

Efficiently Learning Global Quantum Channels with Local Tomography

Zidu Liu^{1,*} and Dominik S. Wild^{1,2}

¹*Max Planck Institute of Quantum Optics, 85748, Garching, Germany*

²*Munich Center for Quantum Science and Technology (MCQST), 80799, Munich, Germany*



(Received 13 March 2026; revised 11 August 2026; accepted 21 September 2026; published 22 October 2026)

Scalable characterization of quantum processors is crucial for mitigating noise and imperfections. While randomized measurement protocols enable efficient access to local observables, inferring a globally consistent description of multi-qubit processes remains challenging. Here we introduce a local-to-global reconstruction framework for one-dimensional multi-qubit states and channels. The method is efficient provided that correlations, as quantified by the conditional mutual information, decay exponentially. In particular, we prove that under this assumption, the required number of samples scales polynomially with the system size and the inverse of the desired global reconstruction error, which we quantify by the trace-norm distance between the true and the reconstructed Choi states. Our approach is based on combining local shadow tomography with locally optimal recovery maps obtained by convex optimization. We supplement these rigorous guarantees by studying the performance of the protocol numerically for a system evolving under a local Lindbladian and a noisy, shallow circuit. By employing a tensor network representation, we reconstruct channels acting on up to 50 qubits and accurately recover global diagnostics such as the process fidelity, the Choi state purity, and Pauli-weight-resolved process matrix elements. Our work thus extends the powerful toolbox of local shadow tomography to scalable channel characterization with access to global properties.

DOI: [10.1103/51mt-ghpn](https://doi.org/10.1103/51mt-ghpn)

I. INTRODUCTION

The development of quantum technologies has been progressing at a rapid pace. Leading quantum computing platforms have now reached sizes of more than 100 qubits [1–4] and have carried out computational tasks that are beyond the reach of classical computers [5–8]. The ability to run quantum circuits with minimal noise on these devices crucially relies on their accurate characterization and calibration. This poses a formidable practical challenge since the systems are too large for complete tomography or full device-scale modeling and simulation. The challenge is expected to become even more significant as devices continue to grow in size and complexity on the path toward quantum error correction and fault-tolerant quantum computation.

Randomized schemes such as randomized benchmarking [9–11] and shadow tomography [12–18] address part of this challenge. However, randomized benchmarking probes only a restricted family of structured random

sequences and reports an averaged error rate, which may hide important gate- and context-dependent effects such as crosstalk between particular operations. While shadow tomography provides access to a larger class of observables, it necessitates deeper circuits to probe global properties. In this work, we overcome this limitation by making a physically motivated assumption that allows us to reconstruct the global state from local measurements, thereby providing access to nonlocal properties including the purity, entanglement entropy, and individual components of a process matrix.

The key assumption underlying our approach is a locality condition formulated in terms of the conditional mutual information (CMI). We focus on one-dimensional systems and assume that the CMI decays exponentially with the size of the conditioning subsystem. This assumption can be rigorously justified for Gibbs states of local Hamiltonians [19–21] and we expect it to hold more generally for shallow circuits in the presence of weak noise. The formal role of this assumption is to allow us to patch together information from small subsystems with controlled error, resulting in the protocol summarized in Fig. 1. We first perform tomography on subsystems that are large compared to the decay length of the CMI. From this information, we then reconstruct the global state using so-called recovery maps, each of which appends a single site to the current reconstruction while acting only on a small conditioning region formed by

