result can be interpreted as follows: if a noise map satisfies Corollary III.2.1, then there exists a QRes-Net associated to 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]. Then, there always exist a unitary, stochastic unravelling $\{U_\phi\}_\phi$, such that $\mathcal{E}(\rho) = \mathbb{E}_\phi\{U_\phi\rho U_\phi^\dagger\}$ [40]. From Proposition IV.2, the variance of the action of the ensemble is lower-bounded by the variance of the channel, and hence it provides a BP free QResNet for weak enough noise. Interestingly, Proposition IV.2 can be equally applied if the ensemble is not unitary, extending the framework of small angle initialization to non-unitary quantum models (e.g. those based on linear combination of unitaries (LCU) or analogous techniques [41]).

B. Non-unital noise and entanglement

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

$$\mathcal{E}_c(\rho) = (1-p)\mathcal{E}(\rho) + p\tilde{\rho}, \quad (21)$$

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 to \mathcal{E}_c . Intuitively, we can think of the resulting channel $\tilde{\Phi}_\theta$ as the repetition of L layers, each made up of the composition of $\tilde{\Phi}(\rho) = (1-p)\rho + p\tilde{\rho}$ and of $\mathcal{E} \circ \mathcal{U}_{\theta_i}$. Clearly $\tilde{\rho}$ is the unique fixed-point of $\tilde{\Phi}$, while since 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

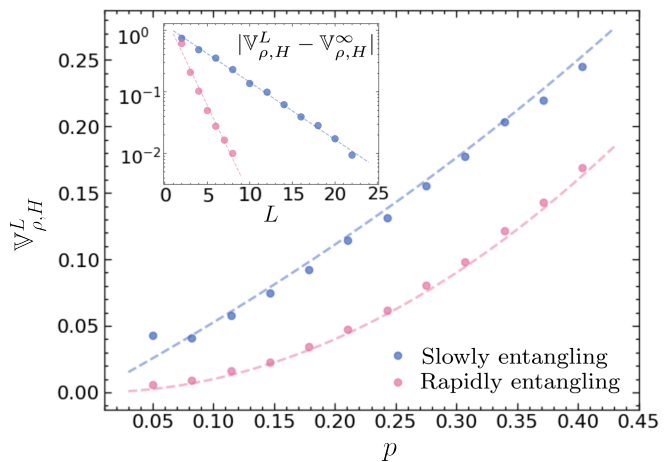


FIG. 3. Scaling of $\mathbb{V}_{\rho,H}^L$ as a function of the noise strength and the entangling power of the intermediate channel. The main figure illustrates the scaling of $\mathbb{V}_{\rho,H}^\infty$ with noise strength p for both rapidly entangling (pink) and slowly entangling (light blue) channels, using $L = 8$ and $L = 20$, respectively. The dotted lines represent the theoretical predictions of Eq. (25) and Eq. (24). The inset verifies the exponential convergence of $\mathbb{V}_{\rho,H}^L$ to $\mathbb{V}_{\rho,H}^\infty$ 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.

variance in such models. However, depending on the relative strength of the two effects, the behaviour of $\mathbb{V}_{\rho,H}^\infty$ as a function of p may vary greatly. Theorem III.1 allows to quantify the variance scaling in the limit of a rapidly entangling and slowly entangling channel, i.e. $\beta \rightarrow \infty$ and $\beta \rightarrow 0$ 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, $\mathbb{V}_{\rho,H}^\infty$ is given by the following Lemma.

Lemma IV.1. *Let \mathcal{E}_c be a quantum channel of type Eq. (21), with $0 < p \leq 1$ and let T be the transition matrix of \mathcal{E}^\dagger . Then*

$$\mathbb{V}_{\rho,H}^\infty = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (\mathbb{1} - p)T)^{-1}\ell_H). \quad (22)$$

In particular, if T is a projection, then

$$\mathbb{V}_{\rho,H}^\infty = \left(\frac{p}{2-p} - p^2\right)(\ell_{\tilde{\rho}}, T\ell_H) + p^2(\ell_{\tilde{\rho}}, \ell_H). \quad (23)$$

The first thing to notice is that the dependence on the initial state of $\mathbb{V}_{\rho,H}^\infty$ is completely lost: the component associated to 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 IV.1 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

$$\mathbb{V}_{\rho,H}^{\infty} = \frac{p}{2-p}(\ell_{\bar{\rho}}, \ell_H). \quad (24)$$

The result shows that first term in Eq. (23) dominates, suggesting that $\mathbb{V}_{\rho,H}^{\infty}$ 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 [31]. 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. (23) dominates, and we get a *quadratic* dependence on p up to an exponentially vanishing correction, namely

$$\mathbb{V}_{\rho,H}^{\infty} = p^2(\ell_{\bar{\rho}}, \ell_H) + O(4^{-n}). \quad (25)$$

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 by the formalism introduced in this work. We utilize PennyLane [42] to construct and optimize the PQC described in Appendix F, and the results are shown in Fig. 3. The plots show both the quadratic and linear scaling with p predicted by our model, as well as an exponential decay in the difference $|\mathbb{V}_{\rho,H}^L - \mathbb{V}_{\rho,H}^{\infty}|$ 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 $\mathbb{V}_{\rho,H}^{\infty}$, while the action of the noise is still too weak to erase the contribution of the first layers, hence deviating from the predicted behaviour. This same phenomenon is not observed in the rapidly entangling case, as entanglement production already exponentially suppresses those contributions.

V. DISCUSSION AND OUTLOOK

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 [15–17], this approach inevitably fails in the general setting of non-unitary circuits, where the group structure description is lost.

In this work, we employed non-negative matrix theory to derive a general formulation for the variance $\mathbb{V}_{\rho,H}^{\infty}$ in the deep circuit limit, for circuit composed of local 2-designs interleaved by arbitrary quantum channels. The

observed structure of $\mathbb{V}_{\rho,H}^{\infty}$ 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 non-contractive ones. This indicates that a mixed configuration of ideal and noisy qubits could potentially outperform purely ideal or purely noisy systems in terms of variance. 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 logical qubits is still constrained [43, 44]. Furthermore, this approach may be beneficial in characterizing variational quantum algorithms executed on hardware, where different qubits may experience varying error rates [39]. In this context, $\mathbb{V}_{\rho,H}^{\infty}$ can be used to approximate $\mathbb{V}_{\rho,H}^L$ in regimes where L is sufficiently large to significantly impact some qubits but not yet others. Additionally, we introduced a general lower bound on the variance of noisy circuits. This bound is derived by restricting the mixing speed of $\mathbb{V}_{\rho,H}^L$ to prevent it from reaching its asymptotic limit.

We subsequently employ this approach to introduce QResNets, an initialization strategy analogous to small angle initializations [18–21] 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 unravellings of sufficiently weak noise maps. Notably, since these unravellings are not necessarily unitary, this approach can yield BP-free architectures beyond unitary circuits, encompassing more complex models such as those employing linear combinations of unitaries (LCU) [41].

Lastly, we numerically investigate the scaling of $\mathbb{V}_{\rho,H}^{\infty}$ in the presence of non-unital noise, demonstrating how the introduction of extensive entanglement can exacerbate noise-induced concentration.

Future research may extend our findings by relaxing the local 2-design property of the unitary layer. This would broaden the applicability of our results to a wider range of quantum circuits, beyond hardware-specific designs. Furthermore, recent studies have established a connection between the absence of concentration and classical simulability for both ideal [45] and noisy [31] 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: our framework offers 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 towards 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).

Appendix A: Locality and locality transfer matrix properties

In order to show the main properties of locality vectors and locality transfer matrices (LTM), 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 $P_0^{(m)} = \mathbb{1}/\sqrt{d_m}$, $P_{j_m}^{(m)} = P_{j_m}^{(m)\dagger} \forall j_m$, $\text{Tr}[P_{j_m}^{(m)} P_{k_m}^{(m)}] = \delta_{j_m k_m}$. 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 labelled by the multi-index $j = (j_1, \dots, j_M)$, where the entries j_m refer to an element of the local bases, and hence $j_m \in \{0, \dots, d_m^2 - 1\}$. 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 non trivially 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

$$\tilde{\delta}_{i,\kappa} = \prod_{m=0}^M \tilde{\delta}_{i_m, \kappa_m} = \begin{cases} 1 & \text{if } \kappa_m = 0 \Leftrightarrow i_m = 0 \forall m \\ 0 & \text{otherwise} \end{cases} \quad (\text{A1})$$

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

Lemma A.1. *The indicator function $\tilde{\delta}_{i,b}$ has the following properties:*

$$\sum_i \tilde{\delta}_{i,\kappa} = d_\kappa, \quad \sum_\kappa \tilde{\delta}_{i,\kappa} = 1, \quad \sum_{\kappa \in K} \tilde{\delta}_{i,\kappa} \tilde{\delta}_{i,\lambda} = \tilde{\delta}_{i,\lambda} \sum_{\kappa \in K} \delta_{\kappa,\lambda} \quad (\text{A2})$$

