Estimates of loss function concentration in noisy parametrized quantum circuits

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

[1]*Department of Physics, University of Trieste, Strada Costiera 11, 34151 Trieste, Italy*
[2]*Istituto Nazionale di Fisica Nucleare, Trieste Section, Via Valerio 2, 34127 Trieste, Italy*
[3]*European Organization for Nuclear Research (CERN), 1211 Geneva, Switzerland*
(Dated: May 5, 2025)

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

I. INTRODUCTION

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

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

are small-angle initializations [19–23], which leverage the restriction of the domain Θ of θ to limit the expressive power of the circuit, and consequently avoid concentration.

In addition to that, the presence of noise is also deemed detrimental, as it often gives rise to noise-induced barren plateau (NIBP) [24, 25] and symmetry breaking [26, 27]. The former exacerbates the BP effect, producing a deterministic concentration. In this case, an efficient loss-minimizing direction is exponentially hard to estimate, *regardless* of the parameter's initialization choice. These phenomena are mostly linked to decoherence. Ref. [28] offers a review of the subject.

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

In this work, we approach the problem through a novel formulation grounded in non-negative matrix theory, which enables a unified and general description of these dynamics. This framework not only addresses the previously uncharted aspects of the problem but also seamlessly encompasses both unital and non-unital noise settings. We quantify the deep circuit variance $\mathbb{V}_{\rho,H}^{\infty}$ for layered circuits composed of local 2-designs interleaved with

arXiv:2410.01893v2 [quant-ph] 2 May 2025

general noise maps, providing a comprehensive picture. More specifically, we are able to recover known results for unitary circuits and strictly contractive noise maps as limiting cases, while unveiling the emergence of a more complex phenomenon, namely *absorption*, in the intermediate case. Crucially, this can occur only with noise maps that are not strictly contractive—that is, maps that are insufficiently strong to fully erase all quantum information from the system. It is worth emphasizing that this class encompasses both unital and non-unital noise channels, illustrating that unitality of the noise alone is not necessarily detrimental for PQCs. The proposed framework also facilitates a systematic analysis of the noise resilience properties of PQCs, providing a foundation for principled BP mitigation strategies. Notably, we demonstrate a direct correlation between a circuit's resilience to noise and its capacity for enhanced expressivity under small-angle initialization-like schemes.

II. MOTIVATING EXAMPLES

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

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

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

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

where each layer is composed of a unitary part $\mathcal{U}_{\theta_l}$ and an arbitrary quantum channel $\mathcal{E}_l$. If we denote by $\mathcal{B}$ the space of bounded operators acting on $\mathcal{H}$, both such components can be regarded as linear, completely positive functions mapping $\mathcal{B}$ to itself. Furthermore, the intermediate channels $\mathcal{E}_l$ will be deemed strictly-contractive if the restriction to the Hermiatian, traceless hyperplane $\mathbb{H}_0 \subset \mathcal{B}$ of its adjoint action $\mathcal{E}_l^\dagger$ has an induced Schatten 1-norm $\|\mathcal{E}_l^\dagger|_{\mathbb{H}_0}\|_1$ strictly less than unity, $\|\mathcal{E}_l^\dagger|_{\mathbb{H}_0}\|_1 < 1$, and contractive otherwise, i.e. $\|\mathcal{E}_l^\dagger|_{\mathbb{H}_0}\|_1 = 1$. According to this definition, contractive channels preserve the 1-norm of certain operators in $\mathcal{B}$, while possibly being strictly contractive for others. Moreover, unitary channels, which preserve all norms across their entire domain,

are a special case of contractive channels that are never strictly contractive. We refer to Appendix D for more details.

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

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

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

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

III. GENERAL QUANTUM CHANNELS

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

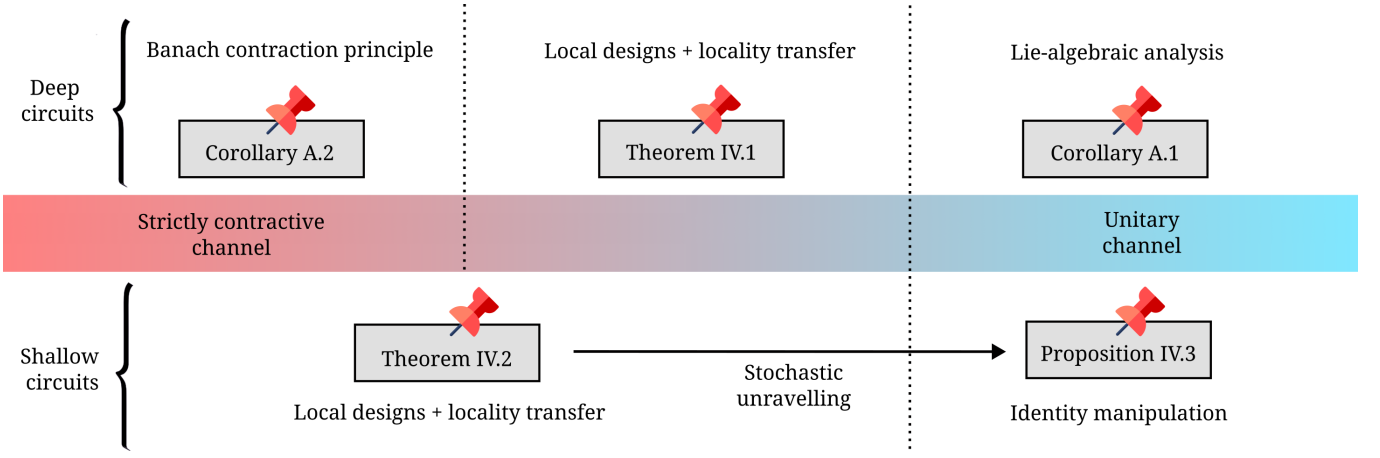$$\mathcal{B} = \bigoplus_{\kappa \in \{0,1\}^M} \mathcal{B}_\kappa \qquad (3)$$

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 [24, 25] and unitary [16–18] channels, Theorem IV.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 Appendix A. The lower part focuses on "shallow" circuits, where a lower bound on $\mathbb{V}_{\rho,H}^L$ is established using Theorem IV.2. This extends previous works on brickwork circuits [13, 32], enabling initialization strategies such as small angle initializations [21, 22] to be represented as stochastic unravellings of noise maps. In Section IV D we expand upon their applicability, showing how both unitary and non-unitary QResNets can be derived thanks to Proposition IV.3.

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

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

for some Hermitian, orthonormal basis $\{B_j\}_{j=1}^{d_\kappa}$ of $\mathcal{B}_\kappa$. Clearly, from the definition, a κ-local operator A has locality vector $(\ell_A)_\lambda = \delta_{\kappa,\lambda} \|A\|_2^2$, where $\|\cdot\|_2$ is the Hilbert-Schmidt norm. We remark that similar quantities are not new in the context of PQCs, and in fact the vector ℓ_A is analogous to the purity measures defined in [16, 18]. Based on this, a derived notion of *locality preservation* can be introduced for linear maps acting on $\mathcal{B}$. In particular, given a map $\Lambda : \mathcal{B} \to \mathcal{B}$ and two subspaces $\mathcal{B}_\kappa$ and $\mathcal{B}_\lambda$, we can measure the degree to which Λ is able to put them in communication. The more subspaces are connected, the less locality preserving the map Λ will be. This idea is captured formally in the following definition of a locality transfer matrix (LTM).

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

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

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

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

With this description in mind, we assume that $\mathcal{U}_\theta$ of Eq. (2) is locality preserving, and hence describes an ideal operation limited to the local subsystems $\mathcal{H}_m$, while each channel $\mathcal{E}_l$ encodes both the operations which entangle the subsystems as well as any interaction between the system and the environment. This formally captures the idea of a variational quantum algorithm running on a real, noisy, device, where quantum computation can be 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, $\mathcal{U}_{\theta_l}$ is deep enough to form an approximate 2-design over the local unitary groups $U(d_m)$. This assumption is justified by the fact that such local operations are relatively inexpensive, and that the necessary depth can be very small even for large systems, since in general it depends on d_m rather than d.

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

Hermitian adjoint of one another, i.e.

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

We refer the reader to Appendix B for a more detailed discussion. Having introduced the main tools, we now characterize the scaling of the variance $\mathbb{V}_{\rho,H}^L$ of $\mathcal{L}_{\rho,H}$ for quantum channels described by Eq. (2).

IV. 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 work.

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 IV.1 (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*

$$\mathbb{E}_\theta \left\{ \mathrm{Tr} \left[\Phi_\theta(\rho) H \right]^2 \right\} = \left(\ell_\rho, \prod_{l=1}^L T_l \, \ell_H \right) \tag{7}$$

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 the fact that, under the assumptions of Section II, each layer $\mathcal{U}_{\theta_l}$ forms a global unitary 1-design, Proposition IV.1 allows a direct calculation of $\mathbb{V}_{\rho,H}^L$ as a function of the LTMs of the intermediate channels as

$$\mathbb{V}_{\rho,H}^L = \left(\ell_\rho, \prod_{l=1}^L T_l \, \ell_H \right) - \frac{\mathrm{Tr} \left[H \right]^2}{d^2}. \tag{8}$$

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

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 H 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 IV.1 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, $\mathrm{Tr}\left[H\right] = 0$. In this case, Eq. (8) becomes $\mathbb{V}_{\rho,H}^L = \left(\ell_\rho, T^L \ell_H \right)$. Note that, by construction, T is a non-negative matrix, namely $T_{\kappa,\lambda} \geq 0 \ \forall \kappa, \lambda$, and thus it can always be expressed in canonical, block-upper triangular form, where each of the diagonal blocks is irreducible [35]. Throughout this work, irreducible components of T will be regarded as *essential* if they cannot lead outside the block, and will be denoted by T_z. Otherwise, an irreducible block will be deemed *inessential*, and their collection will be denoted by Q. A useful pictorial representation of these possibilities is shown in Fig. 2, together with the matrix canonical form. A foundational reference for this definition can be found in Ref. [35], while a more detailed discussion is proposed in Appendix D.