*Contact author: zidu.liu@mpq.mpg.de

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

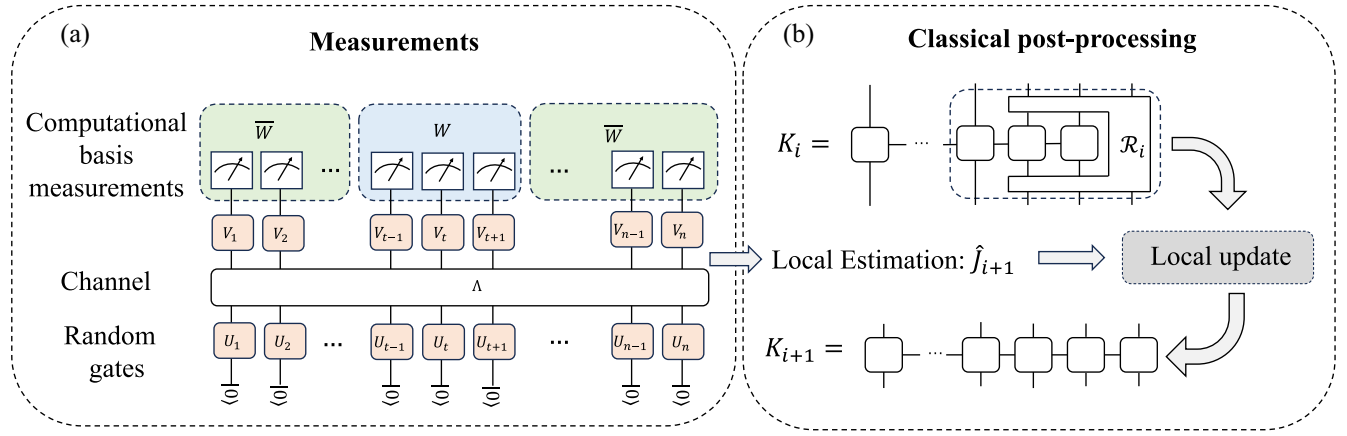


FIG. 1. Protocol for process matrix reconstruction. (a) In each experimental shot, we prepare $|0\rangle^{\otimes n}$, apply independently sampled single-qubit gates $U_1, \dots, U_n$ before the unknown channel Λ and $V_1, \dots, V_n$ prior to measurement in the computational basis. (b) Local-to-global stitching. The measurement records are first converted into estimates $\hat{J}_i$ of the reduced Choi states on overlapping windows of $2w + 1$ consecutive sites, where w denotes the reconstruction width. The global Choi state is then assembled one site at a time: at step i , the current reconstruction K_i , which is supported on sites $\{1, \dots, i\}$ and stored as an MPO, is extended to site $i + 1$ by a recovery map $\mathcal{R}_i$ acting only on the w sites $\{i - w + 1, \dots, i\}$. The map $\mathcal{R}_i$ is determined by the convex optimization in Eq. (5), which matches the marginal of the extended state on the window $\{i - 2w + 1, \dots, i + 1\}$ to the tomographic estimate $\hat{J}_{i+1}$. Iterating until site n is reached yields the global estimate $\hat{J} = K_n$.

the w preceding sites. We refer to w as the reconstruction width. The exponential decay of the CMI with the size of the conditioning region ensures that a reconstruction width growing only logarithmically with the system size and the inverse target error suffices, which underlies the efficiency of the protocol. We prove that $\text{poly}(n, 1/\varepsilon)$ samples are sufficient to reconstruct a (mixed) quantum state on n qubits with an error less than ε in trace norm.

When applied to quantum channels, our guarantees are correspondingly expressed in terms of the trace-norm distance between the normalized Choi states. Operationally, this metric quantifies the distinguishability of the two channels for the fixed product of Bell states used as input in the Choi representation, and it controls quantities such as the process fidelity and individual process-matrix elements. This metric is weaker than the worst-case diamond-norm distance, into which it can be converted at the expense of a dimensional factor that is exponential in the system size in general, although the factor can be made polynomial for structured channels such as Pauli channels (see Appendix D). The reconstructed state can be represented by a matrix product operator (MPO) [22–24], which permits the efficient computation of a wide variety of properties.

We showcase the efficiency of the approach by applying it to the problem of process tomography. In this context, our method recovers the Choi state corresponding to the probed quantum channel [25]. When available, process tomography is a powerful tool to gain fine-grained information about quantum gates, which may help to identify the cause of noise and imperfections. In practice, however, process tomography is limited to very small

systems owing to the fact that the process matrix describing a channel acting on n qubits is composed of 16^n parameters [26,27]. We are able to accurately reconstruct the process matrix of a channel acting on 50 qubits from estimates of reduced channels acting on blocks of 3 adjacent qubits and observe that our method works well even beyond the regime where the rigorous error bounds apply. We anticipate that the protocol will be most useful for the fine-grained characterization of few-layer operations in one-dimensional architectures, for example, a layer of parallel entangling gates in a trapped-ion or superconducting-qubit chain. In this setting, our method can detect crosstalk correlations between gates executed in parallel and, via global quantities like the Choi state purity, separate coherent from incoherent errors. This is complementary to randomized benchmarking discussed above, which reports averaged error rates, and to Pauli-noise learning [28], which efficiently characterizes the stochastic Pauli components but does not retain arbitrary coherent errors.