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. \square

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 $\Lambda : \mathcal{B} \rightarrow \mathcal{B}$ defined in the main text as

$$(\ell_A)_\kappa = \sum_j \text{Tr}[P_j A]^2 \tilde{\delta}_{j,\kappa} \quad (\text{A3})$$

and

$$T_{\kappa,\lambda} = \frac{1}{d_\lambda} \sum_{i,j} \text{Tr}[P_i \Lambda(P_j)]^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} \quad (\text{A4})$$

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

Lemma A.2. *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 $\mathcal{E} : \mathcal{B} \rightarrow \mathcal{B}$. 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

$$\begin{aligned} (\ell_A)_\kappa &= \sum_j \text{Tr}[P_j A]^2 \tilde{\delta}_{j,\kappa} = \sum_j \text{Tr} \left[\sum_i \text{Tr}[B_i P_j] B_i A \right]^2 \tilde{\delta}_{j,\kappa} \\ &= \sum_j \sum_{i,i'} \text{Tr}[B_i P_j] \text{Tr}[B_{i'} P_j] \text{Tr}[B_i A] \text{Tr}[B_{i'} A] \tilde{\delta}_{j,\kappa} \\ &= \sum_{i,i'} \left(\sum_j \text{Tr}[B_i P_j] \text{Tr}[B_{i'} P_j] \tilde{\delta}_{j,\kappa} \right) \text{Tr}[B_i A] \text{Tr}[B_{i'} A] \\ &= \sum_i \text{Tr}[B_i A]^2 \tilde{\delta}_{i,\kappa} \end{aligned} \quad (\text{A5})$$

where the last equality is due to

$$\begin{aligned} &\sum_j \text{Tr}[B_i P_j] \text{Tr}[B_{i'} P_j] \tilde{\delta}_{j,\kappa} \\ &= \prod_{m=1}^M \left(\sum_{j_m=1}^{d_m} \text{Tr}[B_{i'_m} P_{j_m}] \text{Tr}[B_{i_m} P_{j_m}] \tilde{\delta}_{j_m, \kappa_m} \right) \quad (\text{A6}) \\ &= \prod_{m=1}^M (\text{Tr}[B_{i'_m} B_{i_m}] \tilde{\delta}_{i_m, \kappa_m}) = \delta_{i,i'} \tilde{\delta}_{i,\kappa} \end{aligned}$$

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 . \square

Thanks to the formulation of Eq. (A4), the relation between the LTM of a map Λ and the Hermitian adjoint Λ^\dagger with respect to the Hilbert-Schmidt scalar product can be seen. In particular, we have

$$\begin{aligned} d_\lambda T_{\kappa,\lambda} &= \sum_{i,j} \text{Tr}[P_i \Lambda(P_j)]^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} \\ &= \sum_{i,j} \text{Tr}[\Lambda^\dagger(P_i) P_j]^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} = d_\kappa T_{\lambda,\kappa}^\dagger, \end{aligned} \quad (\text{A7})$$

which can be compactly written in matrix form as $TD = (T^\dagger D)^t$, with $D_{\kappa,\lambda} = d_\kappa \delta_{\kappa,\lambda}$. For 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.

$$(a, b) = a^t D^{-1} b = \sum_{\kappa} \frac{a_\kappa b_\kappa}{d_\kappa}. \quad (\text{A8})$$

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

Appendix B: Proof of Proposition III.1 and Proposition III.2

In this section we provide a proof for the building blocks of the main results of this work. Note that, what follows hinge on the structure of the circuit Φ_θ provided in the main text, which we recall is composed of L layers of interleaved unitary and noise channels

$$\Phi_\theta = \mathcal{U}_{\theta_{L+1}} \circ \mathcal{E}_L \circ \mathcal{U}_{\theta_{L-1}} \cdots \circ \mathcal{E}_1 \circ \mathcal{U}_{\theta_1}. \quad (\text{B1})$$

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

Definition B.1. (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. $U_\theta = \otimes_m U_{\theta_m}^{(m)}$ each acting solely on \mathcal{H}_m , and additionally

$$\int_{\Theta} d\theta U_\theta^{(m) \otimes t} \otimes U_\theta^{(m) * \otimes t} = \int_{V \in U(d_m)} d\mu(V) V^{(m) \otimes t} \otimes V^{(m) * \otimes t} \quad (\text{B2})$$

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 B.1 (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*

$$\mathbb{E}_\theta \{ \text{Tr} [AU_\theta^\dagger B U_\theta] \} = \frac{\text{Tr} [A] \text{Tr} [B]}{d} \quad (\text{B3})$$

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 $a_i = \text{Tr} [P_i A]$, and consequently $a_0 = \frac{1}{\sqrt{d}} \text{Tr} [A]$ (respectively for B). Then we have the

following chain of equalities:

$$\begin{aligned} & \int_{\Theta} \prod_{m=1}^M d\theta \text{Tr} [AU_\theta^\dagger B U_\theta] \\ &= \sum_{i,j} a_i b_j \int_{\Theta} \prod_{m=1}^M d\theta \text{Tr} [P_i U_\theta^\dagger P_j U_\theta] \\ &= \sum_{i,j} a_i b_j \prod_{m=1}^M \int_{U \in U(d_m)} d\mu(U) \text{Tr}_m [P_{i_m}^{(m)} U^\dagger P_{j_m}^{(m)} U] \\ &= \sum_{i,j} a_i b_j \prod_{m=1}^M \frac{1}{d_m} \text{Tr}_m [P_{i_m}^{(m)}] \text{Tr}_m [P_{j_m}^{(m)}] \\ &= \sum_{i,j} a_i b_j \prod_{m=1}^M \delta_{0,i_m} \delta_{0,j_m} = a_0 b_0 \end{aligned} \quad (\text{B4})$$

where Tr_m denotes the partial trace over the m -th subsystem. \square

Regarding the second moment, it can be computed using the following Lemma.

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

$$\begin{aligned} & \mathbb{E}_\theta \{ \text{Tr} [P_i U_\theta^\dagger P_j U_\theta] \text{Tr} [P_k U_\theta^\dagger P_l U_\theta] \} = \\ &= \delta_{ij} \delta_{kl} \prod_{m=1}^M \left(\delta_{0,i_m} \delta_{0,j_m} + (1 - \delta_{0,i_m})(1 - \delta_{0,j_m}) \frac{1}{d_m^2 - 1} \right) \end{aligned} \quad (\text{B5})$$

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

$$\begin{aligned} & \int_{U \in U(d)} d\mu(U) \text{Tr} [AU^\dagger B U] \text{Tr} [CU^\dagger D U] \\ &= \frac{1}{d^2 - 1} (\text{Tr} [A] \text{Tr} [B] \text{Tr} [C] \text{Tr} [D] + \text{Tr} [AC] \text{Tr} [BD]) \\ &+ \frac{1}{d(d^2 - 1)} (\text{Tr} [AC] \text{Tr} [C] \text{Tr} [D] + \text{Tr} [A] \text{Tr} [C] \text{Tr} [BD]) \end{aligned} \quad (\text{B6})$$

Based on Eq. (B6), the result follows from direct integration:

$$\begin{aligned} & \prod_{m=1}^M \int d\mu(U) \text{Tr} [P_{i_m}^{(m)} U^\dagger P_{j_m}^{(m)} U] \text{Tr} [P_{k_m}^{(m)} U^\dagger P_{l_m}^{(m)} U] \\ &= \prod_{m=1}^M \frac{1}{d_m^2 - 1} \left(\text{Tr} [P_{i_m}^{(m)}] \text{Tr} [P_{j_m}^{(m)}] \text{Tr} [P_{k_m}^{(m)}] \text{Tr} [P_{l_m}^{(m)}] \right. \\ &+ \text{Tr} [P_{i_m}^{(m)} P_{k_m}^{(m)}] \text{Tr} [P_{j_m}^{(m)} P_{l_m}^{(m)}] \left. \right) \\ &- \frac{1}{d_m(d_m^2 - 1)} \left(\text{Tr} [P_{i_m}^{(m)} P_{k_m}^{(m)}] \text{Tr} [P_{j_m}^{(m)}] \text{Tr} [P_{l_m}^{(m)}] \right. \\ &+ \text{Tr} [P_{i_m}^{(m)}] \text{Tr} [P_{k_m}^{(m)}] \text{Tr} [P_{j_m}^{(m)} P_{l_m}^{(m)}] \left. \right) \end{aligned} \quad (\text{B7})$$

which exploiting the orthogonality conditions becomes:

$$\begin{aligned}
&= \prod_{m=1}^M \frac{1}{d_m^2 - 1} (d_m^2 \delta_{0i_m} \delta_{0j_m} \delta_{0k_m} \delta_{0l_m} + \delta_{i_m k_m} \delta_{j_m l_m} + \\
&\quad - \delta_{0i_m} \delta_{0k_m} \delta_{j_m l_m} - \delta_{i_m k_m} \delta_{0j_m} \delta_{0l_m}) \\
&= \prod_{m=1}^M \delta_{0i_m} \delta_{0j_m} \delta_{0k_m} \delta_{0l_m} + \\
&\quad + \frac{1}{d_m^2 - 1} (\delta_{i_m k_m} - \delta_{0i_m} \delta_{0k_m}) (\delta_{j_m l_m} - \delta_{0j_m} \delta_{0l_m}) \\
&= \delta_{jk} \delta_{jl} \prod_{m=1}^M \left(\delta_{0i_m} \delta_{0j_m} + (1 - \delta_{0i_m})(1 - \delta_{0j_m}) \frac{1}{d_m^2 - 1} \right) \tag{B8}
\end{aligned}$$

concluding the proof. \square

In particular, Lemma B.2 can be used to compute the variance of loss functions computed as expectation values $\mathcal{L}_{\rho, H} = \text{Tr}[U_\theta(A)B]$, i.e. in the absence of intermediate channels \mathcal{E}_i .