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 $(v_z)_\kappa = 1 \ \forall \kappa$. In contrast, its contribution vanishes under repeated applications of the channel when the condition is not satisfied. This suggests that, as $L \to \infty$, the general form for the variance reads

$$\mathbb{V}_{\rho,H}^\infty = \sum_{z \, | \, \rho(T_z)=1} (\ell_\rho, w_z)(\ell_H)_z + (\ell_\rho, w_z)(A\ell_H)_z, \tag{9}$$

where w_z denotes the right eigenvector of the leading eigenvalue of T_z and A is a matrix of the same shape as
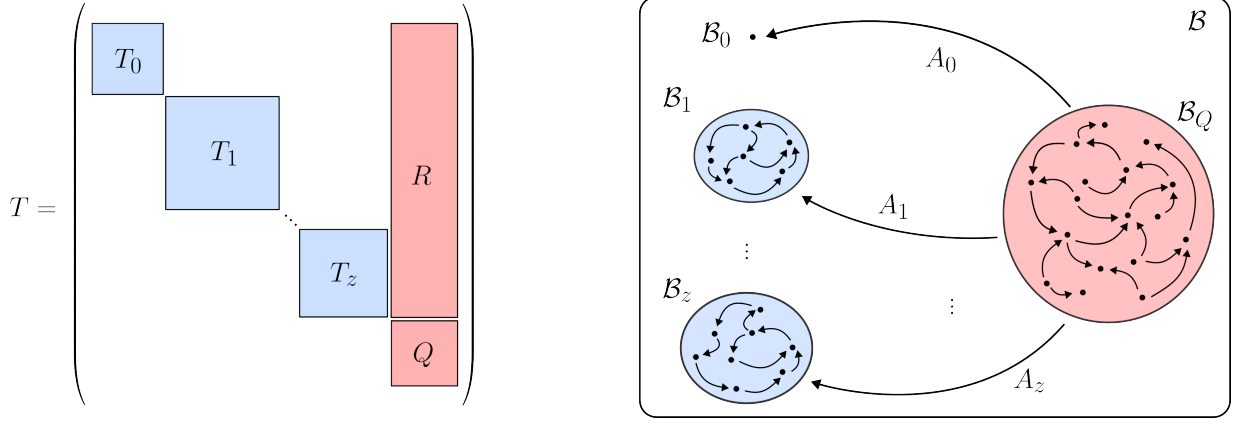
FIG. 2. Graphical representations of the stochastic process describing $\mathbb{V}_{\rho,H}^L$. On the left, we show the structure of the general locality transfer matrix (LTM), highlighting the decomposition into irreducible components [35]. Light blue blocks represent irreducible, essential components 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 $\mathcal{B}_\kappa$. Here, each dot represents a single subspace, while the arrows represent the adjoint action of the channel $\mathcal{E}$. Essential and inessential components share the same colour code of T.

R in Fig. 2. Such reasoning is formally encapsulated in the following Theorem.

Theorem IV.1 (Deep circuit variance). *Let $\rho, H \in \mathcal{B}$ and let Φ_θ a be layered quantum channel as in described in Section II. Then the Cesàro average of $\mathbb{V}_{\rho,H}^L$ converges to Eq. (9), and we have*

$$\left| \frac{1}{L} \sum_{l=0}^{L} \mathbb{V}_{\rho,H}^l - \mathbb{V}_{\rho,H}^\infty \right| \in O\left(e^{-\beta L} \|H\|_2^2 \right), \qquad (10)$$

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

$$\left| \mathbb{V}_{\rho,H}^L - \mathbb{V}_{\rho,H}^\infty \right| \in O\left(e^{-\beta L} \|H\|_2^2 \right), \qquad (11)$$

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

As a special case, if the intermediate channel can be reduced unitarily to a tensor product of single qubit channels, then also w_z of Eq. (9) can be explicitly computed, yielding the following Corollary.

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

$$\mathbb{V}_{\rho,H}^\infty = \sum_{z \mid \rho(T_z)=1} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z (A\ell_H)_z}{d_z}. \qquad (12)$$

The proof of Theorem IV.1 and Corollary IV.1 can be found in Appendix D.

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 to. This is put into an effective example and further discussed Appendix H. For simplicity, in the following, we will assume *aperiodicity* for each irreducible block.

We start the analysis of Theorem IV.1 by unpacking Eq. (12). In this equation, $\mathbb{V}_{\rho,H}^\infty$ is shown to depend on four important quantities, namely the invariant subspaces $\mathcal{B}_z$ of $\Phi_\theta^\dagger$, 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 present. Failure to achieve this will cause the emergence of noise-induced concentration in the sense of Ref. [24], as shown in Appendix A. This extends the notion of alignment already introduced in the literature [28] for ρ, H and unitary circuits. Moreover, we observe that such subspaces act as *attractors* for the variance, as each summand not only depends on the local components of ρ and H, but also on the components of H belonging to $\mathcal{E}(\mathcal{B}_z) \cap \mathcal{B}_Q$. In this sense, A can be interpreted as an absorption matrix, which quantifies the extent of the contribution of such terms. We remark that this contribution is always non-negative, and it is intimately related to the contractivity properties of $\mathcal{E}^\dagger$. 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.

As shown in Appendix A, Eq. (12) can be used to recover deep circuit limit variance in the unitary and

strictly contractive setting, consistent with existing literature [16, 17, 32, 34]. In particular, regarding strictly contractive, non-unital noise, we show how previous results can be expressed as the contribution to the absorption term of $\mathcal{B}_0$, which always forms a norm-preserving subspace due to trace preservation of Φ_θ. However, it is crucial to remark that this specific contribution is not due to the retention of any computational power to the PQC, since the dependence on the initial state is completely lost, but rather to the competing effects between the drive of $\mathcal{U}_\theta$ and $\mathcal{E}$ towards the respective, different fixed points. An example of this phenomenon is provided in Appendix G, where we study how extensive entanglement can further exacerbate the issue. This is in stark contrast with the general case discussed above, i.e. the absorption terms of $\mathcal{B}_z$, $z > 0$, as, although more difficult to realize, they keep a non-trivial dependence on the initial state, and as such it can be said to genuinely avoid concentration if d_z scales appropriately.

The results showed so far were based on the analysis of the dominant eigenvectors of T. When L is not deep enough to reach convergence to its asymptotic limit, we need to characterize better the 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 IV.1 are sufficiently loose, namely $\beta L \in O(\log n)$. This implies either that the circuit is *shallow*, i.e. there are not enough layers to reach the asymptotic value $\mathbb{V}_{\rho,H}^\infty$, 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 L. As this speed is related to the amount of correlations with respect to the partition introduced by the channel, we study this scenario in the limit of $\mathcal{E}^\dagger$ being close to locality preserving. This idea is formally captured in the following Theorem, which extends its application to generally non-homogeneous channels.

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

$$\mathbb{V}_{\rho,H}^L \geq \alpha^L(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{13}$$

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

Depending on the scaling of α and the dimensions d_m of the subsystems, Eq. (13) can provide a meaningful

lower bound. For instance, focusing on the case of subsystems with constant dimension, we have the following Corollary.

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

(a) $\alpha > 0$, $\alpha \in \Omega(1)$ and $L \in O(\log n)$,

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

is satisfied, then

$$\mathbb{V}_{\rho,H}^L \geq F(n)(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{14}$$

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

The conditions of Corollary IV.2 reflect the aforementioned scenarios; in particular, condition (a) ensures the absence of concentration for *shallow* circuits, both unitary and noisy. Specifically, this holds true whenever $0 < \alpha \in \Omega(1)$, indicating that the intermediate channel does not become increasingly rapidly entangling, as the problem size grows. As a notable example, this condition is satisfied for brickwork circuits, equipped with local noise and a local observable [32]. Similarly, condition (b) reflects the absence of concentration for *effectively shallow* quantum channels. A significant example is that of 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 [21, 22], which similarly hinge on identity manipulation. In fact, while the primary concern of Theorem IV.2 is noise, it could still be regarded as a theoretical foundation of such initialization strategies. In the following Section, we show a deep relation between these types of smart initializations for noiseless circuits and the properties of noisy layers.

D. Connection with small angle initializations

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

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

Proposition IV.2 (QResNet). *Let $\mathcal{E}_\phi(\cdot) = e^{i\phi G}\cdot e^{-i\phi G}$ be a unitary entangling gate, 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*

$$\mathbb{V}_{\rho,H}^L \geq F(n)(\ell_\rho, \ell_H), \tag{15}$$

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

This result represents an application of Corollary IV.2 (*b*) in the case of unitary, parameterized intermediate channels. Note that Theorem IV.2, which is the backbone of Proposition IV.2, is not limited to unitary circuits, but is applicable to generic quantum channels. Indeed, if we take $\mathcal{E}$ to be a noise model, Corollary IV.2 (*b*) may be analogously interpreted as a condition on the noise rates to avoid concentration. This showcases a connection between QResNets that can avoid BP and noise models not strong enough to cause NIBP. Formally, this is captured by the following Proposition.

Proposition IV.3 (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*

$$\mathbb{E}_\phi\{T_\phi\} \geq T, \tag{16}$$

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

This result can be interpreted as follows: if a noise map satisfies Corollary IV.2, then there exists a QResNet 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.3, the variance of the model obtained from the ensemble by promoting ϕ to variational parameters is lower-bounded by the variance of the channel, and hence it provides a BP free QResNet for weak enough noise. Interestingly, Proposition IV.3 can be equally applied if the ensemble is not unitary, which can happen if the noise channel $\mathcal{E}$ is non-unital. This effectively extends 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].

Finally, it is important to note that stochastic unravellings are not unique, allowing multiple QResNets (unitary or non-unitary) to be derived from the same channel. Additionally, different choices among such models may lead to distinct variances [42], all bounded from below by the variance of $\mathcal{E}$. Indeed, any stochastic unravelling is connected to the physical channel solely through their first moment, while higher-order moments are relevant only in the context of the associated QResNet, as they do not represent otherwise measurable quantities.

V. DISCUSSION

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

In particular, within this framework, the variance is formulated in terms of the channel's locality transfer matrix (LTM), which in turn enables the precise computation of $\mathbb{V}_{\rho,H}^\infty$ in the deep-circuit limit. Specifically, 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 through the non-reversible action of the quantum noise channel. Such result is supported by numerical simulations, provided in Appendix G and Appendix F.

This indicates that a mixed configuration of ideal and

noisy qubits, as well as a partial realization of error correction, could potentially outperform purely ideal or purely noisy systems in terms of variance. This is especially important in the early stages of fault-tolerant quantum computing, where a large quantity of noisy qubits can be utilized, but the availability of ideal qubits is still constrained [43, 44]. This may also be beneficial in characterizing variational quantum algorithms executed on hardware, where different qubits may experience varying error rates [39]. In this context, $\mathbb{V}_{\rho,H}^{\infty}$ 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.

Furthermore, this approach clarifies the role of unitality in noise-induced barren plateaus, shifting the focus on the more relevant aspect of *contractivity* and *alignment* between invariant subspaces of the unitary layers $\mathcal{U}_\theta$ and strictly contractive subspaces of the noise map $\mathcal{E}$.

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 [19–22] that can effectively mitigate the occurrence of barren plateaus. Moreover, we demonstrate that an analogous procedure can be applied to noise maps, establishing a formal connection between weak noise and QResNets. This enables us to derive BP-free QResNets as stochastic 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].

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

Appendix A: Limiting cases of the deep circuit formula

Starting from the analysis of Section IV, we show how we can use Eq. (12) to 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. (12) vanishes for unitary dynamics, which is reversible by definition. Formally, we have the following Corollary of Theorem IV.1.