Before proceeding with the details of our work, we briefly comment on connections to related studies. Learning a general n -qubit channel in diamond norm is generally intractable. Recent results on the query and sample complexity of quantum process tomography establish costs that grow exponentially with the number of qubits [29–32]. Self-consistent gate-set tomography reconstructs the complete error structure of a gate set, but its cost also grows exponentially with the number of qubits [33]. In the setting of process tomography, the present framework accesses regimes beyond those previously treated for unitary channels [34], Pauli channels [28] and for channels

described by Kraus operators with low Pauli weight [35]. Moreover, when used for process tomography, our method does not require an assumption of Markovianity, in contrast to Lindbladian tomography [36,37]. Related machine-learning approaches have inferred reduced quantum channels on NISQ devices and local effective generators from noisy measurements, but not a globally consistent reconstruction of the full channel [38,39].

From a tensor-network perspective, our protocol can be viewed as a procedure for recovering an MPO representation of a quantum state under the assumption of exponentially decaying CMI, hence complementing other lines of work on MPO tomography. Previous methods for local-to-global reconstruction recover one-dimensional mixed states in MPO form from local marginals under an invertibility condition [40] and were demonstrated experimentally on up to 14 trapped-ion qubits [41]. Subsequent work [42] compared Petz-recovery-map-based and matrix-reconstruction-based methods. Compared to this line of research, we treat general quantum channels rather than states, working in the Choi representation. We also obtain end-to-end guarantees in which local process shadows feed directly into the reconstruction, with trace-norm and sample-complexity bounds under the exponentially decaying CMI assumption. Heuristic schemes have also been proposed that directly fit tensor-network models to experimental data [43–45]. There have additionally been several works providing rigorous guarantees under different assumptions. One such work assumes approximate factorization conditions that can be phrased in terms of Rényi-type approximate Markov behavior, and its theoretical guarantees are primarily expressed in Hilbert-Schmidt norm rather than trace norm [46]. Another recent work is based on a formal assumption concerning the spectral properties of MPO tensors, which, however, lacks a simple physical interpretation [47]. The reconstruction in the case of exponentially decaying CMI has been previously considered for the Belavkin-Staszewski relative entropy [48] and for the von Neumann entropy [49]. In contrast to the latter work, our analysis uses a locally optimal recovery strategy, and we provide tighter error bounds leading to a quadratic improvement in the scaling of the sample complexity. In a similar spirit, Gao and Yu [50] established sample-optimal tomography of quantum Markov chains by means of Petz recovery maps. Their analysis applies to states satisfying an exact Markov condition, whereas we allow for approximate Markovianity in the form of exponentially decaying CMI and provide an end-to-end local-shadow protocol for general channels.

II. SETTING

We consider an unknown noisy quantum channel Λ acting on n spins, each represented by a local Hilbert space $\mathbb{C}^d$. For simplicity, we will focus on the case of qubits

($d = 2$), although our results can be readily generalized to any constant local dimension d . Let $\sigma_0 = \mathbb{I}$ and $\sigma_1, \sigma_2, \sigma_3$ denote the single-qubit Pauli operators. For any $\mu \in \{0, 1, 2, 3\}^n$, we define the n -qubit Pauli operator $P_\mu = \sigma_{\mu_1} \otimes \cdots \otimes \sigma_{\mu_n}$. We will often represent the channel in terms of the process matrix χ , which is the unique $4^n \times 4^n$ matrix such that

$$\Lambda(\rho) = \sum_{\alpha\beta} \chi_{\alpha\beta} P_\alpha \rho P_\beta. \quad (1)$$