Proposition B.1. *Let $A, B \in \mathcal{B}$, and let $\{U_\theta\}_{\theta \in \Theta}$ be a unitary ensemble forming a local 2-design. Then*

$$\mathbb{E}_\theta \left\{ \text{Tr} [AU_\theta^\dagger BU_\theta]^2 \right\} = (\ell_A, \ell_B), \tag{B9}$$

where (\cdot, \cdot) is the scalar product defined in Eq. (A8).

Proof. Let $\{P_j\}$ be a local basis for the system, and consider the respective decompositions of A and B . By Lemma B.2 we have

$$\begin{aligned}
\mathbb{E}_\theta \left\{ \text{Tr} [AU_\theta^\dagger BU_\theta]^2 \right\} &= \sum_{i,j} a_i^2 b_j^2 \prod_{m=1}^M \left(\delta_{0i_m} \delta_{0j_m} + \right. \\
&\quad \left. + (1 - \delta_{0i_m})(1 - \delta_{0j_m}) \frac{1}{d_m^2 - 1} \right) \tag{B10}
\end{aligned}$$

In the following, it will be convenient to recast the product on the right-hand side of Eq. (B10) into the equivalent formulation

$$\begin{aligned}
&\prod_{m=1}^M \left(\delta_{0i_m} \delta_{0j_m} + (1 - \delta_{0i_m})(1 - \delta_{0j_m}) \frac{1}{d_m^2 - 1} \right) = \\
&= \sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa} \prod_{m=1}^M (\delta_{0i_m} \delta_{0j_m})^{1-\kappa_m} (1 - \delta_{0i_m})^{\kappa_m} (1 - \delta_{0j_m})^{\kappa_m} \tag{B11}
\end{aligned}$$

where the binary vectors $\kappa \in \{0,1\}^M$ identify all possible sets S_κ introduced in Appendix A and $d_\kappa = \prod_{m=1}^M (d_m^2 - 1)^{\kappa_m}$. Putting it back into Eq. (B10) we get

$$\begin{aligned}
&\sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa} \sum_{i,j} a_i^2 b_j^2 \prod_{m=1}^M (\delta_{0i_m} \delta_{0j_m})^{1-\kappa_m} (1 - \delta_{0i_m})^{\kappa_m} (1 - \delta_{0j_m})^{\kappa_m} \\
&= \sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa} \left(\sum_i a_i^2 \prod_{m=1}^M \delta_{0i_m}^{1-\kappa_m} (1 - \delta_{0i_m})^{\kappa_m} \right) \times \\
&\quad \times \left(\sum_j b_j^2 \prod_{m=1}^M \delta_{0j_m}^{1-\kappa_m} (1 - \delta_{0j_m})^{\kappa_m} \right) \\
&= \sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa} \left(\sum_i a_i^2 \tilde{\delta}_{i,\kappa} \right) \left(\sum_j b_j^2 \tilde{\delta}_{j,\kappa} \right) \\
&= \sum_{\kappa \in \{0,1\}^M} \frac{(\ell_A)_\kappa (\ell_B)_\kappa}{d_\kappa} = (\ell_A, \ell_B) \tag{B12}
\end{aligned}$$

from which the Proposition follows. \square

Applying Proposition B.1 for an initial state ρ and an observable H yields Proposition III.1 of the main text. This can be extended to the general case introducing the action of the intermediate channels \mathcal{E}_i , and in particular we get the following Proposition.

Proposition B.2. *Let $A, B \in \mathcal{B}$ and $\Lambda : \mathcal{B} \rightarrow \mathcal{B}$ 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*

$$\mathbb{E}_{\theta_1, \theta_2} \left\{ \text{Tr} [AU_{\theta_1}^\dagger \Lambda(V_{\theta_2} B V_{\theta_2}^\dagger) U_{\theta_1}]^2 \right\} = (\ell_A, T \ell_B) \tag{B13}$$

where T is the locality transfer matrix associated to Λ .

Proof. Let $\tilde{B}_{\theta_2} = \Lambda(V_{\theta_2} B V_{\theta_2}^\dagger)$. By Proposition B.2, we have

$$\begin{aligned}
\mathbb{E}_{\theta_2} \mathbb{E}_{\theta_1} \left\{ \text{Tr} [AU_{\theta_1}^\dagger B_{\theta_2} U_{\theta_1}]^2 \right\} &= \mathbb{E}_{\theta_2} \left\{ (\ell_A, \ell_{\tilde{B}_{\theta_2}}) \right\} \\
&= \sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa} (\ell_A)_\kappa \mathbb{E}_{\theta_2} \left\{ (\ell_{\tilde{B}_{\theta_2}})_\kappa \right\} \tag{B14}
\end{aligned}$$

Expanding the definition on the last term with respect to the local basis $\{P_j\}_j$, and applying again Proposition B.2 we get

$$\begin{aligned}
\mathbb{E}_{\theta_2} \left\{ (\ell_{\tilde{B}_{\theta_2}})_{\kappa} \right\} &= \sum_i \mathbb{E}_{\theta_2} \left\{ \text{Tr} [P_i \tilde{B}_{\theta_2}]^2 \right\} \tilde{\delta}_{i,\kappa} \\
&= \sum_i \mathbb{E}_{\theta_2} \left\{ \text{Tr} [\Lambda^\dagger(P_i) V_{\theta_2}^\dagger B V_{\theta_2}]^2 \right\} \tilde{\delta}_{i,\kappa} \\
&= \sum_{\lambda \in \{0,1\}^M} \frac{1}{d_\lambda} \sum_{i,j,k} \text{Tr} [P_i \Lambda(P_j)]^2 \text{Tr} [P_k B]^2 \tilde{\delta}_{j,\lambda} \tilde{\delta}_{k,\lambda} \tilde{\delta}_{i,\kappa} \\
&= \sum_{\lambda \in \{0,1\}^M} \left(\frac{1}{d_\lambda} \sum_{i,j} \text{Tr} [P_i \Lambda(P_j)]^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} \right) \times \\
&\quad \times \left(\sum_k \text{Tr} [P_k B]^2 \tilde{\delta}_{k,\lambda} \right) \\
&= \sum_{\lambda \in \{0,1\}^M} T_{\kappa,\lambda} (\ell_B)_\lambda = (T \ell_B)_\kappa
\end{aligned} \tag{B15}$$

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

Iterated application of Proposition B.2 for an initial state ρ , an observable H , and a general intermediate quantum channel \mathcal{E} , yields Proposition III.2 of the main text.

Appendix C: Proof of Theorem III.1

The proof of Theorem III.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 sake of clarity and completeness we recall them in the following.

1. Preliminaries

In this section, 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 section, we briefly recap on the main results on non-negative matrix theory useful in the proof of Theorem III.1 of the main text. 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 C.1 (Non-negative matrix). A $n \times n$ matrix T is said to be non-negative if each entry $(T)_{ij} \geq 0$.

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

Definition C.2 (Irreducible matrix). A $n \times n$ non-negative matrix T is said to be *irreducible* if for two arbitrary indices $i, j = 1, \dots, n$, 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 C.1 (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. $v_i, w_i > 0 \forall i = 1, \dots, n$,
- (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 $P = w v^t$ 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 C.2 (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*

$$y^t T \leq s y^t \tag{C1}$$

component-wise. 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 C.3 (Asymptotic behaviour of irreducible matrices). *Let T be a $n \times n$ non-negative, irreducible matrix. Then the Cesàro average of T converges, and we have*

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L T^l / r^l = P \quad (\text{C2})$$

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

$$\lim_{L \rightarrow \infty} T^L / r^L = P \quad (\text{C3})$$

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 relabelling 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. (C4).

$$T = \begin{pmatrix} \boxed{T_0} & & & & & & \boxed{R_0} \\ & \boxed{T_1} & & & & & \boxed{R_1} \\ & & \ddots & & & & \vdots \\ & & & \boxed{T_z} & & & \boxed{R_z} \\ & & & & & & \boxed{Q} \end{pmatrix} \quad (\text{C4})$$

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 section, 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 C.3. (Schatten norm) Given $A \in \mathbb{M}_n$ and $p \in [1, \infty]$, we define the Schatten p -norm $\|\cdot\|_p : \mathbb{M}_n \rightarrow \mathbb{R}$ as

$$\|A\|_p = \text{Tr} \left[\left(\sqrt{A^\dagger A} \right)^p \right]^{1/p}. \quad (\text{C5})$$

A crucial property of Schatten norms is Hölder inequality, namely $|\text{Tr} [A^\dagger B]| \leq \|A\|_p \|B\|_q$, $\forall A, B \in \mathbb{M}_n, \forall p, q$ 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 C.3, an induced norm on linear operators acting on \mathbb{M}_n can be defined.