Corollary A.1 (Deep, unitary circuits). *Let $\rho, H \in \mathcal{B}$ and let Φ_θ a be layered quantum channel as described in Section II, where $\mathcal{E}(\cdot) = W \cdot W^\dagger$, $W \in U(d)$ is an arbitrary unitary transformation. Then we have*

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

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

This Corollary contains many interesting properties of the variance in the deep circuit limit. First, it captures the necessity of *alignment* between ρ, H and Φ_θ in order to achieve a substantial variance, i.e. 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 $\mathcal{U}_\theta$ and $\mathcal{E}$, whose invariant subspaces need to align in order to keep the dimension d_z of $\mathcal{B}_z$ from being exponentially large. Indeed, while $\mathcal{B}_0$ is always an invariant subspace satisfying Corollary A.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

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

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

As a complementary remark, we point out that, conversely to the above, the first term in Eq. (12) 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 of Theorem IV.1.

Corollary A.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*

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

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

Note that, even if Eq. (A3) 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 A.2, and therefore exhibit some form of concentration. However, Corollary A.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 $\mathcal{U}_\theta$. According to the method introduced here, this can clearly be interpreted as a consequence of the interaction between noise and unitary layers, which effectively "spread" the contractive effect of $\mathcal{E}$ to the whole irreducible component. Unital channels will suffer most severely from NIBP, since in that case the absorption term in Eq. (A3) vanishes. Contrarily, as pointed out in the literature [32, 34], non-unital channels may avoid the exponential concentration. Here, we have shown how these results obtained in the literature can be seen as the contribution to the absorption term of $\mathcal{B}_0$ which cannot be strictly contractive due to trace preservation of Φ_θ.

Appendix B: Locality and locality transfer matrix properties

In order to show the main properties of locality vectors and 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 $P_0^{(m)} = \mathbb{1}/\sqrt{d_m}$, $P_{j_m}^{(m)} = P_{j_m}^{(m)\dagger}\,\forall j_m$,$\mathrm{Tr}\left[P_{j_m}^{(m)}P_{k_m}^{(m)}\right] = \delta_{j_m k_m}$. Starting from these, we can build an orthonormal basis $\{P_j\}_j$ for the whole space by means of tensor products. Each basis element will be labelled by the multi-index $j = (j_1,...,j_M)$, where the entries j_m refer to an element of the local bases, and hence $j_m \in \{0,...,d_m^2-1\}$. 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

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

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

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

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

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

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

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

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

and

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

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

Lemma B.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 $\mathcal{E} : \mathcal{B} \to \mathcal{B}$. the corresponding locality transfer matrix T is uniquely defined.*

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

$$\sum_j \mathrm{Tr}\,[P_j A]^2\,\tilde{\delta}_{j,\kappa} = \sum_j \mathrm{Tr}\left[\sum_i \mathrm{Tr}\,[B_i P_j]\,B_i A\right]^2 \tilde{\delta}_{j,\kappa}$$

$$= \sum_j \sum_{i,i'} \mathrm{Tr}\,[B_i P_j]\,\mathrm{Tr}\,[B_{i'} P_j]\,\mathrm{Tr}\,[B_i A]\,\mathrm{Tr}\,[B_{i'} A]\,\tilde{\delta}_{j,\kappa}$$

$$= \sum_{i,i'} \left(\sum_j \mathrm{Tr}\,[B_i P_j]\,\mathrm{Tr}\,[B_{i'} P_j]\,\tilde{\delta}_{j,\kappa}\right) \mathrm{Tr}\,[B_i A]\,\mathrm{Tr}\,[B_{i'} A]$$

$$= \sum_i \mathrm{Tr}\,[B_i A]^2\,\tilde{\delta}_{i,\kappa}$$

$$\tag{B5}$$

where the last equality is due to

$$\sum_j \text{Tr}\left[B_i P_j\right]\text{Tr}\left[B_{i'}P_j\right]\tilde{\delta}_{j,\kappa} =$$

$$= \prod_{m=1}^{M}\left(\sum_{j_m=1}^{d_m}\text{Tr}\left[B_{i'_m}P_{j_m}\right]\text{Tr}\left[B_{i_m}P_{j_m}\right]\tilde{\delta}_{j_m,\kappa_m}\right) \quad \text{(B6)}$$

$$= \prod_{m=1}^{M}\left(\text{Tr}\left[B_{i'_m}B_{i_m}\right]\tilde{\delta}_{i_m,\kappa_m}\right) = \delta_{i,i'}\tilde{\delta}_{i,\kappa}$$

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. $\qquad\square$

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

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

$$= \sum_{i,j}\text{Tr}\left[\Lambda^\dagger(P_i)P_j\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda} = d_\kappa T_{\lambda,\kappa}^\dagger, \quad \text{(B7)}$$

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

$$(a,b) = a^t D^{-1}b = \sum_\kappa \frac{a_\kappa b_\kappa}{d_\kappa}. \quad \text{(B8)}$$

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

Appendix C: Proof of Proposition IV.1

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

$$\Phi_\theta = \mathcal{U}_{\theta_{L+1}}\circ\mathcal{E}_L\circ\mathcal{U}_{\theta_{L-1}}\cdots\circ\mathcal{E}_1\circ\mathcal{U}_{\theta_1}. \quad \text{(C1)}$$

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

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

$$\int_\Theta d\theta U_\theta^{(m)\otimes t}\otimes U_\theta^{(m)*\otimes t} = \int_{V\in U(d_m)}d\mu(V)V^{(m)\otimes t}\otimes V^{(m)*\otimes t} \quad \text{(C2)}$$

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 C.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*

$$\mathbb{E}_\theta\left\{\text{Tr}\left[AU_\theta^\dagger BU_\theta\right]\right\} = \frac{\text{Tr}\left[A\right]\text{Tr}\left[B\right]}{d} \quad \text{(C3)}$$

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

$$\int_\Theta \prod_{m=1}^{M}d\theta\,\text{Tr}\left[AU_\theta^\dagger BU_\theta\right] =$$

$$= \sum_{i,j}a_i b_j\int_\Theta\prod_{m=1}^{M}d\theta\,\text{Tr}\left[P_i U_\theta^\dagger P_j U_\theta\right]$$

$$= \sum_{i,j}a_i b_j\prod_{m=1}^{M}\int_{U\in U(d_m)}d\mu(U)\,\text{Tr}_m\left[P_{i_m}^{(m)}U^\dagger P_{j_m}^{(m)}U\right]$$

$$= \sum_{i,j}a_i b_j\prod_{m=1}^{M}\frac{1}{d_m}\text{Tr}_m\left[P_{i_m}^{(m)}\right]\text{Tr}_m\left[P_{j_m}^{(m)}\right]$$

$$= \sum_{i,j}a_i b_j\prod_{m=1}^{M}\delta_{0,i_m}\delta_{0,j_m} = a_0 b_0$$

$$\text{(C4)}$$

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

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

Lemma C.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*

$$\mathbb{E}_\theta\left\{\text{Tr}\left[P_i U_\theta^\dagger P_j U_\theta\right]\text{Tr}\left[P_k U_\theta^\dagger P_l U_\theta\right]\right\} =$$

$$= \delta_{ij}\delta_{kl}\prod_{m=1}^{M}\left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0j_m})\frac{1}{d_m^2-1}\right) \quad \text{(C5)}$$

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

$$\int_{U\in U(d)}d\mu(U)\,\text{Tr}\left[AU^\dagger BU\right]\text{Tr}\left[CU^\dagger DU\right]$$

$$= \frac{1}{d^2-1}\left(\text{Tr}\left[A\right]\text{Tr}\left[B\right]\text{Tr}\left[C\right]\text{Tr}\left[D\right] + \text{Tr}\left[AC\right]\text{Tr}\left[BD\right]\right)$$

$$+ \frac{1}{d(d^2-1)}\left(\text{Tr}\left[AC\right]\text{Tr}\left[C\right]\text{Tr}\left[D\right] + \text{Tr}\left[A\right]\text{Tr}\left[C\right]\text{Tr}\left[BD\right]\right)$$

$$\text{(C6)}$$

Based on Eq. (C6), the result follows from direct integration:

$$\prod_{m=1}^{M} \int_{U \in U(d_m)} d\mu(U) \operatorname{Tr}\left[P_{i_m}^{(m)} U^\dagger P_{j_m}^{(m)} U\right] \times$$
$$\times \operatorname{Tr}\left[P_{k_m}^{(m)} U^\dagger P_{l_m}^{(m)} U\right]$$
$$= \prod_{m=1}^{M} \frac{1}{d_m^2 - 1}\Big(\operatorname{Tr}\left[P_{i_m}^{(m)}\right]\operatorname{Tr}\left[P_{j_m}^{(m)}\right]\operatorname{Tr}\left[P_{k_m}^{(m)}\right]\operatorname{Tr}\left[P_{l_m}^{(m)}\right]$$
$$+ \operatorname{Tr}\left[P_{i_m}^{(m)} P_{k_m}^{(m)}\right]\operatorname{Tr}\left[P_{j_m}^{(m)} P_{l_m}^{(m)}\right]\Big)$$
$$- \frac{1}{d_m(d_m^2 - 1)}\Big(\operatorname{Tr}\left[P_{i_m}^{(m)} P_{k_m}^{(m)}\right]\operatorname{Tr}\left[P_{j_m}^{(m)}\right]\operatorname{Tr}\left[P_{l_m}^{(m)}\right]$$
$$+ \operatorname{Tr}\left[P_{i_m}^{(m)}\right]\operatorname{Tr}\left[P_{k_m}^{(m)}\right]\operatorname{Tr}\left[P_{j_m}^{(m)} P_{l_m}^{(m)}\right]\Big)$$
$$= \prod_{m=1}^{M} \frac{1}{d_m^2 - 1}(d_m^2\, \delta_{0 i_m} \delta_{0 j_m} \delta_{0 k_m} \delta_{0 l_m}$$
$$+ \delta_{i_m k_m}\delta_{j_m l_m} - \delta_{0 i_m}\delta_{0 k_m}\delta_{j_m l_m} - \delta_{i_m k_m}\delta_{0 j_m}\delta_{0 l_m})$$
$$= \prod_{m=1}^{M} \delta_{0 i_m}\delta_{0 j_m}\delta_{0 k_m}\delta_{0 l_m}$$
$$+ \frac{1}{d_m^2 - 1}(\delta_{i_m k_m} - \delta_{0 i_m}\delta_{0 k_m})(\delta_{j_m l_m} - \delta_{0 j_m}\delta_{0 l_m})$$
$$= \delta_{ik}\delta_{jl} \prod_{m=1}^{M}\left(\delta_{0 i_m}\delta_{0 j_m} + (1 - \delta_{0 i_m})(1 - \delta_{0 j_m})\frac{1}{d_m^2 - 1}\right) \tag{C7}$$

$\square$

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

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

$$\mathbb{E}_\theta\left\{\operatorname{Tr}\left[AU_\theta^\dagger B U_\theta\right]^2\right\} = (\ell_A, \ell_B), \tag{C8}$$