It is further convenient to use the Choi-Jamiołkowski representation of a channel [25,51]. To this end, we consider a doubled local Hilbert space $\mathbb{C}^d \otimes \mathbb{C}^d$, which supports the maximally entangled state $|\Phi\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^d |k\rangle_{\text{in}} \otimes |k\rangle_{\text{out}}$, where the subscripts serve as labels for the Hilbert spaces. Given a channel Λ , we define the Choi state

$$J(\Lambda) = (\mathcal{I}_{\text{in}} \otimes \Lambda_{\text{out}})(|\Phi\rangle\langle\Phi|), \quad (2)$$

where $|\Phi\rangle^{(n)} = |\Phi\rangle^{\otimes n}$ and $\mathcal{I}$ is the identity map. By the Choi-Jamiołkowski isomorphism, the Choi state is uniquely determined by the channel and vice versa. Moreover, for any completely positive and trace-preserving (CPTP) channel, the Choi state is positive and satisfies $\text{tr}_{\text{out}}[J(\Lambda)] = \mathbb{I}_{\text{in}}/d^n$. The process matrix χ corresponds to a representation of a Choi state in the orthonormal basis $|\alpha\rangle = (P_{\alpha,\text{in}} \otimes \mathbb{I}_{\text{out}})|\Phi\rangle^{(n)}$, i.e., $\chi_{\alpha\beta} = \langle\alpha|J(\Lambda)|\beta\rangle$.

Experimentally, we access Λ via the protocol shown in Fig. 1. In each shot we prepare the state $|0\rangle^{\otimes n}$, to which we apply independently sampled single-qubit gates $U = \bigotimes_{j=1}^n U_j$. We then apply the channel Λ followed by another layer of single-qubit gates $V = \bigotimes_{j=1}^n V_j$ before measuring each qubit in the computation basis to obtain an outcome bitstring $\mathbf{b}$. Each shot is thus fully described by the triple $(U, V, \mathbf{b})$. Our protocol works, in principle, for any set of single-qubit unitaries that is tomographically complete. For concreteness, we choose each U_i such that it prepares an eigenstate of the three single-qubit Pauli operators. Similarly, each V_j rotates the basis of the subsequent measurement into the eigenbasis of one of the three Pauli operators.

To keep the protocol scalable, we avoid manipulating the full n -qubit channel directly and instead reconstruct reduced channels on small, overlapping windows. For a window $W \subseteq \{1, \dots, n\}$ with complement $\bar{W}$, we define the reduced channel $\Lambda^{(W)}$ as the action of Λ on W when $\bar{W}$ is initialized in the maximally mixed state and finally traced out. Explicitly, for any state σ supported on W ,

$$\Lambda^{(W)}(\sigma) = \text{tr}_{\bar{W}} \left[\Lambda \left(\sigma \otimes \frac{1}{d^{|\bar{W}|}} \mathbb{I}_{\bar{W}} \right) \right]. \quad (3)$$

The corresponding Choi state satisfies $J(\Lambda^{(W)}) = \text{tr}_{\bar{W}}[J(\Lambda)]$, where the trace is performed over both the “in” and the “out” spaces associated with subsystem $\bar{W}$. We estimate the reduced Choi state or, equivalently, the reduced channel $\Lambda^{(W)}$ described by the process matrix $\chi^{(W)}$ using process classical shadows under local Pauli-basis randomization [16,17] (see Appendix C 1 for the explicit single-shot estimators and the reconstruction formula).

In practice, it is only possible to determine the reduced channels of small subsystems as the sample complexity and the computational complexity of the tomography increase exponentially with the size of W . Our goal is to nevertheless reconstruct the global process matrix χ by gluing together the locally reconstructed windows. In general, however, local marginals need not have a unique global extension (the quantum state/channel marginal problem [52–56]) so an additional assumption is necessary. We will show below that under the following assumption that the CMI decays exponentially with the size of the conditioning system, it is possible to efficiently reconstruct the global process matrix.

To formally state the assumption, we introduce further notation (see Appendix A for details). Given a state ρ supported on three disjoint subsystems A , B , and C , the CMI between A and C conditioned on B is defined as $I_\rho(A : C|B) = S_\rho(AB) + S_\rho(BC) - S_\rho(B) - S_\rho(ABC)$, where $S_\rho(A)$ is the von Neumann entropy of the reduced state $\rho(A)$ on subsystem A . When $\rho = J(\Lambda)$ is a Choi state and the subsystems are specified in terms of the physical sites, it is understood that both the “in” and “out” parts are included for each site. Our key assumption may thus be stated as follows.