Definition C.4. (Induced norm) Given a linear operator $\Lambda : \mathbb{M}_n \rightarrow \mathbb{M}_n$ we define the induced $p \rightarrow q$ norm as

$$\|\Lambda\|_{p \rightarrow q} := \sup_{A: \|A\|_p = 1} \|\Lambda(A)\|_q. \quad (\text{C6})$$

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 C.5. (Contractivity of linear maps) A linear operator $\Lambda : \mathbb{M}_n \rightarrow \mathbb{M}_n$ is said to be *contractive* with respect to the p norm if $\|\Lambda\|_{p \rightarrow p} \leq 1$, and similarly to be *strictly contractive* if $\|\Lambda\|_{p \rightarrow p} < 1$.

Linear maps that are also CPTP are known to always be contractive with respect to the 1-norm [46, 47], but are in general not contractive for other p -norms. This property, together with Hölder’s inequality, allow putting an upper bound on the value of the variance of an arbitrary, layered quantum circuit.

Lemma C.4 (Hölder’s inequality for variances). *Let $A, B \in \mathbb{M}_n$ be hermitian matrices, and $\Phi_\theta : \mathbb{M}_n \rightarrow \mathbb{M}_n$ be a parameterized quantum channel. In particular, consider the L layered map of type Eq. (B1). Then*

$$|\text{Tr} [\Phi_\theta^L(A)B]| \leq \|\mathcal{E}\|_{p \rightarrow p}^L \|A\|_p \|B\|_q \quad \forall \theta \in \Theta \quad (\text{C7})$$

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:

$$\begin{aligned}
|\mathrm{Tr} [\Phi_\theta^L(A)B]| &\leq \|\Phi_\theta^L(A)\|_p \|B\|_q \\
&= \|U_{\theta_L} \mathcal{E}(\Phi_\theta^{L-1}(A)) U_{\theta_L}^\dagger\|_p \|B\|_q \\
&= \|\mathcal{E}(\Phi_\theta^{L-1}(A))\|_p \|B\|_q \\
&\leq \|\mathcal{E}\|_{p \rightarrow p} \|\Phi_\theta^{L-1}(A)\|_p \|B\|_q \quad \forall \theta \in \Theta
\end{aligned} \tag{C8}$$

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. \square

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 [48]). 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. $\|\mathcal{N}\|_{\mathbb{H}_0} \|p \rightarrow p\| \leq 1 \forall p$ if $n = 2$. For single qubit channels we can even be more explicit, as showed in the following Lemma.

Lemma C.5 (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*

$$\mathcal{N}'(P_0) = P_0 + \sum_{i>0} t_i P_i, \quad \mathcal{N}'(P_i) = \lambda_i P_i \quad \forall i > 0 \tag{C9}$$

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 [49, 50] that any single qubit quantum channel can be cast in the canonical form of Eq. (C9) by means of a change of basis. Furthermore, the constraint on the parameters follow, analogously to [31], by considering that any single qubit state must have bounded purity, namely $\mathrm{Tr}[\rho^2] \leq 1$, 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

$$\mathrm{Tr}[\rho^2] = \frac{1 + \alpha_i^2}{2} \leq 1, \quad \mathrm{Tr}[\mathcal{N}'(\rho)^2] = \frac{1 + \sum_{i>0} (t_i + \lambda_i \alpha_i)^2}{2} \leq 1 \tag{C10}$$

respectively, which concludes the proof. \square

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

Theorem C.6 (Kadison-Schwarz inequality [51]). *Let $A, B \in \mathbb{H}_0$, and $\Lambda : \mathbb{M}_n \rightarrow \mathbb{M}_n$ be a CPU map. Then*

$$\Lambda(A)\Lambda(B) \leq \Lambda(AB). \tag{C11}$$

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 $\mathbb{V}_{\rho, H}^\infty$ 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 C.7 (Limiting form of T). *Let $\Lambda : \mathcal{B} \rightarrow \mathcal{B}$ be a CPU map and T be the corresponding LTM. Then T and T^L take the form*

$$T = \begin{pmatrix} \boxed{T_0} & & & & \boxed{R_0} \\ & \boxed{T_1} & & & \boxed{R_1} \\ & & \ddots & & \vdots \\ & & & \boxed{T_z} & \boxed{R_z} \\ & & & & \boxed{Q} \end{pmatrix}, \tag{C12}$$

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

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

Proof. We start by putting T into the canonical form of Eq. (C4). 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

$$\begin{aligned}
A_z^{(2)} &= T_z R_z + R_z Q \\
A_z^{(L)} &= T_z^{L-1} R_z + A^{(L-1)} Q \\
&= T_z^{L-1} R_z + \sum_{l=0}^{L-2} T_z^l R_z Q^{L-1-l} = \sum_{l=0}^{L-1} T_z^l R_z Q^{L-1-l}
\end{aligned} \tag{C13}$$

which gives the proposition. \square

As already shown in Lemma C.4, the value of the variance is upper bounded by $\|H\|_\infty^2$. Since by Proposition B.2, 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 C.1 (Spectral radius of T). *Let $\Lambda : \mathcal{B} \rightarrow \mathcal{B}$ 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, also Q can be cast in canonical block upper triangular form by means of a basis permutation, where each diagonal block is irreducible.

$$Q = \begin{pmatrix} \boxed{Q_1} & \boxed{*} & \dots & \boxed{*} \\ & \boxed{Q_2} & & \boxed{*} \\ & & \ddots & \vdots \\ & & & \boxed{Q_k} \end{pmatrix} \tag{C14}$$

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 A.2, 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 $P_i^2 = \mathbf{1}/d \forall i$. Exploiting Eq. (A4), we can compute the column sum of T as

$$\begin{aligned}
\sum_\kappa T_{\kappa,\lambda} &= \sum_\kappa \frac{1}{d_\lambda} \sum_{i,j} \text{Tr} [P_i \Lambda(P_j)]^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} \\
&= \frac{1}{d_\lambda} \sum_{i,j} \text{Tr} [P_i \Lambda(P_j)]^2 \tilde{\delta}_{j,\lambda} \left(\sum_\kappa \tilde{\delta}_{i,\kappa} \right) \\
&= \frac{1}{d_\lambda} \sum_{i,j} \text{Tr} [P_i \Lambda(P_j)]^2 \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \text{Tr} [\Lambda(P_j)^2] \tilde{\delta}_{j,\lambda} \\
&\leq \frac{1}{d_\lambda} \sum_j \text{Tr} [\Lambda(P_j^2)] \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \tilde{\delta}_{j,\lambda} = 1
\end{aligned} \tag{C15}$$

where the third and last equality follow from Lemma A.1, 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. (C15) show *sub-stochasticity* of T . Indeed, this condition can be recast in vector form as $v^t T \leq v^t$, where $v_\kappa = \mathbf{1} \forall \kappa$. 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

$$\sum_\kappa (T - Q_k)_{\kappa,\lambda} > 0, \tag{C16}$$

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 $r_{Q_k} < 1$. Putting everything together, one gets $\rho(Q) < 1$. \square

When analysing 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 C.7.1. *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 $(v_z)_\kappa = \mathbf{1} \forall \kappa$ 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. \square

Intuitively, the blocks which 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 C.2. *Let T be a LTM, and let T_z be an irreducible block with $r_z < 1$. Then, as $L \rightarrow \infty$, $\|T_z^L\| \rightarrow 0$ and $\|A_z^{(L)}\| \rightarrow 0$ exponentially fast for any matrix norm $\|\cdot\|$. Similarly, also $\|Q^L\| \rightarrow 0$.*

Proof. The proposition can be proven using Gelfand's formula. In particular since $\lim_{L \rightarrow \infty} \|T_z^L\|^{1/L} = r_z$, we can always bound $\|T_z^L\| \leq K\tau^L$, for some constant $K > 0$ and $\tau = r_z + \epsilon < 1$ for an arbitrarily small ϵ . In the same way, by Proposition C.1 a similar result can be obtained for Q . Finally, the absorption term $A_z^{(L)}$ can also be bounded using Lemma C.7. In that case we have

$$\begin{aligned} \|A_z^{(L)}\| &\leq \sum_{l=0}^{L-1} \|T_z^L\| \|R_z\| \|Q^{L-1-l}\| \\ &\leq \|R_z\| K_T K_Q \tau^l \kappa^{L-1-l} \leq K_A \alpha^L \end{aligned} \quad (\text{C17})$$

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 sub-additivity and sub-multiplicativity of the matrix norm $\|\cdot\|$. \square

3. Proof of Theorem III.1

As stated in the main text, the limiting value of quantum circuits of type Eq. (B1) 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

$$\mathbb{V}_{\rho, H}^\infty = \sum_{z>0} (\ell_\rho, w_z) (\ell_H)_z + (\ell_\rho, w_z) (A_z \ell_H) \quad (\text{C18})$$

with some normalized, strictly positive vector w_z , and absorption matrices A_z . Indeed, the following shows that this is the case.