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

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

$$\mathbb{E}_\theta\left\{\operatorname{Tr}\left[AU_\theta^\dagger B U_\theta\right]^2\right\} =$$
$$\sum_{i,j} a_i^2 b_j^2 \prod_{m=1}^{M}\left(\delta_{0 i_m}\delta_{0 j_m} + (1 - \delta_{0 i_m})(1 - \delta_{0 j_m})\frac{1}{d_m^2 - 1}\right) \tag{C9}$$

In the following, it will be convenient to recast the product on the right-hand side of Eq. (C9) into the equivalent formulation

$$\prod_{m=1}^{M}\left(\delta_{0 i_m}\delta_{0 j_m} + (1 - \delta_{0 i_m})(1 - \delta_{0 i_m})\frac{1}{d_m^2 - 1}\right) =$$
$$= \sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa} \prod_{m=1}^{M}(\delta_{0 i_m}\delta_{0 j_m})^{1 - \kappa_m}(1 - \delta_{0 i_m})^{\kappa_m}(1 - \delta_{0 j_m})^{\kappa_m} \tag{C10}$$

where the binary vectors $\kappa \in \{0,1\}^M$ identify all possible sets S_κ introduced in Appendix B and $d_\kappa = \prod_{m=1}^{M}(d_m^2 - 1)^{\kappa_m}$. Putting it back into Eq. (C9) we get

$$\sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa} \sum_{i,j} a_i^2 b_j^2 \prod_{m=1}^{M}(\delta_{0 i_m}\delta_{0 j_m})^{1 - \kappa_m} \times$$
$$\times (1 - \delta_{0 i_m})^{\kappa_m}(1 - \delta_{0 j_m})^{\kappa_m}$$
$$= \sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa}\left(\sum_i a_i^2 \prod_{m=1}^{M} \delta_{0 i_m}^{1 - \kappa_m}(1 - \delta_{0 i_m})^{\kappa_m}\right) \times$$
$$\times \left(\sum_j b_j^2 \prod_{m=1}^{M} \delta_{0 j_m}^{1 - \kappa_m}(1 - \delta_{0 j_m})^{\kappa_m}\right)$$
$$= \sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa}\left(\sum_i a_i^2 \tilde{\delta}_{i,\kappa}\right)\left(\sum_j b_j^2 \tilde{\delta}_{j,\kappa}\right)$$
$$= \sum_{\kappa \in \{0,1\}^M} \frac{(\ell_A)_\kappa (\ell_B)_\kappa}{d_\kappa} = (\ell_A, \ell_B) \tag{C11}$$

from which the Proposition follows.

$\square$

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

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

$$\mathbb{E}_{\theta_1, \theta_2}\left\{\operatorname{Tr}\left[AU_{\theta_1}^\dagger \Lambda(V_{\theta_2} B V_{\theta_2}^\dagger) U_{\theta_1}\right]^2\right\} = (\ell_A, T\ell_B) \tag{C12}$$

where T is the locality transfer matrix associated to Λ.

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

$$\mathbb{E}_{\theta_2}\mathbb{E}_{\theta_1}\left\{\operatorname{Tr}\left[AU_{\theta_1}^\dagger \tilde{B}_{\theta_2} U_{\theta_1}\right]^2\right\}$$
$$= \mathbb{E}_{\theta_2}\left\{(\ell_A, \ell_{\tilde{B}_{\theta_2}})\right\}$$
$$= \sum_{\kappa \in \{0,1\}^M} \frac{1}{d_\kappa}(\ell_A)_\kappa\, \mathbb{E}_{\theta_2}\left\{(\ell_{\tilde{B}_{\theta_2}})_\kappa\right\} \tag{C13}$$

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

$$\mathbb{E}_{\theta_2}\left\{(\ell_{\tilde{B}_{\theta_2}})_\kappa\right\} = \sum_i \mathbb{E}_{\theta_2}\left\{\mathrm{Tr}\left[P_i\tilde{B}_{\theta_2}\right]^2\right\}\tilde{\delta}_{i,\kappa}$$

$$= \sum_i \mathbb{E}_{\theta_2}\left\{\mathrm{Tr}\left[\Lambda^\dagger(P_i)V_{\theta_2}^\dagger BV_{\theta_2}\right]^2\right\}\tilde{\delta}_{i,\kappa}$$

$$= \sum_{\lambda\in\{0,1\}^M}\frac{1}{d_\lambda}\sum_{i,j,k}\mathrm{Tr}\left[P_i\Lambda(P_j)\right]^2\mathrm{Tr}\left[P_kB\right]^2\tilde{\delta}_{j,\lambda}\tilde{\delta}_{k,\lambda}\tilde{\delta}_{i,\kappa}$$

$$= \sum_{\lambda\in\{0,1\}^M}\left(\frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\left[P_i\Lambda(P_j)\right]^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda}\right)\times$$

$$\times\left(\sum_k\mathrm{Tr}\left[P_kA\right]^2\tilde{\delta}_{k,\lambda}\right)$$

$$= \sum_{\lambda\in\{0,1\}^M}T_{\kappa,\lambda}(\ell_B)_\lambda = (T\ell_B)_\kappa$$

$$\text{(C14)}$$

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

Iterated application of Proposition C.2 for an initial state ρ, an observable H, and a general intermediate quantum channel $\mathcal{E}$, yields Eq. (7) of the main text.

Appendix D: Proof of Theorem IV.1

The proof of Theorem IV.1 is based on the characterization of the general LTM for the Hermitian adjoint $\mathcal{E}^\dagger$ of arbitrary quantum channel. To do so, several aspects of non-negative matrix theory, as well as the contractivity properties of $\mathcal{E}^\dagger$ are key. For 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 IV.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 D.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 D.2 (Irreducible matrix). A $n \times n$ non-negative matrix T is said to be *irreducible* if for two arbitrary indices $i,j = 1,...n$, 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 D.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. $v_i, w_i > 0 \ \forall i = 1,...n$,*

(c) *v and w are unique up to a scalar multiple, and hence can be taken to be normalized, i.e. $v^t w = 1$,*

(d) *$r \geq |\lambda|$, for all eigenvalue λ of T,*

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

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

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

Theorem D.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*

$$y^t T \leq s y^t \qquad \text{(D1)}$$

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 D.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*

$$\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} T^l / r^l = P \qquad (D2)$$

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

$$\lim_{L \to \infty} T^L / r^L = P \qquad (D3)$$

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. (D4).

$$T = \begin{pmatrix} \boxed{T_0} & & & & \boxed{R_0} \\ & \boxed{T_1} & & & \boxed{R_1} \\ & & \ddots & & \vdots \\ & & & \boxed{T_z} & \boxed{R_z} \\ & & & & \boxed{Q} \end{pmatrix} \qquad (D4)$$

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

$$\|A\|_p = \mathrm{Tr} \left[\left(\sqrt{A^\dagger A} \right)^p \right]^{1/p}. \qquad (D5)$$

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

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

$$\|\Lambda\|_{p \to q} := \sup_{A : \|A\|_p = 1} \|\Lambda(A)\|_q. \qquad (D6)$$

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

Linear maps that are also CPTP are known to always be contractive with respect to the 1-norm [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 D.1 (Hölder's inequality for variances)**.** *Let $A, B \in \mathbb{M}_n$ be hermitian matrices, and $\Phi_\theta : \mathbb{M}_n \to \mathbb{M}_n$ be a parameterized quantum channel. In particular, consider the L layered map of type Eq. (2). Then*

$$\left| \mathrm{Tr}\left[\Phi_\theta^L(A) B \right] \right| \leq \|\mathcal{E}\|_{p \to p}^L \|A\|_p \|B\|_q \quad \forall \theta \in \Theta \qquad (D7)$$

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

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

$$\begin{aligned} \left| \mathrm{Tr}\left[\Phi_\theta^L(A) B \right] \right| &\leq \|\Phi_\theta^L(A)\|_p \|B\|_q \\ &= \|U_{\theta_L} \mathcal{E}\left(\Phi_\theta^{L-1}(A) \right) U_{\theta_L}^\dagger \|_p \|B\|_q \\ &= \|\mathcal{E}\left(\Phi_\theta^{L-1}(A) \right)\|_p \|B\|_q \\ &\leq \|\mathcal{E}\|_{p \to p} \|\Phi_\theta^{L-1}(A)\|_p \|B\|_q \quad \forall \theta \in \Theta \end{aligned} \qquad (D8)$$

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. $\qquad\square$

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

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

$$\mathcal{N}'(P_0) = P_0 + \sum_{i>0} t_i P_i, \quad \mathcal{N}'(P_i) = \lambda_i P_i \,\forall i > 0 \qquad (D9)$$

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. (D9) by means of a change of basis. Furthermore, the constraint on the parameters follow, analogously to [32], by considering that any single qubit state must have bounded purity, namely $\text{Tr}\left[\rho^2\right] \leq 1$, and since $\mathcal{N}'$ is a channel, the same must hold for $\mathcal{N}'(\rho)$. Since any qubit state can be decomposed in terms of $\{P_i\}_i$ as $\rho = 1/\sqrt{2}P_0 + 1/\sqrt{2}\sum_{i>0}\alpha_i P_i$, $\alpha_i \in \mathbb{R}$, we get

$$\text{Tr}\left[\rho^2\right] = \frac{1 + \alpha_i^2}{2} \leq 1, \quad \text{Tr}\left[\mathcal{N}'(\rho)^2\right] = \frac{1 + \sum_{i>0}(t_i + \lambda_i \alpha_i)^2}{2} \leq 1$$
$$(D10)$$

respectively, which concludes the proof. $\qquad\square$

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

Theorem D.4 (Kadison-Schwarz inequality [51]). *Let $A, B \in \mathbb{H}_0$, and $\Lambda : \mathbb{M}_n \to \mathbb{M}_n$ be a CPU map. Then*

$$\Lambda(A)\Lambda(B) \leq \Lambda(AB). \qquad (D11)$$

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 D.3 (Limiting form of T). *Let $\Lambda : \mathcal{B} \to \mathcal{B}$ be a CPU map and T be the corresponding LTM. Then T and T^L take the form*

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

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

$$(D12)$$

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. (D4). 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

$$\begin{aligned} A_z^{(2)} &= T_z R_z + R_z Q \\ A_z^{(L)} &= T_z^{L-1} R_z + A^{(L-1)} Q \\ &= T_z^{L-1} R_z + \sum_{l=0}^{L-2} T_z^l R_z Q^{L-1-l} \qquad (D13) \\ &= \sum_{l=0}^{L-1} T_z^l R_z Q^{L-1-l} \end{aligned}$$

which gives the proposition. $\qquad\square$

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

Proposition D.1 (Spectral radius of T). *Let $\Lambda : \mathcal{B} \to \mathcal{B}$ be a CPU map and T be the corresponding LTM. Then T is contractive in the sense of the spectral radius, i.e. $\rho(T) \le 1$. 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.

$$Q = \begin{pmatrix} Q_1 & * & \cdots & * \\ & Q_2 & & * \\ & & \ddots & \vdots \\ & & & Q_k \end{pmatrix} \tag{D14}$$