Definition 1. (Exponentially decaying CMI). Consider three subsystems $A, B, C \subset \{1, 2, \dots, n\}$ with the definite ordering $i_A < i_B < i_C$ for all $i_A \in A, i_B \in B, i_C \in C$. We say that a channel Λ acting on sites $\{1, 2, \dots, n\}$ has exponentially decaying CMI if there exist constants a and ξ such that for any such subsystems the Choi state $J(\Lambda)$ satisfies

$$I_{J(\Lambda)}(A : C|B) \leq a e^{-|B|/\xi}. \quad (4)$$

III. LEARNING PROTOCOL

States satisfying the exponentially decaying CMI condition are commonly referred to as approximate Markov states [57], in reference to exact Markov states for which the CMI vanishes. Given an exact Markov state ρ for which $I_\rho(A : C|B) = 0$, there exists a so-called Petz recovery map $\mathcal{R}_{B \rightarrow BC}$ acting on subsystem B and extending it to BC such that $\rho(ABC) = (\mathcal{I}_A \otimes \mathcal{R}_{B \rightarrow BC})[\rho(AB)]$ [58,59]. Crucially, the Petz recovery map can be expressed in terms of the reduced state $\rho(BC)$ and does not require knowledge about subsystem A . A similar map exists for approximate Markov

states but the recovery is no longer exact, incurring an error proportional to $\sqrt{I_\rho(A : C|B)}$ [60,61].

Our strategy is to use such recovery maps to reconstruct the Choi states describing the global channel Λ . We proceed sequentially, adding one site at each step. Concretely, given the reduced Choi state supported on sites $\{1, \dots, i\}$, we apply a recovery map $\mathcal{R}_i$ to extend the support to $\{1, \dots, i+1\}$. The recovery map is assumed to only act on sites $\{i-w+1, \dots, i\}$. We refer to w as the reconstruction width. In the notation of the previous paragraph, the subsystems A , B , and C are given by $A = \{1, \dots, i-w\}$, $B = \{i-w+1, \dots, i\}$, and $C = \{i+1\}$. Owing to the exponentially decaying CMI, the reconstruction error can be made arbitrarily small by increasing w .

A key technical challenge is to find suitable recovery maps. If the reduced Choi state on sites $\{i-w+1, \dots, i, i+1\}$ were known exactly, we could directly compute the rotated Petz map introduced in Ref. [61], which guarantees a bounded reconstruction error. In practice, however, the reduced state can only be known approximately owing to the uncertainty of local tomography with a finite number of samples. A recent result [49] overcomes this challenge by showing that the rotated Petz map can be made stable against shot noise. Here, we pursue a different approach because the computation of the rotated Petz map is cumbersome. Moreover, despite its formal guarantees, the rotated Petz map is often far from optimal and may perform worse than even the much simpler standard Petz recovery map [62,63].

Motivated by these shortcomings, we use a simple numerical procedure to determine a locally optimal recovery map. We denote by $\hat{J}_i$ the empirical estimates of the reduced Choi state on sites $\{i-2w, \dots, i\}$ obtained from process classical tomography. We further let K_i be the reconstructed Choi state on sites $\{1, \dots, i\}$. Defining the subsystems $L_i = \{1, \dots, i-2w\}$, $A_i = \{i-2w+1, \dots, i-w\}$, $B_i = \{i-w+1, \dots, i\}$, and $C_i = \{i+1\}$, as well as the union $\bar{A}_i = L_i \cup A_i$, the locally optimal recovery map satisfies

$$\mathcal{R}_i^* = \arg \min_{\mathcal{R}_i \in \mathcal{C}(B_i \rightarrow B_i C_i)} \|\hat{J}_{i+1} - (\mathcal{I}_{\bar{A}_i} \otimes \mathcal{R}_i)[\text{tr}_{L_i}(K_i)]\|_1, \quad (5)$$

where $\mathcal{C}(B_i \rightarrow B_i C_i)$ is the set of CPTP maps from subsystem B_i to $B_i \cup C_i$. In other words, instead of constructing the recovery map analytically from a Petz-type formula, we define it variationally: $\mathcal{R}_i^*$ is the CPTP map, acting on the conditioning region B_i only, that best reproduces the tomographic window estimate $\hat{J}_{i+1}$ when applied to the current reconstruction K_i . It is computed as a convex optimization over the Choi matrix of the map, as discussed below. Having obtained $\mathcal{R}_i^*$ we iterate the procedure for $K_{i+1} = (\mathcal{I}_{\bar{A}_i} \otimes \mathcal{R}_i^*)[K_i]$,

ALGORITHM 1. Sequential construction of the global process matrix from randomized measurements.