Theorem C.8 (Deep circuit variance). *Let $\rho, H \in \mathcal{B}$ and let Φ_θ a be layered quantum channel as in described in the main text. Then the Cesàro average of $\mathbb{V}_{\rho, H}^L$ converges, and we have*

$$\left| \frac{1}{L} \sum_{l=0}^L \mathbb{V}_{\rho, H}^l - \mathbb{V}_{\rho, H}^\infty \right| \in O(e^{-\beta L} \|H\|_2^2), \quad (\text{C19})$$

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

$$|\mathbb{V}_{\rho, H}^L - \mathbb{V}_{\rho, H}^\infty| \in O(e^{-\beta L} \|H\|_2^2), \quad (\text{C20})$$

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

Proof. Thanks to Proposition C.2, 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 non-negative 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 $\lim_{N \rightarrow \infty} T_z^{dN} = T_z^{(d\infty)}$ exists, this does not imply that $\lim_{L \rightarrow \infty} T_z^L$ does. In fact, different subsequences might have different limiting values, and in particular $\lim_{N \rightarrow \infty} T_z^{dN+m} = T_z^{(d\infty)} T^m$, which is different for all $m = 0, \dots, d-1$. In the periodic scenario then, T_z^L does not have a limit, and the only convergent quantity is the Cesàro average, i.e.

$$P_z = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L T_z^l = T_z^{(d\infty)} \frac{1}{d} \sum_{m=0}^{d-1} T_z^m \quad (\text{C21})$$

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

Proposition C.3. *Given a LTM with a periodic irreducible block T_z of period d , then*

$$\lim_{N \rightarrow \infty} A^{(dN+m)} = T^{(d\infty)} A^{(m)} + A^{(d\infty)} Q^m \quad (\text{C22})$$

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

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

$$\begin{aligned} A^{(dN)} &= \sum_{l=0}^{Nd-1} T_z^l R_z Q^{dN-1-l} \\ &= \sum_{m=0}^{d-1} \sum_{n=0}^{N-1} T_z^{nd+m} R_z Q^{(N-1)d-nd+d-1-m} \\ &= \sum_{m=0}^{d-1} \sum_{n=0}^{N-1} T_z^{nd} T_z^m R_z Q^{d-1-m} (Q^d)^{N-n} \\ &= \sum_{n=0}^{N-1} T_z^{nd} A^{(d)} (Q^d)^{N-n} \\ &= T_z^{(d\infty)} A^{(d)} \sum_{n=0}^{N-1} (Q^d)^{N-n} + \sum_{n=0}^{N-1} \Delta_z^{(n)} A^{(d)} (Q^d)^{N-n} \end{aligned} \quad (\text{C23})$$

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 $T_{zm}^N \rightarrow P_{zm}$ 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, $\|\Delta_z^{(N)}\| \leq \sum_{m=0}^d \|\Delta_{zm}^{(N)}\| \leq K_z \tau_z^N$, where $\tau = \max_m \{\tau_{zm}\} < 1$. Together with Proposition C.1, this implies that, the last term in Eq. (C23) approaches zero with the same exponential speed, similarly to what happens in Proposition C.2. Putting everything together,

and considering that $\sum_{n=0}^{N-1} X^n \rightarrow (\mathbb{1} - X)^{-1} \forall X$ such that $\rho(X) < 1$, we have

$$A^{(d\infty)} = T^{(d\infty)} A^{(d)} (\mathbb{1} - Q^d)^{-1} \quad (\text{C24})$$

At this point, by induction similarly to Lemma C.7, one can easily show that

$$A^{(dN+m)} = T^{(dN)} A^{(m)} + A^{(dN)} Q^m \quad (\text{C25})$$

which yields the proposition. \square

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

$$A_z = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L A^{(l)} = \frac{1}{d} \sum_{m=0}^{d-1} T^{(d\infty)} A^{(m)} + A^{(d\infty)} Q^m. \quad (\text{C26})$$

Recalling that the right eigenvector v_z can be explicitly calculated when $\rho(T_z) = 1$ (see Corollary C.7.1), 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 Eq. (C19) and Eq. (C20) 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

$$\begin{aligned} \left| \frac{1}{L} \sum_{l=0}^L \mathbb{V}_{\rho, H}^l - \mathbb{V}_{\rho, H}^\infty \right| &= \left| \left(\ell_\rho, \frac{1}{L} \sum_{l=0}^L T^l - T^\infty \ell_H \right) \right| \\ &\leq \left\| \frac{1}{L} \sum_{l=0}^L T^l - T^\infty \right\| \sqrt{(\ell_\rho, \ell_\rho)(\ell_H, \ell_H)}. \end{aligned} \quad (\text{C27})$$

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

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 C.4 (Single qubit noise). *Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered quantum channel as in described in the main text. 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 $\mathcal{N} = \mathcal{N}_1 \otimes \dots \otimes \mathcal{N}_n$ is a composition of single qubit quantum channels. Then*

$$\mathbb{V}_{\rho, H}^\infty = \sum_z \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z (A\ell_H)_z}{d_z}. \quad (\text{C28})$$

Proof. Without loss of generality, we can consider the single qubit channels \mathcal{N}_m to be in their normal form of Lemma C.5. In particular, this holds due to the invariance of the LTM with respect to changes of local bases (Lemma A.2). 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 $\text{Tr}[\mathcal{N}_m^\dagger(P_{j_m})^2] = t_{j_m}^2 + \lambda_{j_m}^2 \leq 1$ by Lemma C.5. 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 C.7.1 we know that $\sum_\kappa (T_z)_{\kappa, \lambda} = 1$. By Eq. (C15) this implies

$$\begin{aligned} \sum_\kappa (T_z)_{\kappa, \lambda} &= \frac{1}{d_\lambda} \sum_j \text{Tr}[(W^\dagger \mathcal{N}^\dagger(P_j) W)^2] \tilde{\delta}_{j, \lambda} \\ &= \frac{1}{d_\lambda} \sum_j \text{Tr}[\mathcal{N}^\dagger(P_j)^2] \tilde{\delta}_{j, \lambda} \\ &= \frac{1}{d_\lambda} \sum_j \prod_{m=1}^M \text{Tr}[\mathcal{N}_m^\dagger(P_{j_m})^2] \tilde{\delta}_{j, \lambda} = 1. \end{aligned} \quad (\text{C29})$$

Since each term in the product is upper-bounded by 1, Eq. (C29) implies $\text{Tr}[\mathcal{N}_m^\dagger(P_{j_m})^2] = t_{j_m}^2 + \lambda_{j_m}^2 = 1 \forall j_m$. This condition is only compatible with Lemma C.5 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 $W P_j W^\dagger$ with respect to the normalized Pauli basis, we can get

$$\begin{aligned} \sum_\kappa (T_z)_{\kappa, \lambda}^\dagger &= \frac{1}{d_\lambda} \sum_j \text{Tr}[\mathcal{N}(W P_j W^\dagger)^2] \tilde{\delta}_{j, \lambda} \\ &= \frac{1}{d_\lambda} \sum_j \text{Tr} \left[\left(\sum_i \text{Tr}[P_i W P_j W^\dagger] \mathcal{N}(P_i) \right)^2 \right] \tilde{\delta}_{j, \lambda} \\ &= \frac{1}{d_\lambda} \sum_j \text{Tr} \left[\left(\sum_i \text{Tr}[P_i W P_j W^\dagger] \prod_m \lambda_{i_m} P_i \right)^2 \right] \tilde{\delta}_{j, \lambda} \\ &= \frac{1}{d_\lambda} \sum_{i, j} \prod_m \lambda_{i_m}^2 \text{Tr}[P_i W P_j W^\dagger]^2 \tilde{\delta}_{j, \lambda} \\ &= \frac{1}{d_\lambda} \sum_j \text{Tr}[W P_j^2 W^\dagger] \tilde{\delta}_{j, \lambda} = 1. \end{aligned} \quad (\text{C30})$$

This allows to compute the right eigenvector w_z of the leading eigenvalue of T_z . Using Eq. (A7), we have indeed

$$\sum_\lambda (T_z)_{\kappa, \lambda} d_\lambda = \sum_\lambda (T_z)_{\lambda, \kappa}^\dagger d_\kappa = d_\kappa \quad (\text{C31})$$

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. \square

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

Corollary C.8.1 (Unitary circuits). *For unitary circuits of type Eq. (B1), we have*

$$\mathbb{V}_{\rho,H}^{\infty} = \sum_{z>0} \frac{(\ell_{\rho})_z (\ell_H)_z}{d_z} \quad (\text{C32})$$

Proof. This form of the variance is a special case of Proposition C.4, putting $\mathcal{N}(\rho) = \rho$, and noting that the absorption terms must vanish. In particular, this follows from Eq. (C15), observing that unitary channels saturate Kadison-Schwarz inequality, which combined with Corollary C.7.1 imply $Q = 0$. \square

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 there by trace preservation.