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 B.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 $P_i^2 = \mathbb{1}/d \, \forall i$. Exploiting Eq. (B4), we can compute the column sum of T as

$$\begin{aligned} \sum_\kappa T_{\kappa,\lambda} &= \sum_\kappa \frac{1}{d_\lambda} \sum_{i,j} \text{Tr}\left[P_i \Lambda(P_j)\right]^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} \\ &= \frac{1}{d_\lambda} \sum_{i,j} \text{Tr}\left[P_i \Lambda(P_j)\right]^2 \tilde{\delta}_{j,\lambda} \left(\sum_\kappa \tilde{\delta}_{i,\kappa}\right) \\ &= \frac{1}{d_\lambda} \sum_{i,j} \text{Tr}\left[P_i \Lambda(P_j)\right]^2 \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \text{Tr}\left[\Lambda(P_j)^2\right] \tilde{\delta}_{j,\lambda} \\ &\le \frac{1}{d_\lambda} \sum_j \text{Tr}\left[\Lambda(P_j^2)\right] \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \tilde{\delta}_{j,\lambda} = 1 \end{aligned}$$
$$\tag{D15}$$

where the third and last equality follow from Lemma B.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. (D15) show *sub-stochasticity* of T. Indeed, this condition can be recast in vector form as $v^t T \le v^t$, where $v_\kappa = 1 \forall \kappa$. In particular, this holds for all irreducible blocks in the diagonal, which by the Subinvariance theorem implies $r_z \le 1$ and $r_{Q_k} \le 1$, giving $\rho(T) \le 1$. 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

$$\sum_\kappa (T - Q_k)_{\kappa,\lambda} > 0, \tag{D16}$$

which implies that $\exists \lambda$ s.t. $\sum_\kappa (Q_k)_{\kappa,\lambda} < 1$. Written in matrix form, this reads $v^t Q_k \le v^t$, $v^t Q_k \ne v^t$, which by the Subinvariance theorem, implies $r_{Q_k} \ne 1$. 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 D.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 $(v_z)_\kappa = 1 \forall \kappa$ 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 D.2. *Let T be a LTM, and let T_z be an irreducible block with $r_z < 1$. Then, as $L \to \infty$, $\|T_z^L\| \to 0$ and $\|A_z^{(L)}\| \to 0$ exponentially fast for any matrix norm $\|\cdot\|$. Similarly, also $\|Q^L\| \to 0$.*

Proof. The proposition can be proven using Gelfand's formula. In particular since $\lim_{L\to\infty} \|T_z^L\|^{1/L} = r_z$, we can always bound $\|T_z^L\| \le K\tau^L$, for some constant $K > 0$ and $\tau = r_z + \epsilon < 1$ for an arbitrarily small ϵ. In the same way, by Proposition D.1 a similar result can be obtained for Q. Finally, the absorption term $A_z^{(L)}$ can also be bounded using Lemma D.3. In that case we have

$$\begin{aligned} \|A_z^{(L)}\| &\le \sum_{l=0}^{L-1} \|T_z^L\| \|R_z\| \|Q^{L-1-l}\| \\ &\le \|R_z\| K_T K_Q \tau^l \kappa^{L-1-l} \le K_A \alpha^L \end{aligned} \tag{D17}$$

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 IV.1

As stated in the main text, the limiting value of quantum circuits of type Eq. (2) is obtained by studying the spectral properties of the LTM of the intermediate channel in the Heisenberg picture. In particular, the previous discussion suggests a limiting value for the variance of the form

$$\mathbb{V}_{\rho,H}^{\infty} = \sum_{z\,|\,\rho(T_z)=1} (\ell_\rho, w_z)(\ell_H)_z + (\ell_\rho, w_z)(A_z \ell_H) \quad \text{(D18)}$$

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

Theorem D.5 (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 to Eq. (D18), and we have*

$$\left| \frac{1}{L} \sum_{l=0}^{L} \mathbb{V}_{\rho,H}^l - \mathbb{V}_{\rho,H}^{\infty} \right| \in O\left(e^{-\beta L} \|H\|_2^2 \right), \quad \text{(D19)}$$

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

$$\left| \mathbb{V}_{\rho,H}^L - \mathbb{V}_{\rho,H}^{\infty} \right| \in O\left(e^{-\beta L} \|H\|_2^2 \right), \quad \text{(D20)}$$

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

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

$$P_z = \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} T_z^l = T_z^{(d\infty)} \frac{1}{d} \sum_{m=0}^{d-1} T_z^m \quad \text{(D21)}$$

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

$$\lim_{N\to\infty} A^{(dN+m)} = T^{(d\infty)} A^{(m)} + A^{(d\infty)} Q^m \quad \text{(D22)}$$

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

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

$$
\begin{aligned}
A^{(dN)} &= \sum_{l=0}^{Nd-1} T_z^l R_z Q^{dN-1-l} \\
&= \sum_{m=0}^{d-1} \sum_{n=0}^{N-1} T_z^{nd+m} R_z Q^{(N-1)d-nd+d-1-m} \\
&= \sum_{m=0}^{d-1} \sum_{n=0}^{N-1} T_z^{nd} T_z^m R_z Q^{d-1-m} (Q^d)^{N-n} \\
&= \sum_{n=0}^{N-1} T_z^{nd} A^{(d)} (Q^d)^{N-n} \\
&= T_z^{(d\infty)} A^{(d)} \sum_{n=0}^{N-1} (Q^d)^{N-n} + \sum_{n=0}^{N-1} \Delta_z^{(n)} A^{(d)} (Q^d)^{N-n}
\end{aligned}
$$
$$\text{(D23)}$$

where $\Delta_z^{(N)} = T_z^{dN} - T_z^{(d\infty)}$. Since T_z^d is block diagonal, and each of the d blocks T_{zm} is irreducible and aperiodic, we have that $T_{zm}^N \to P_{zm}$ exponentially fast, i.e. $\|T_{zm}^N - P_{zm}\| = \|\Delta_{zm}^{(N)}\| < K_{zm}\tau_{zm}^N$, for some $K > 0$ and $\tau < 1$. Then, $\|\Delta_z^{(N)}\| \le \sum_{m=0}^{d} \|\Delta_{zm}^{(N)}\| \le K_z \tau_z^N$, where $\tau = \max_m\{\tau_{zm}\} < 1$. Together with Proposition D.1, this implies that, the last term in Eq. (D23) approaches zero with the same exponential speed, similarly to what happens in Proposition D.2. Putting everything together, and considering that $\sum_{n=0}^{N-1} X^n \to (\mathbb{1} - X)^{-1} \,\forall X$ such that $\rho(X) < 1$, we have

$$A^{(d\infty)} = T^{(d\infty)} A^{(d)} (\mathbb{1} - Q^d)^{-1} \quad \text{(D24)}$$

At this point, by induction similarly to Lemma D.3, one can easily show that

$$A^{(dN+m)} = T^{(dN)} A^{(m)} + A^{(dN)} Q^m \quad \text{(D25)}$$

which yields the proposition. □

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

$$A_z = \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} A^{(l)} = \frac{1}{d} \sum_{m=0}^{d-1} T^{(d\infty)} A^{(m)} + A^{(d\infty)} Q^m. \quad \text{(D26)}$$

Recalling that the right eigenvector v_z can be explicitly calculated when $\rho(T_z) = 1$ (see Corollary D.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. (D19) and Eq. (D20) 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

$$\left| \frac{1}{L} \sum_{l=0}^{L} \mathbb{V}_{\rho,H}^l - \mathbb{V}_{\rho,H}^\infty \right| = \left| \left(\ell_\rho, \frac{1}{L} \sum_{l=0}^{L} T^l - T^\infty \ell_H \right) \right|$$
$$\leq \left\| \frac{1}{L} \sum_{l=0}^{L} T^l - T^\infty \right\| \sqrt{(\ell_\rho, \ell_\rho)(\ell_H, \ell_H)}.$$
$$(D27)$$

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. □

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 D.4 (Single qubit noise). *Let $\rho, H \in \mathcal{B}$ and let Φ_θ a be 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 $\mathcal{N} = \mathcal{N}_1 \otimes ... \otimes \mathcal{N}_n$ is a composition of single qubit quantum channels. Then*

$$\mathbb{V}_{\rho,H}^\infty = \sum_z \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z (A\ell_H)_z}{d_z}. \qquad (D28)$$

Proof. Without loss of generality, we can consider the single qubit channels $\mathcal{N}_m$ to be in their normal form of Lemma D.2. In particular, this holds due to the invariance of the LTM with respect to changes of local bases (Lemma B.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 $\text{Tr}\left[\mathcal{N}_m^\dagger(P_{j_m})^2\right] = t_{j_m}^2 + \lambda_{j_m}^2 \leq 1$ by Lemma D.2. 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 D.1 we know that $\sum_\kappa (T_z)_{\kappa,\lambda} = 1$. By Eq. (D15) this implies

$$\sum_\kappa (T_z)_{\kappa,\lambda} = \frac{1}{d_\lambda} \sum_j \text{Tr}\left[(W^\dagger \mathcal{N}^\dagger(P_j)W)^2 \right] \tilde{\delta}_{j,\lambda}$$
$$= \frac{1}{d_\lambda} \sum_j \text{Tr}\left[\mathcal{N}^\dagger(P_j)^2 \right] \tilde{\delta}_{j,\lambda} \qquad (D29)$$
$$= \frac{1}{d_\lambda} \sum_j \prod_{m=1}^M \text{Tr}\left[\mathcal{N}_m^\dagger(P_{j_m})^2 \right] \tilde{\delta}_{j,\lambda} = 1.$$

Since each term in the product is upper-bounded by 1, Eq. (D29) implies $\text{Tr}\left[\mathcal{N}_m^\dagger(P_{j_m})^2\right] = t_{j_m}^2 + \lambda_{j_m}^2 = 1 \, \forall j_m$. This condition is only compatible with Lemma D.2 if $t_{j_m}^2 = 0$ and $\lambda_{j_m}^2 = 1$. This result can now be used to show that, the adjoint of $T_z^\dagger$ of the LTM T_z, must also be column stochastic. Indeed, if we consider the expansion of $WP_jW^\dagger$ with respect to the normalized Pauli basis, we can get

$$\sum_\kappa (T_z)_{\kappa,\lambda}^\dagger = \frac{1}{d_\lambda} \sum_j \text{Tr}\left[\mathcal{N}(WP_jW^\dagger)^2 \right] \tilde{\delta}_{j,\lambda}$$
$$= \frac{1}{d_\lambda} \sum_j \text{Tr}\left[\left(\sum_i \text{Tr}\left[P_i W P_j W^\dagger\right] \mathcal{N}(P_i) \right)^2 \right] \tilde{\delta}_{j,\lambda}$$
$$= \frac{1}{d_\lambda} \sum_j \text{Tr}\left[\left(\sum_i \text{Tr}\left[P_i W P_j W^\dagger\right] \prod_m \lambda_{i_m} P_i \right)^2 \right] \tilde{\delta}_{j,\lambda}$$
$$= \frac{1}{d_\lambda} \sum_{i,j} \prod_m \lambda_{i_m}^2 \text{Tr}\left[P_i W P_j W^\dagger\right]^2 \tilde{\delta}_{j,\lambda}$$
$$= \frac{1}{d_\lambda} \sum_j \text{Tr}\left[W P_j^2 W^\dagger \right] \tilde{\delta}_{j,\lambda} = 1.$$
$$(D30)$$

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

$$\sum_\lambda (T_z)_{\kappa,\lambda} d_\lambda = \sum_\lambda (T_z)_{\lambda,\kappa}^\dagger d_\kappa = d_\kappa \qquad (D31)$$

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

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

Corollary D.2 (Unitary circuits). *For unitary circuits of type Eq. (2), we have*

$$\mathbb{V}_{\rho,H}^\infty = \sum_{z>0} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} \qquad (D32)$$

Proof. This form of the variance is a special case of Proposition D.4, putting $\mathcal{N}(\rho) = \rho$, and noting that the absorption terms must vanish. In particular, this follows from Eq. (D15), observing that unitary channels saturate Kadison-Schwarz inequality, which combined with Corollary D.1 imply $Q = 0$. □

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

Corollary D.3 (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*

$$\mathbb{V}^\infty_{\rho,H} = \frac{(A\ell_H)_0}{d}. \tag{D33}$$

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

Proof. This form of the variance is a special case of Eq. (D18), where all absorbing components vanish, and we have $T = Q$. In particular, this follows from Eq. (D15), observing that the above condition implies that for $\mathcal{E}^\dagger$ Kadison-Schwarz inequality is strict, which combined with Corollary D.1 imply $T_z = 0 \,\forall z > 0$, leaving only T_0. Finally, the corollary follows from the normalization condition $\mathrm{Tr}\,[\rho] = 1$ on ρ, which ensures $(\ell_\rho)_0 = 1/d$. $\square$

Appendix E: Variance lower bounds

In this section we prove the general variance lower bounds outlined in Section IV.

1. Proof of Theorem IV.2

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

Theorem E.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*

$$\mathbb{V}^L_{\rho,H} \geq \alpha^L(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{E1}$$

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

$$
\begin{aligned}
(\ell_\rho, T_l \ell_H) &= \sum_{\kappa,\lambda} \frac{(\ell_\rho)_\kappa T_{\kappa,\lambda}(\ell_H)_\lambda}{d_\kappa} \\
&\geq \sum_{\kappa \in K} \frac{(\ell_\rho)_\kappa T_{\kappa,\kappa}(\ell_H)_\kappa}{d_\kappa} \\
&\geq \alpha_l \sum_{\kappa \in K} \frac{(\ell_\rho)_\kappa (\ell_H)_\kappa}{d_\kappa} = \alpha_l(\ell_\rho, \ell_{\mathcal{K}(H)})
\end{aligned}
\tag{E2}
$$

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

(a) $\alpha > 0$, $\alpha \in \Omega(1)$ and $L \in O(\log n)$,

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

is satisfied, then

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{E3}$$

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

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

$$
\begin{aligned}
-\log(F(n)) &< -L\log\left(1 - C\frac{\log(n)}{L}\right) \\
&= C\log(n)\left(1 + C\frac{\log(n)}{2L} + O\left(\frac{\log^2(n)}{L^2}\right)\right) \in O(\log(n))
\end{aligned}
\tag{E4}
$$

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

2. Small-angle initializations lower bounds

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.3, 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*

$$\mathbb{E}_\phi\{T_\phi\} \geq T, \tag{E5}$$

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

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

$$
\begin{aligned}
\mathrm{Tr}\,[\mathcal{E}(A)B]^2 &= \mathrm{Tr}\,[A\mathcal{E}^\dagger(B)]^2 \\
&= \mathbb{E}_\phi\left\{\mathrm{Tr}\left[A\mathcal{E}_\phi^\dagger(B)\right]\right\}^2 \\
&\leq \mathbb{E}_\phi\left\{\mathrm{Tr}\left[A\mathcal{E}_\phi^\dagger(B)\right]^2\right\} \quad \forall A, B
\end{aligned}
\tag{E6}
$$

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

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

$$
\begin{aligned}
\mathbb{V}_{\rho,H}^L &= \mathbb{E}_\phi \mathbb{E}_\theta \left\{ \mathrm{Tr}\left[\Phi_{\theta,\phi}(\rho)H\right]^2 \right\} \\
&= \mathbb{E}_\phi \left\{ (\ell_\rho, T_\phi \ell_H) \right\} = (\ell_\rho, \mathbb{E}_\phi\{T_\phi\}\ell_H).
\end{aligned}
\tag{E7}
$$

Combining this observation with Corollary E.1, we can get the QResNet lower bound of Proposition IV.2.

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

$$
\mathbb{V}_{\rho,H}^L \geq F(n)(\ell_\rho, \ell_H)
\tag{E8}
$$

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

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 definiton, we have

$$
\begin{aligned}
T_{\kappa,\kappa} &= \frac{1}{d_\kappa} \sum_{i,j} \mathrm{Tr}\left[\mathbb{E}_\phi\{P_i \mathcal{E}_\phi^\dagger(P_j)\}\right]^2 \tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\kappa} \\
&\geq \frac{1}{d_\kappa} \sum_i \mathrm{Tr}\left[\mathbb{E}_\phi\{P_i \mathcal{E}_\phi^\dagger(P_i)\}\right]^2 \tilde{\delta}_{i,\kappa} \\
&= \frac{1}{d_\kappa} \sum_i \mathrm{Tr}\left[\mathbb{E}_\phi\{P_i e^{-i\phi G} P_i e^{i\phi G}\}\right]^2 \tilde{\delta}_{i,\kappa}
\end{aligned}
\tag{E9}
$$

Since $\sigma^2 \to 0$ as $n \to \infty$, 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. (E9), we get

$$
\begin{aligned}
T_{\kappa,\kappa} &\approx \frac{1}{d_\kappa} \sum_i \mathrm{Tr}\Big[\mathbb{E}_\phi\{P_i(\mathbb{1} + i\phi G - \phi^2 G^2/2)P_i(\mathbb{1} + \\
&\quad + i\phi G - \phi^2 G^2/2)\}\Big]^2 \tilde{\delta}_{i,\kappa} \\
&= \frac{1}{d_\kappa} \sum_i \left(1 - \mathbb{E}_\phi\{\phi^2\}\left(\mathrm{Tr}\left[G^2 P_i^2\right] - \mathrm{Tr}\left[P_i G P_i G\right]\right)\right)^2 \tilde{\delta}_{i,\kappa} \\
&\geq 1 - 4\|G\|_2^2\sigma^2 \quad \forall \kappa \in \{0,1\}^M
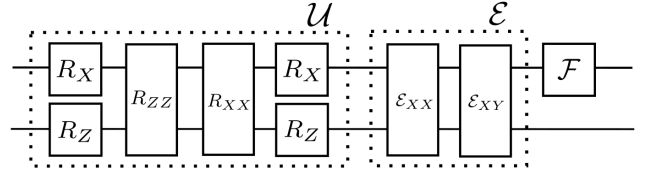\end{aligned}
\tag{E10}
$$

FIG. 3. Single layer structure used in the numerical examples. The circuit acts on a two-qubit register and is divided into three blocks, each representing either unitary gates ($\mathcal{U}$) or noise channels ($\mathcal{E}$ and $\mathcal{F}$).

Exploiting Corollary E.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. $\qquad\square$

Appendix F: NIBP-free landscapes under unital noise

To illustrate the emergence of non-trivial behavior in the loss variance $\mathbb{V}_{\rho,H}$ under unital noise channels, we focus on two representative examples. Both involve small quantum circuits composed of $n = 2$ qubits and correlated noise. In these cases, the presence of non-unitary layers is either irrelevant or beneficial to the overall magnitude of the variance.

We begin with the circuit depicted in Fig. 3, which features a unitary layer $\mathcal{U}$ as shown, combined with a fixed noise channel $\mathcal{E}_l = \mathcal{E} = \mathcal{E}_{XX} \circ \mathcal{E}_{XY} \; \forall l$. Specifically, $\mathcal{E}$ is defined as the composition of two dephasing-like channels on the XX and XY bases respectively, namely $\mathcal{E}_{XX}(\rho) = (1-p)\rho + pXX\rho XX$ and $\mathcal{E}_{XY}(\rho) = (1-p)\rho + pXY\rho XY$.

By analyzing the generators of the parameterized gates R_X and R_{ZZ} (i.e., X and ZZ), one can verify that they commute with both the other rotations appearing in $\mathcal{U}$ (i.e., R_Z and R_{XX}) and with the noise channel for any choice of parameters. Consequently, their action can be moved to the end of the circuit, leaving a mixed channel at the beginning. Notably, the state $\rho = |00\rangle\langle 00| + |11\rangle\langle 11|$ is a fixed point of this mixed channel, as it is singularly as fixed point of each of the remainig operations. As a result, by choosing ρ as the initial state, the circuit is effectively equivalent to its purely unitary counterpart and thus free of NIBP. This is confirmed by an analytical computation of the variance via Eq. (9), which yields a non-vanishing result that is independent of the number of layers. In particular, Fig. 4 numerically corroborates the prediction of $\mathbb{V}_{\rho,H}^L = 1$ for $H = ZZ + XI + YZ$.

Furthermore, by substituiting $\mathcal{U}$ with $\mathcal{U}'$, obtained by replacing the first RX gate with an RY gate, Fig. 4 shows how the NIBP phenomenon reappears, highlighting again the crucial role in the interplay between noise and circuit structure.
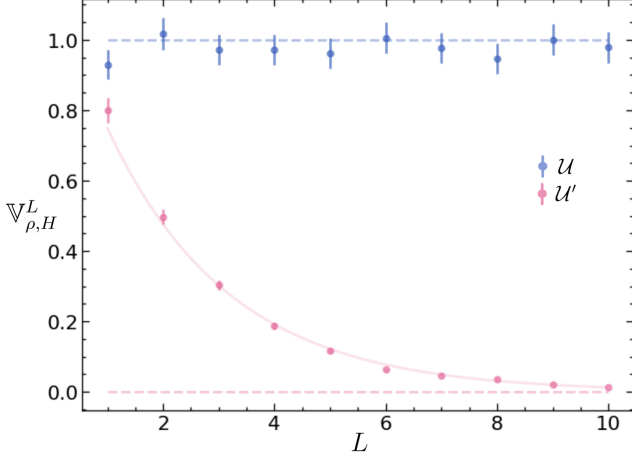
FIG. 4. Example of NIBP-free unital, non-unitary channel. Dotted lines represent the theoretical deep-circuit limiting values for the variance in the case of layers defined by $\mathcal{U} \circ \mathcal{E}$ (blue) and $\mathcal{U}' \circ \mathcal{E}$ (pink), where $\mathcal{U}'$ is obtained from $\mathcal{U}$ in Fig. 3 by substituting the first RX gate with an RY gate. The solid pink line is an exponential fit of the convergence to the limiting value.