Corollary C.8.2 (Noise-induced concentration). *Let \mathcal{E} be a quantum channel, 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*

$$\mathbb{V}_{\rho,H}^{\infty} = \frac{(A\ell_H)_0}{d}. \quad (\text{C33})$$

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

Proof. This form of the variance is a special case of Eq. (C18), where all absorbing components vanish, and we have $T = Q$. In particular, this follows from Eq. (C15), observing that the above condition implies that for \mathcal{E}^{\dagger} Kadison-Schwarz inequality is strict, which combined with Corollary C.7.1 imply $T_z = 0 \forall z > 0$, leaving only T_0 . Finally, the corollary follows from the normalization condition $\text{Tr}[\rho] = 1$ on ρ , which ensures $(\ell_{\rho})_0 = 1/d$. \square

Appendix D: Proof of Theorem III.2

In this section we employ Proposition III.2 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 D.1. *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*

$$\mathbb{V}_{\rho,H}^L \geq \alpha^L (\ell_{\rho}, \ell_{\mathcal{K}(H)}), \quad (\text{D1})$$

where $\mathcal{K}(\cdot) = \sum_{\kappa \in K} \sum_j \text{Tr}[P_j \cdot] P_j \tilde{\delta}_{j,\kappa}$ 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

$$\begin{aligned} (\ell_{\rho}, T_l \ell_H) &= \sum_{\kappa,\lambda} \frac{(\ell_{\rho})_{\kappa} T_{\kappa,\lambda} (\ell_H)_{\lambda}}{d_{\kappa}} \\ &\geq \sum_{\kappa \in K} \frac{(\ell_{\rho})_{\kappa} T_{\kappa,\kappa} (\ell_H)_{\kappa}}{d_{\kappa}} \\ &\geq \alpha_l \sum_{\kappa \in K} \frac{(\ell_{\rho})_{\kappa} (\ell_H)_{\kappa}}{d_{\kappa}} = \alpha_l (\ell_{\rho}, \ell_{\mathcal{K}(H)}) \end{aligned} \quad (\text{D2})$$

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

Despite its simplicity, Theorem D.1 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 D.1.1. *Let $\mathcal{H} = \otimes_{m=1}^M \mathcal{H}_m$, $d_m \in \Theta(1)$. 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

$$\mathbb{V}_{\rho,H}^L \geq F(n) (\ell_{\rho}, \ell_{\mathcal{K}(H)}), \quad (\text{D3})$$

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

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

$$\begin{aligned} -\log(F(n)) &< -L \log\left(1 - C \frac{\log(n)}{L}\right) = \\ &= C \log(n) \left(1 + C \frac{\log(n)}{2L} + O\left(\frac{\log^2(n)}{L^2}\right)\right) \in O(\log(n)) \end{aligned} \quad (\text{D4})$$

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

Appendix E: Additional material on the applications

In this section we provide the proofs of all the results state in Applications section in the main text, concerning both small angle initializations and the effect of entanglement on noise induced concentration.

1. Small angle initializations

In order to prove the general lower bounds on small angle initializations provided in the main text, it is useful to start from Proposition IV.2, as it is the fundamental building block in this type of proofs. We recall it for convenience.

Proposition E.1. *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 to 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*

$$\mathbb{E}_\phi\{T_\phi\} \geq T, \quad (\text{E1})$$

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

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

$$\begin{aligned} \text{Tr}[\mathcal{E}(A)B]^2 &= \text{Tr}[A\mathcal{E}^\dagger(B)]^2 \\ &= \mathbb{E}_\phi^\dagger\left\{\text{Tr}[A\mathcal{E}_\phi^\dagger(B)]\right\}^2 \\ &\leq \mathbb{E}_\phi^\dagger\left\{\text{Tr}[A\mathcal{E}_\phi^\dagger(B)]^2\right\} \quad \forall A, B \end{aligned} \quad (\text{E2})$$

which follows from the observation that $f(\phi) = \text{Tr}[A\mathcal{E}_\phi^\dagger(B)] \in \mathbb{R}$, 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 $f(\phi) = \text{Tr}[AK^\dagger BK]$ 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. \square

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 $\mathbb{E}_\phi\{T_\phi\}$. In particular, it is easily verified that, if ϕ is sampled independently of the other parameters, then

$$\begin{aligned} \mathbb{V}_{\rho, H}^L &= \mathbb{E}_\phi \mathbb{E}_\theta \left\{ \text{Tr}[\Phi_{\theta, \phi}(\rho)H]^2 \right\} \\ &= \mathbb{E}_\phi \left\{ (\ell_\rho, T_\phi \ell_H) \right\} = (\ell_\rho, \mathbb{E}_\phi\{T_\phi\} \ell_H). \end{aligned} \quad (\text{E3})$$

Combining this observation with Corollary D.1.1, we can get the QResNet lower bound of Proposition IV.1.

Proposition E.2 (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 $L \in \Omega(\log^{1+\epsilon} n)$*

$$\mathbb{V}_{\rho, H}^L \geq F(n)(\ell_\rho, \ell_H) \quad (\text{E4})$$

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

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

$$\begin{aligned} T_{\kappa, \kappa} &= \frac{1}{d_\kappa} \sum_{i, j} \text{Tr} \left[\mathbb{E}_\phi \{ P_i \mathcal{E}_\phi^\dagger(P_j) \} \right]^2 \tilde{\delta}_{i, \kappa} \tilde{\delta}_{j, \kappa} \\ &\geq \frac{1}{d_\kappa} \sum_i \text{Tr} \left[\mathbb{E}_\phi \{ P_i \mathcal{E}_\phi^\dagger(P_i) \} \right]^2 \tilde{\delta}_{i, \kappa} \\ &= \frac{1}{d_\kappa} \sum_i \text{Tr} \left[\mathbb{E}_\phi \{ P_i e^{-i\phi G} P_i e^{i\phi G} \} \right]^2 \tilde{\delta}_{i, \kappa} \end{aligned} \quad (\text{E5})$$

Since $\sigma^2 \rightarrow 0$ as $n \rightarrow \infty$, to find the asymptotic behaviour 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. (E5), we get

$$\begin{aligned} T_{\kappa, \kappa} &\approx \\ &\frac{1}{d_\kappa} \sum_i \text{Tr} \left[\mathbb{E}_\phi \{ P_i (\mathbb{1} + i\phi G - \phi^2 G^2/2) P_i (\mathbb{1} + i\phi G - \phi^2 G^2/2) \} \right]^2 \tilde{\delta}_{i, \kappa} \\ &= \frac{1}{d_\kappa} \sum_i (1 - \mathbb{E}_\phi\{\phi^2\}) (\text{Tr}[G^2 P_i^2] - \text{Tr}[P_i G P_i G])^2 \tilde{\delta}_{i, \kappa} \\ &\geq 1 - 4\|G\|_2^2 \sigma^2 \quad \forall \kappa \in \{0, 1\}^M \end{aligned} \quad (\text{E6})$$

Exploiting Corollary D.1.1, 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 absence of concentration on a unitary QResNet, provided that \mathcal{P} is chosen as an initialization probability by Proposition E.1. \square

2. Non-unital noise and entanglement

As shown in Appendix C, explicit computation of the absorption matrix $A = R(\mathbb{1} - Q)^{-1}$ is a non-trivial task, which should be tackled on a case-by-case basis. Indeed, in the non-unital case, this term describes the complex phenomenon arising from the interaction of two competing effects, which drive the system towards 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, non-unital noise as a function of its strength. To do so, let us consider a non-unital map of the form

$$\mathcal{E}_c(\rho) = (1-p)\mathcal{E}(\rho) + p\tilde{\rho}, \quad (\text{E7})$$

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 to \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 $\Phi(\rho) = (1-p)\rho + p\tilde{\rho}$ and of $\mathcal{E} \circ \mathcal{U}_{\theta_i}$. Then we have the following Lemma.