As a second example, we change initial state, setting it to $\rho = \mathbb{1} \otimes |0\rangle \langle 0|$, while keeping H. By virtue of this choice, the variance of the unitary circuit is now identically zero, as shown in Fig. 5. Here, we consider a modified noise channel $\mathcal{E}' = \mathcal{E} \circ \mathcal{F}$, where $\mathcal{F}$ is a single-qubit channel applied to the first qubit, defined as $\mathcal{F}(\rho) = (1-p)\rho + p\mathcal{A}(\rho)$, and

$$
\begin{aligned}
\mathcal{A}(\rho) = &\frac{q}{\sqrt{2}} \operatorname{Tr}[X\rho] X + \frac{q}{\sqrt{2}} \operatorname{Tr}[X\rho] Y \\
&+ (1-q) \operatorname{Tr}[Y\rho] Y \\
&+ \frac{q}{\sqrt{2}} \operatorname{Tr}[Z\rho] Z + \frac{q}{\sqrt{2}} \operatorname{Tr}[Z\rho] Y
\end{aligned} \tag{F1}
$$

By computing its Choi matrix, one can confirm that $\mathcal{F}$ is completely positive and trace-preserving (CPTP), and therefore constitutes a valid quantum channel whenever $0 \le p \le 1$ and $0 \le q \le 1/2$. Although $\mathcal{E}'$ is still not strictly contractive, it is *contractive enough* to satisfy Corollary A.2, and thus provably induces the NIBP effect, as corroborated numerically in Fig. 5. Interestingly however, the behavior of $\mathbb{V}_{\rho,H}$ under this channel is not monotonic with respect to the circuit depth: it initially increases before eventually decaying to zero in the deep-circuit limit, thus temporarily enhancing the variance. This observation motivates the consideration of a non-uniform noise profile, where $\mathcal{E}_l \ne \mathcal{E}_{l'}$ for some l, l', defined as:

$$
\mathcal{E}_l'' = \begin{cases} \mathcal{E}' & \text{if } l \le const \\ \mathcal{E} & \text{otherwise} \end{cases} \tag{F2}
$$

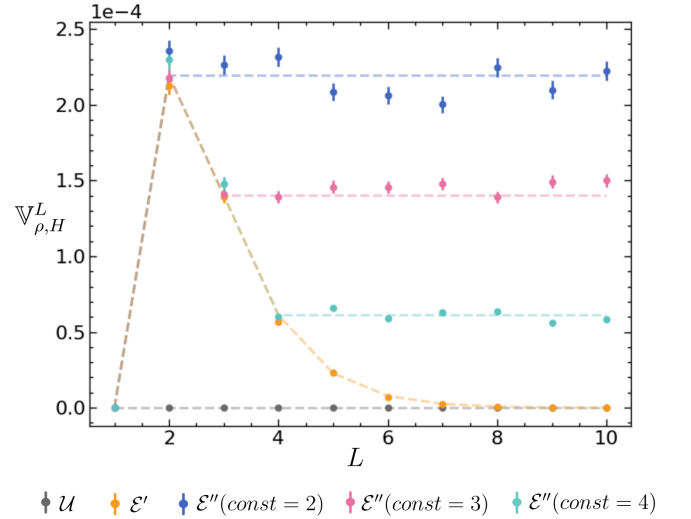In such a setup, the circuit benefits from the absence of concentration observed in the first example, as well

FIG. 5. Variance improvment with unital channels. Dotted lines represent the theoretical prediction given by Eq. (7), while points indicate the numerically estimated circuit variances.

as the transient variance enhancement from the second, converging to non-zero value dependent on the constant rather than the number of layers. This hybrid effect is clearly demonstrated in Fig. 5. The mechanism at play here is at the core of the *absorption* effects shown in the main text.

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

Appendix G: Non-unital noise and entanglement

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

1. Numerics

Exploiting the knowledge about the structure of the absorption matrix A, derived from Theorem IV.1, it is possible to study the scaling of the variance $\mathbb{V}_{\rho,H}^\infty$ 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

$$
\mathcal{E}_c(\rho) = (1-p)\mathcal{E}(\rho) + p\tilde{\rho}, \tag{G1}
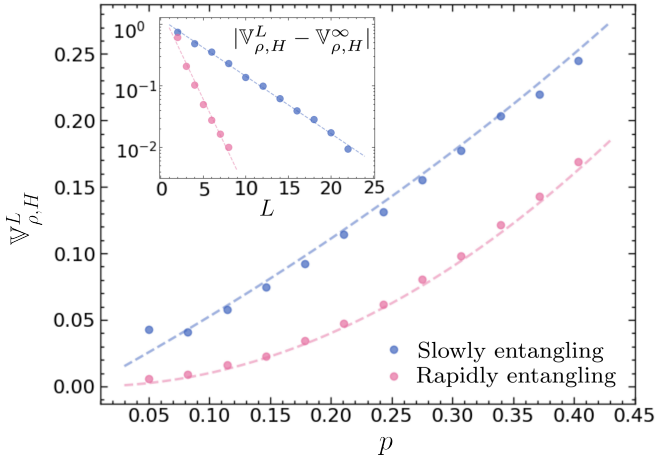$$

FIG. 6. 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. (G5) and Eq. (G4) 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.

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 $\tilde{\Phi}(\rho) = (1 - p)\rho + p\tilde{\rho}$ and of $\mathcal{E} \circ \mathcal{U}_{\theta_l}$. Clearly $\tilde{\rho}$ is the unique fixed-point of $\tilde{\Phi}$, while 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 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 IV.1 allows to quantify the variance scaling in the limit of a rapidly entangling and slowly entangling channel, i.e. $\beta \to \infty$ and $\beta \to 0$ respectively. In particular, for both limits, the LTM approaches a projection, namely the dominant eigenprojection in the former and the identity in the latter. In these cases, $\mathbb{V}_{\rho,H}^{\infty}$ is given by the following Lemma.

Lemma G.1. *Let $\mathcal{E}_c$ be a quantum channel of type Eq. (G1), with $0 < p \leq 1$ and let T be the transition matrix of $\mathcal{E}^\dagger$. Then*

$$\mathbb{V}_{\rho,H}^{\infty} = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1}\ell_H). \quad (G2)$$

In particular, if T is a projection, then

$$\mathbb{V}_{\rho,H}^{\infty} = \left(\frac{p}{2-p} - p^2\right)(\ell_{\tilde{\rho}}, T\ell_H) + p^2(\ell_{\tilde{\rho}}, \ell_H). \quad (G3)$$

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 G.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

$$\mathbb{V}_{\rho,H}^{\infty} = \frac{p}{2-p}(\ell_{\tilde{\rho}}, \ell_H). \quad (G4)$$

The result shows that first term in Eq. (G3) 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 [32]. Conversely, in the opposite limit, the dependence on p is more complex, as it now depends on the irreducible components T_z. For simplicity, if we consider the case of a highly expressive ansatz, we may take T to have only one irreducible component T_1. In this case, the last term in Eq. (G3) dominates, and we get a *quadratic* dependence on p up to an exponentially vanishing correction, namely

$$\mathbb{V}_{\rho,H}^{\infty} = p^2(\ell_{\tilde{\rho}}, \ell_H) + O(4^{-n}). \quad (G5)$$

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 [52] to construct and optimize the PQC described in the following.

Regarding the initial state, we use $\rho = (|0\rangle\langle 0|)^{\otimes n}$ for simplicity. As for the channel, we consider maps $\mathcal{E}_c$ of the family $\mathcal{E}_c = \mathcal{N} \circ \mathcal{E}$, $\mathcal{E}_c(\rho) = (1-p)\mathcal{E}(\rho) + p\tilde{\rho}$. In particular, $\mathcal{E}$ is a unitary, entangling channel depicted in Fig. 7 and $p \in (0,1]$ represents the noise strength of the noise map $\mathcal{N}(\rho) = (1-p)\rho + p\tilde{\rho}$ with fixed point $\tilde{\rho}$. Specifically, we fix $\tilde{\rho}$ to be a highly entangled, pure state, i.e. the GHZ state $\tilde{\rho} = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})(\langle 0|^{\otimes n} + \langle 1|^{\otimes n})/2$.

Concerning H, we fix $H = h\sum_{k=1}^{n} 2^{n/2} Z_k \otimes Z_{k+1}$, as it represents the simplest observable involving all n qubits while having a non-vanishing normalization factor $(\ell_{\tilde{\rho}}, \ell_H)$. In particular, as shown in Appendix G 3, 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. Finally, the entangling part is chosen according to Fig. 7. While the number of layers L needed to reach convergence to $\mathbb{V}_{\rho,H}^{\infty}$ is logarithmic, as shown in Theorem IV.1 and numerically assessed in the inset of Fig. 6, the mixing speed varies depending
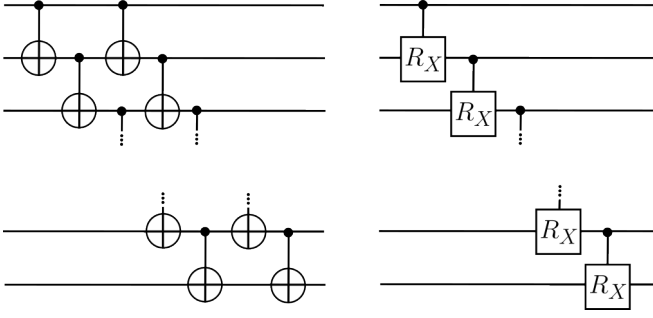
FIG. 7. Entangling unitaries used in the examples. On the left, the rapidly entangling configuration is composed of a double cascade of CNOT gates, while on the right, the slowly entangling one is composed of a single cascade of controlled RX gates, where $RX(\theta) = e^{i\theta X/2}$ and X is the Pauli X gate. In particular, we fixed $\theta = \pi/20$.

on the entangling part. For this reason, $L = 8$ layers are sufficient in the rapidly entangling case, but $L = 20$ are necessary in the slowly entangling case. Finally, all simulations are performed using $n = 10$ qubits. The results are shown in Fig. 6. The plots show both the quadratic and linear scaling with p predicted by our model, as well as an exponential decay in the difference $|\mathbb{V}^L_{\rho,H} - \mathbb{V}^\infty_{\rho,H}|$ for fixed noise rates. A slight deviation from the predicted scaling is observed in the slowly entangling setting for $p \approx 0$. This effect can be explained by the finite amount of layers used: the slow speed of convergence imposed by the condition $T \approx \mathbb{1}$ prevents the variance to reach the asymptotic limit $\mathbb{V}^\infty_{\rho,H}$, while the action of the noise is still too weak to erase the contribution of the first layers, hence deviating from the predicted behaviour. This same phenomenon is not observed in the rapidly entangling case, as entanglement production already exponentially suppresses those contributions.