Proposition E.3. Let \mathcal{E}_c be a quantum channel of type Eq. (E7), with $0 < p \leq 1$ and let T be the transition matrix of \mathcal{E}^\dagger . Then

$$\mathbb{V}_{\rho,H}^\infty = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1} \ell_H). \quad (\text{E8})$$

In particular, if T is a projection, then

$$\mathbb{V}_{\rho,H}^\infty = \left(\frac{p}{2-p} - p^2 \right) (\ell_{\tilde{\rho}}, T \ell_H) + p^2 (\ell_{\tilde{\rho}}, \ell_H). \quad (\text{E9})$$

Proof. The first result follows directly from Theorem C.8, in particular Corollary C.8.2, by computation of A_0 . In particular, we can explicitly compute R_0 element wise as

$$\begin{aligned} (R_0)_{0,\kappa} &= \frac{1}{d_\kappa} \sum_j \text{Tr} \left[\frac{\mathbb{1}}{\sqrt{d}} \mathcal{E}_c^\dagger(P_j) \right]^2 \tilde{\delta}_{j,\kappa} \\ &= \frac{1}{d_\kappa} \sum_j \text{Tr} \left[\mathcal{E}_c \left(\frac{\mathbb{1}}{\sqrt{d}} \right) P_j \right]^2 \tilde{\delta}_{j,\kappa} \\ &= \frac{1}{d_\kappa} \sum_j \left((1-p) \text{Tr} \left[\Phi \left(\frac{\mathbb{1}}{\sqrt{d}} \right) P_j \right] + p \sqrt{d} \text{Tr} [\tilde{\rho} P_j] \right)^2 \tilde{\delta}_{j,\kappa} \\ &= dp^2 \frac{1}{d_\kappa} \sum_j \text{Tr} [\tilde{\rho} P_j]^2 \tilde{\delta}_{j,\kappa} = dp^2 (\ell_{\tilde{\rho}})_\kappa \end{aligned} \quad (\text{E10})$$

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

$$(A_0)_{0,\lambda} = dp^2 (\ell_{\tilde{\rho}})_\kappa (1 - (1-p)^2 T)^{-1}_{\kappa,\lambda}. \quad (\text{E11})$$

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

$$(\ell_\rho, A_0 \ell_H) = p^2 (\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1} \ell_H). \quad (\text{E12})$$

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

$$\begin{aligned} (\mathbb{1} - (1-p)^2 T)^{-1} &= (\mathbb{1} - T) + (1 - (1-p)^2)^{-1} T \\ &= (\mathbb{1} - T) + \frac{1}{p(2-p)} T \\ &= \mathbb{1} + \left(\frac{1}{p(2-p)} - 1 \right) T \end{aligned} \quad (\text{E13})$$

which gives the result. \square

Appendix F: Choice of the system for the numerical example

In this section we provide an explicit construction of the noise channels \mathcal{E} , initial state ρ and observable H used to obtain the numerical results showed in Fig. 3.

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. 4 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 $\tilde{\rho} = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})(\langle 0|^{\otimes n} + \langle 1|^{\otimes n})/2$. The entangling part is chosen according to Fig. 4. While the number of layers L needed to reach convergence to $\mathbb{V}_{\rho,H}^\infty$ is logarithmic, as shown in Theorem III.1 and numerically assessed in the inset of Fig. 3, 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. 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 non-vanishing normalization factor $(\ell_{\tilde{\rho}}, \ell_H)$. As shown below, if we fix $h = 9/n$, we have $(\ell_{\tilde{\rho}}, \ell_H) = 1$, which makes the scaling of $\mathbb{V}_{\rho,H}^\infty$ especially easy to check. Indeed, 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

$$\begin{aligned} |0\rangle\langle 0| &= \frac{\mathbb{1} + Z}{\sqrt{2}}, \quad |1\rangle\langle 1| = \frac{\mathbb{1} - Z}{\sqrt{2}}, \\ |0\rangle\langle 1| &= \frac{X + iY}{\sqrt{2}}, \quad |1\rangle\langle 0| = \frac{X - iY}{\sqrt{2}}. \end{aligned} \quad (\text{F1})$$

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

Theorem F.1. 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

$$(A + \omega B)^{\otimes n} = \sum_{j=0}^n \omega^j \sum_{\sigma \in S} \sigma(A^{\otimes(n-j)} \otimes B^{\otimes j}) \quad (\text{F2})$$

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 F.1.1. 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

$$\frac{(A + \omega B)^{\otimes n} + (A - \omega B)^{\otimes n}}{2} = \sum_{j=0}^{\lfloor n/2 \rfloor} \omega^{2j} \sum_{\sigma \in S} \sigma(A^{\otimes(n-2j)} \otimes B^{\otimes 2j}) \quad (\text{F3})$$

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

Proof. This corollary follow directly from Theorem F.1, and noticing that all even-indexed terms in both $(A +$

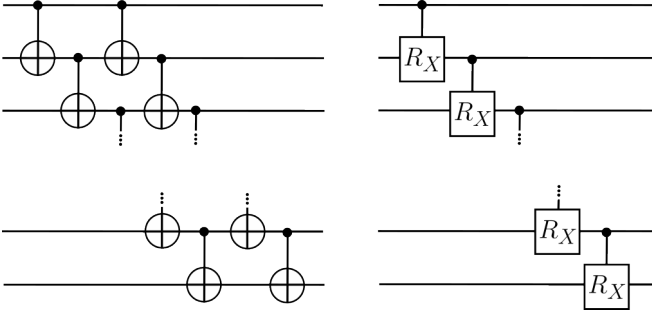


FIG. 4. Entangling unitaries used in the examples. On the left, the rapidly entangling configuration is composed of a double cascade of CNOT gates, while on the right, the slowly entangling one is composed of a single cascade of controlled RX gates, where $RX(\theta) = e^{i\theta X/2}$ and X is the Pauli X gate. In particular, we fixed $\theta = \pi/20$.

$\omega B)^{\otimes n}$ and $(A - \omega B)^{\otimes n}$ are equal, while odd-numbered terms have opposite sign and therefore cancel out. \square

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

$$\tilde{\rho} = \frac{1}{2^{n/2}} \sum_{j=0}^{\lfloor n/2 \rfloor} \sum_{\sigma \in S} \sigma(\mathbb{1}^{\otimes(n-2j)} \otimes Z^{\otimes 2j}) + (-1)^j \sigma(X^{\otimes(n-2j)} \otimes Y^{\otimes 2j}) \quad (\text{F4})$$

As it is clear from Eq. (F4), the fixed point of the channel has a non-vanishing component only on Pauli strings that are either non-trivial on all qubits, or non-trivial *only* in an *even* number of qubits. Then it follows that the simplest observable involving all qubits and with non-vanishing 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. (F4) in the calculations of $(\ell_{\tilde{\rho}}, \ell_H)$. Finally, since each term in the sum is orthogonal, it gives an independent contribution of $1/(d_m^2 - 1)^2 = 1/9$, consequently by choosing $h = 9/n$ we have that $(\ell_{\tilde{\rho}}, \ell_H) = 1$ is normalized.

Appendix G: Example of a circuit with non-convergent variance

As Theorem III.1 suggests, the generic circuit of Eq. (B1) need not have a well-defined value for the deep circuit limit of its variance, namely the limit $\lim_{L \rightarrow \infty} \mathbb{V}_{\rho, H}^L$ no need to exists. As discussed in Appendix C, 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. 5. The simple structure of this circuit allows to explicitly compute $\mathbb{V}_{\rho, H}^L$ as function of L . Assuming $\text{Tr}[H] = 0$, we have

$$\mathbb{V}_{\rho, H}^L = \begin{cases} \frac{(\|\rho\|_2^2 - 1/2)\|H\|_2^2}{3} & \text{if } L \text{ is even} \\ 0 & \text{if } L \text{ is odd} \end{cases} \quad (\text{G1})$$

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 :

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow T_0 = T_2 = (1), \quad T_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (\text{G2})$$

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

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

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L \mathbb{V}_{\rho, H}^l = \frac{(\|\rho\|_2^2 - 1/2)\|H\|_2^2}{6}, \quad (\text{G3})$$

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 .

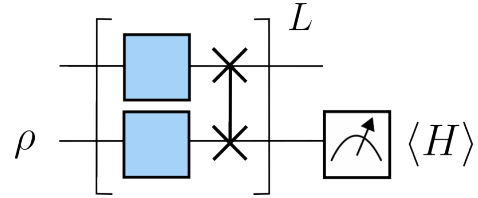


FIG. 5. Simple 2-qubit circuit, designed to have a non-convergent variance. In this construction, the channel \mathcal{E} is chosen to be unitary and in particular to be a SWAP gate. Moreover, both the initial state ρ and the observable H are chosen to be non-trivial only on the second qubit, and are therefore represented as single qubit operators.