2. Calculation of the absorption term

Explicit computation of the absorption matrix $A = R(\mathbb{1} - Q)^{-1}$ is a 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. In particular, here we prove Lemma G.1, which we repeat for convenience.

Lemma G.2. *Let $\mathcal{E}_c$ be a quantum channel of type Eq. (G1), with $0 < p \le 1$ and let T be the transition matrix of $\mathcal{E}^\dagger$. Then*

$$\mathbb{V}^\infty_{\rho,H} = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1}\ell_H). \qquad (G6)$$

In particular, if T is a projection, then

$$\mathbb{V}^\infty_{\rho,H} = \left(\frac{p}{2-p} - p^2\right)(\ell_{\tilde{\rho}}, T\ell_H) + p^2(\ell_{\tilde{\rho}}, \ell_H). \qquad (G7)$$

Proof. The first result follows directly from Theorem D.5, in particular Corollary D.3, by computation of A_0. In particular, we can explicitly compute R_0 element wise as

$$
\begin{aligned}
(R_0)_{0,\kappa} &= \frac{1}{d_\kappa} \sum_j \text{Tr}\left[\frac{\mathbb{1}}{\sqrt{d}}\mathcal{E}_c^\dagger(P_j)\right]^2 \tilde{\delta}_{j,\kappa} \\
&= \frac{1}{d_\kappa} \sum_j \text{Tr}\left[\mathcal{E}_c\left(\frac{\mathbb{1}}{\sqrt{d}}\right)P_j\right]^2 \tilde{\delta}_{j,\kappa} \\
&= \frac{1}{d_\kappa} \sum_j \left((1-p)\text{Tr}\left[\Phi\left(\frac{\mathbb{1}}{\sqrt{d}}\right)P_j\right]\right. \qquad (G8) \\
&\quad \left. + p\sqrt{d}\,\text{Tr}\left[\tilde{\rho}P_j\right]\right)^2 \tilde{\delta}_{j,\kappa} \\
&= dp^2 \frac{1}{d_\kappa} \sum_j \text{Tr}\left[\tilde{\rho}P_j\right]^2 \tilde{\delta}_{j,\kappa} = dp^2(\ell_{\tilde{\rho}})_\kappa
\end{aligned}
$$

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 $A_0 = P_0 R_0 (\mathbb{1} - Q)^{-1}$, and we get

$$(A_0)_{0,\lambda} = dp^2(\ell_{\tilde{\rho}})_\kappa(\mathbb{1} - (1-p)^2 T)^{-1}_{\kappa,\lambda}. \qquad (G9)$$

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

$$(\ell_\rho, A_0\ell_H) = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1}\ell_H). \qquad (G10)$$

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

$$
\begin{aligned}
&(\mathbb{1} - (1-p)^2 T)^{-1} \\
&= (\mathbb{1} - T) + (1 - (1-p)^2)^{-1}T \\
&= (\mathbb{1} - T) + \frac{1}{p(2-p)}T \qquad (G11) \\
&= \mathbb{1} + \left(\frac{1}{p(2-p)} - 1\right)T
\end{aligned}
$$

which gives the result. $\qquad\square$

3. Choice of the system for the numerical example

In this section we give an explicit construction of the noise channels used in the numerical example shown in the main text, as well as the choice and normalization procedure of the observable H used. In particular, the analysis focuses on the case of unitary entangling map Φ

whose transfer matrix is well approximated by a projector, and a highly entangled, pure fixed point, namely the GHZ state $\tilde{\rho} = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})(\langle 0|^{\otimes n} + \langle 1|^{\otimes n})/2$. Thanks to Lemma G.1, we know that the main contribution to $\mathbb{V}_{\rho,H}^{\infty}$ in this setting comes from the term $(\ell_{\tilde{\rho}}, \ell_H)$, so to numerically assess the results, it is useful to choose the observable H in order to normalize this factor. In particular, given the qubit structure, we can express the $\tilde{\rho}$ in terms of the *normalized* Pauli basis $\{\mathbb{1}, X, Y, Z\}^{\otimes n}$, which allows to easily compute the locality vectors. Using the spectral decomposition of the Pauli matrices, it is easy to see that

$$|0\rangle\langle 0| = \frac{\mathbb{1} + Z}{\sqrt{2}}, \quad |0\rangle\langle 1| = \frac{X + iY}{\sqrt{2}},$$
$$|1\rangle\langle 1| = \frac{\mathbb{1} - Z}{\sqrt{2}}, \quad |1\rangle\langle 0| = \frac{X - iY}{\sqrt{2}}. \tag{G12}$$

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

Theorem G.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*

$$(A + \omega B)^{\otimes n} = \sum_{j=0}^{n} \omega^j \sum_{\sigma \in S} \sigma(A^{\otimes(n-j)} \otimes B^{\otimes j}) \tag{G13}$$

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 G.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*

$$\frac{(A + \omega B)^{\otimes n} + (A - \omega B)^{\otimes n}}{2} = \sum_{j=0}^{\lfloor n/2 \rfloor} \omega^{2j} \sum_{\sigma \in S} \sigma(A^{\otimes(n-2j)} \otimes B^{\otimes 2j}) \tag{G14}$$

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

Proof. This corollary follow directly from Theorem G.1, and noticing that all even-indexed terms in both $(A + \omega B)^{\otimes n}$ and $(A - \omega B)^{\otimes n}$ are equal, while odd-numbered terms have opposite sign and therefore cancel out. □

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

$$\tilde{\rho} = \frac{1}{2^{n/2}} \sum_{j=0}^{\lfloor n/2 \rfloor} \sum_{\sigma \in S} \sigma(\mathbb{1}^{\otimes(n-2j)} \otimes Z^{\otimes 2j}) \\ + (-1)^j \sigma(X^{\otimes(n-2j)} \otimes Y^{\otimes 2j}) \tag{G15}$$

As it is clear from Eq. (G15), 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 in *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. (G15) in the calculations of $(\ell_{\tilde{\rho}}, \ell_H)$. Finally, since each term in the sum is orthogonal, it gives an independent contribution of $1/(d^2 - 1)^2 = 1/9$, consequently by choosing $h = 9/n$ we have that $(\ell_{\tilde{\rho}}, \ell_H) = 1$ is normalized.

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

As Theorem IV.1 suggests, the generic circuit of Eq. (2) need not have a well-defined value for the deep circuit limit of its variance, namely the limit $\lim_{L \to \infty} \mathbb{V}_{\rho,H}^L$ no need to exists. As discussed in Appendix D, 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. 8. 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

$$\mathbb{V}_{\rho,H}^L = \begin{cases} \frac{(\|\rho\|_2^2 - 1/2)\|H\|_2^2}{3} & \text{if } L \text{ is even} \\ 0 & \text{if } L \text{ is odd} \end{cases} \tag{H1}$$

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:

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow T_0 = T_2 = (1), \ T_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{H2}$$

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

$$\lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} \mathbb{V}_{\rho,H}^l = \frac{(\|\rho\|_2^2 - 1/2)\|H\|_2^2}{6}, \tag{H3}$$

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.
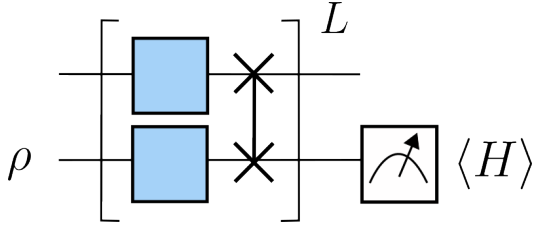
FIG. 8. Simple 2-qubit circuit, designed to have a non-convergent variance. In this construction, the channel $\mathcal{E}$ is chosen to be unitary and in particular to be a SWAP gate. Moreover, both the initial state ρ and the observable H are chosen to be non-trivial only on the second qubit, and are therefore represented as single qubit operators.

[1] R. P. Feynman, International Journal of Theoretical Physiscs **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. Borras, 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. Halimeh, Z. Holmes, S. Kühn, D. Lacroix, R. Lewis, D. Lucchesi, M. L. Martinez, F. Meloni, A. Mezzacapo, S. Montangero, 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. Babbush, 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] E. Cervero Martín, K. Plekhanov, and M. Lubasch, Quantum **7**, 974 (2023).

[13] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Nature Communications **12**, 10.1038/s41467-021-21728-w (2021).

[14] A. V. Uvarov and J. D. Biamonte, Journal of Physics A: Mathematical and Theoretical **54**, 245301 (2021).

[15] S. Thanasilp, S. Wang, N. A. Nghiem, P. Coles, and M. Cerezo, Quantum Machine Intelligence **5**, 10.1007/s42484-023-00103-6 (2023).

[16] 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).

[17] 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).

[18] 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].

[19] C.-Y. Park and N. Killoran, Quantum **8**, 1239 (2024).

[20] C.-Y. Park, M. Kang, and J. Huh, Hardware-efficient ansatz without barren plateaus in any depth (2024), arXiv:2403.04844 [quant-ph].

[21] 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).

[22] 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].

[23] 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].

[24] 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).

[25] M. Schumann, F. K. Wilhelm, and A. Ciani, Quantum Science and Technology **9**, 045019 (2024).

[26] G. Crognaletti, G. Di Bartolomeo, M. Vischi, and L. Loris Viteritti, Quantum Science and Technology **10**, 015048 (2024).

[27] C. Tüysüz, S. Y. Chang, M. Demidik, K. Jansen, S. Vallecorsa, and M. Grossi, PRX Quantum **5**, 030314 (2024).

[28] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, A review of barren plateaus in variational quantum computing (2024), arXiv:2405.00781 [quant-ph].

[29] Y. Quek, D. Stilck França, S. Khatri, J. J. Meyer, and J. Eisert, Nature Physics 10.1038/s41567-024-02536-7 (2024).

[30] R. Takagi, H. Tajima, and M. Gu, Physical Review Letters 10.1103/physrevlett.131.210602 (2023).

[31] K. Tsubouchi, T. Sagawa, and N. Yoshioka, Physical Review Letters 10.1103/physrevlett.131.210601 (2023).

[32] A. A. Mele, A. Angrisani, S. Ghosh, S. Khatri, J. Eisert, D. S. França, and Y. Quek, Noise-induced shallow circuits and absence of barren plateaus (2024), arXiv:2403.13927 [quant-ph].

[33] B. Fefferman, S. Ghosh, M. Gullans, K. Kuroiwa, and K. Sharma, PRX Quantum 5, 030317 (2024).

[34] P. Singkanipa and D. A. Lidar, Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and limit sets (2024), arXiv:2402.08721 [quant-ph].

[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. Sevior, Non-unitary quantum machine learning (2024), arXiv:2405.17388 [quant-ph].

[42] H. M. Wiseman and G. J. Milburn, Phys. Rev. A 47, 642 (1993).

[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).

[52] 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. Isacsson, D. Ittah, S. Jahangiri, 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, Pennylane: Automatic differentiation of hybrid quantum-classical computations (2022), arXiv:1811.04968 [quant-ph].