- [1] R. P. Feynman, *International Journal of Theoretical Physics* **21**, 467 (1982).
- [2] S. Lloyd, *Science* **273**, 1073 (1996).
- [3] A. Di Meglio, K. Jansen, I. Tavernelli, C. Alexandrou,

- S. Arunachalam, C. W. Bauer, K. Borrás, S. Carrazza, A. Crippa, V. Croft, R. de Putter, A. Delgado, V. Dunjko, D. J. Egger, E. Fernández-Combarro, E. Fuchs, L. Funcke, D. González-Cuadra, M. Grossi, J. C. Hal-

- imeh, Z. Holmes, S. Kühn, D. Lacroix, R. Lewis, D. Lucchesi, M. L. Martinez, F. Meloni, A. Mezzacapo, S. Montanero, L. Nagano, V. R. Pascuzzi, V. Radescu, E. R. Ortega, A. Roggero, J. Schuhmacher, J. Seixas, P. Silvi, P. Spentzouris, F. Tacchino, K. Temme, K. Terashi, J. Tura, C. Tüysüz, S. Vallecorsa, U.-J. Wiese, S. Yoo, and J. Zhang, *PRX Quantum* **5**, 037001 (2024).
- [4] E. Fontana, N. Fitzpatrick, D. M. Ramo, R. Duncan, and I. Rungger, *Physical Review A* **104**, 10.1103/physreva.104.022403 (2021).
- [5] M. Vischi, G. Di Bartolomeo, M. Proietti, S. Koudia, F. Cerocchi, M. Dispenza, and A. Bassi, *Phys. Rev. Res.* **6**, 033337 (2024).
- [6] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babush, and H. Neven, *Nature Communications* **9**, 10.1038/s41467-018-07090-4 (2018).
- [7] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, *Quantum* **6**, 824 (2022).
- [8] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, *PRX Quantum* **2**, 040316 (2021).
- [9] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn, *PRX Quantum* **3**, 10.1103/prxquantum.3.020365 (2022).
- [10] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, *PRX Quantum* **3**, 010313 (2022).
- [11] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, *Physical Review X* **11**, 10.1103/physrevx.11.041011 (2021).
- [12] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, *Nature Communications* **12**, 10.1038/s41467-021-21728-w (2021).
- [13] A. V. Uvarov and J. D. Biamonte, *Journal of Physics A: Mathematical and Theoretical* **54**, 245301 (2021).
- [14] S. Thanasilp, S. Wang, N. A. Nghiem, P. Coles, and M. Cerezo, *Quantum Machine Intelligence* **5**, 10.1007/s42484-023-00103-6 (2023).
- [15] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. Ortiz Marrero, M. Larocca, and M. Cerezo, *Nature Communications* **15**, 10.1038/s41467-024-49909-3 (2024).
- [16] E. Fontana, D. Herman, S. Chakrabarti, N. Kumar, R. Yalovetzky, J. Heredge, S. H. Sureshbabu, and M. Pistoia, *Nature Communications* **15**, 10.1038/s41467-024-49910-w (2024).
- [17] N. L. Diaz, D. García-Martín, S. Kazi, M. Larocca, and M. Cerezo, *Showcasing a barren plateau theory beyond the dynamical lie algebra* (2023), arXiv:2310.11505 [quant-ph].
- [18] C.-Y. Park and N. Killoran, *Quantum* **8**, 1239 (2024).
- [19] C.-Y. Park, M. Kang, and J. Huh, *Hardware-efficient ansatz without barren plateaus in any depth* (2024), arXiv:2403.04844 [quant-ph].
- [20] K. Zhang, L. Liu, M.-H. Hsieh, and D. Tao, in *Advances in Neural Information Processing Systems*, edited by A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho (2022).
- [21] Y. Wang, B. Qi, C. Ferrie, and D. Dong, *Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization* (2023), arXiv:2302.06858 [quant-ph].
- [22] R. Puig, M. Drudis, S. Thanasilp, and Z. Holmes, *Variational quantum simulation: a case study for understanding warm starts* (2024), arXiv:2404.10044 [quant-ph].
- [23] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, *Nature Communications* **12**, 10.1038/s41467-021-27045-6 (2021).
- [24] M. Schumann, F. K. Wilhelm, and A. Ciani, *Quantum Science and Technology* **9**, 045019 (2024).
- [25] G. Crognalotti, G. D. Bartolomeo, M. Vischi, and L. L. Viteritti, *Equivariant variational quantum eigensolver to detect phase transitions through energy level crossings* (2024), arXiv:2403.07100 [quant-ph].
- [26] C. Tüysüz, S. Y. Chang, M. Demidik, K. Jansen, S. Vallecorsa, and M. Grossi, *PRX Quantum* **5**, 030314 (2024).
- [27] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, *A review of barren plateaus in variational quantum computing* (2024), arXiv:2405.00781 [quant-ph].
- [28] Y. Quek, D. Stilck França, S. Khatry, J. J. Meyer, and J. Eisert, *Nature Physics* **10.1038/s41567-024-02536-7** (2024).
- [29] R. Takagi, H. Tajima, and M. Gu, *Physical Review Letters* **10.1103/physrevlett.131.210602** (2023).
- [30] K. Tsubouchi, T. Sagawa, and N. Yoshioka, *Physical Review Letters* **10.1103/physrevlett.131.210601** (2023).
- [31] A. A. Mele, A. Angrisani, S. Ghosh, S. Khatry, J. Eisert, D. S. França, and Y. Quek, *Noise-induced shallow circuits and absence of barren plateaus* (2024), arXiv:2403.13927 [quant-ph].
- [32] B. Fefferman, S. Ghosh, M. Gullans, K. Kuroiwa, and K. Sharma, *PRX Quantum* **5**, 030317 (2024).
- [33] P. Singkanipa and D. A. Lidar, *Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and limit sets* (2024), arXiv:2402.08721 [quant-ph].
- [34] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, *Quantum Science and Technology* **7**, 045015 (2022).
- [35] E. Seneta, *Non-negative matrices and Markov chains* (Springer, New York, 2006).
- [36] H.-K. Zhang, S. Liu, and S.-X. Zhang, *Phys. Rev. Lett.* **132**, 150603 (2024).
- [37] C. D. Meyer, *Matrix analysis and applied linear algebra* (Society for Industrial and Applied Mathematics, USA, 2000).
- [38] K. He, X. Zhang, S. Ren, and J. Sun, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016) pp. 770–778.
- [39] G. Di Bartolomeo, M. Vischi, F. Cesa, R. Wixinger, M. Grossi, S. Donadi, and A. Bassi, *Physical Review Research* **5**, 10.1103/physrevresearch.5.043210 (2023).
- [40] S. L. Adler and A. Bassi, *Journal of Physics A: Mathematical and Theoretical* **40**, 15083 (2007).
- [41] J. Heredge, M. West, L. Hollenberg, and M. Seivior, *Non-unitary quantum machine learning* (2024), arXiv:2405.17388 [quant-ph].
- [42] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. Akash-Narayanan, A. Asadi, J. M. Arrazola, U. Azad, S. Banning, C. Blank, T. R. Bromley, B. A. Cordier, J. Ceroni, A. Delgado, O. D. Matteo, A. Dusko, T. Garg, D. Guala, A. Hayes, R. Hill, A. Ijaz, T. Isaacsson, D. Ittah, S. Jhangiri, P. Jain, E. Jiang, A. Khandelwal, K. Kottmann, R. A. Lang, C. Lee, T. Loke, A. Lowe, K. McKiernan, J. J. Meyer, J. A. Montañez-Barrera, R. Moyard, Z. Niu, L. J. O’Riordan, S. Oud, A. Panigrahi, C.-Y. Park, D. Polatajko, N. Quesada, C. Roberts, N. Sá, I. Schoch, B. Shi, S. Shu, S. Sim, A. Singh, I. Strandberg, J. Soni,

- A. Száva, S. Thabet, R. A. Vargas-Hernández, T. Vincent, N. Vitucci, M. Weber, D. Wierichs, R. Wiersema, M. Willmann, V. Wong, S. Zhang, and N. Killoran, [Penylane: Automatic differentiation of hybrid quantum-classical computations](#) (2022), [arXiv:1811.04968 \[quant-ph\]](#).
- [43] J. Preskill, [Quantum](#) **2**, 79 (2018).
- [44] A. Katabarwa, K. Gratsea, A. Caesura, and P. D. Johnson, [PRX Quantum](#) **5**, [10.1103/prxquantum.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](#) (2024), [arXiv:2312.09121 \[quant-ph\]](#).
- [46] M. Raginsky, [Physical Review A](#) **65**, [10.1103/physreva.65.032306](#) (2002).
- [47] A. S. Holevo and R. F. Werner, [Phys. Rev. A](#) **63**, 032312 (2001).
- [48] D. Pérez-García, M. M. Wolf, D. Petz, and M. B. Ruskai, [Journal of Mathematical Physics](#) **47**, [10.1063/1.2218675](#) (2006).
- [49] C. King and M. B. Ruskai, [Minimal entropy of states emerging from noisy quantum channels](#) (2000), [arXiv:quant-ph/9911079 \[quant-ph\]](#).
- [50] M. Beth Ruskai, S. Szarek, and E. Werner, [Linear Algebra and its Applications](#) **347**, 159 (2002).
- [51] R. V. Kadison, [Annals of Mathematics](#) **56**, 494 (1952).