Input: Records $(U, V, \mathbf{b})$; system size n ; reconstruction width w .
Output: Estimate $\hat{J}$ of the global Choi state $J(\Lambda)$.

- 1 Using process classical shadows, obtain an empirical estimate $\hat{J}_{2w+1}$ for the reduced Choi state on sites $\{1, \dots, 2w+1\}$.
- 2 Assign $K_{2w+1} \leftarrow$ the projection of $\hat{J}_{2w+1}$ onto the set of density operators.
- 3 **for** $i = 2w+1$ **to** $n-1$ **do**
- 4 Using process classical shadows on sites, obtain the empirical estimate $\hat{J}_{i+1}$ for the reduced Choi state on sites $\{i-2w+1, \dots, i+1\}$.
- 5 Obtain $\mathcal{R}_i^*$ according to Eq. (5).
- 6 Assign $K_{i+1} \leftarrow (\mathcal{I}_{\bar{A}_i} \otimes \mathcal{R}_i^*)[K_i]$.
- 7 **return** $\hat{J} \leftarrow K_n$.

leading to Algorithm 1. The output of the algorithm can be represented as the sequence of channels $\mathcal{R}_{n-1}^* \circ \dots \circ \mathcal{R}_{2w+1}^*$ applied to the estimate $\hat{J}_{2w+1}$ of the initial reduced Choi state. For efficient computation, we convert this representation into an MPO representation of the Choi. Since each $\mathcal{R}_i^*$ acts on w sites, the bond dimension of the MPO depends exponentially on w . For notational simplicity, our proof initializes the reconstruction at the left edge of the chain. This choice is not fundamental: starting from an interior window, one may apply the one-sided construction separately toward the left and right, with the two directional errors bounded separately and then combined by the triangle inequality. The position of the initial window may therefore be treated as a practical hyperparameter, for example, by selecting it using held-out marginal errors. We do not optimize this choice in the present numerical study.

The Choi-Jamiołkowski isomorphism implies that the set $\mathcal{C}(B_i \rightarrow B_i C_i)$ is convex as the corresponding set of Choi states is convex. Since the cost function in Eq. (5) is also convex, finding $\mathcal{R}_i^*$ constitutes a convex optimization problem, which can be solved efficiently numerically. We show in the next section that Algorithm 1 yields an estimate of the global Choi state with a sample complexity that depends polynomially on the desired 1-norm error and the system size. In practice, one is limited to relatively small values of w since each optimization involves $2w+1$ sites in the doubled Hilbert space. Nevertheless, we show in Sec. IV that the algorithm can perform well even in regimes where the rigorous bounds are not applicable.

We briefly comment on the physicality of the output of Algorithm 1. Because the initial window is projected onto the set of density operators and every subsequent recovery map is CPTP, the final estimate satisfies $\hat{J} \geq 0$ and $\text{Tr} \hat{J} = 1$. It is therefore always a valid quantum state. However, the trace-preserving Choi constraint, $\text{Tr}_{\text{out}} \hat{J} = \mathbb{I}_{\text{in}}/2^n$, is not enforced exactly. Since the trace norm is contractive under the partial trace, Theorem 1 implies $\|\text{Tr}_{\text{out}} \hat{J} - \mathbb{I}_{\text{in}}/2^n\|_1 \leq \|\hat{J} - J(\Lambda)\|_1 \leq \varepsilon$. Thus, the violation of the trace-preserving constraint is at most the global reconstruction error. Linear trace-preserving constraints could be added to the local convex optimizations

without destroying convexity. While the residual loss of the optimization can be used to certify a small global reconstruction error, it remains an open problem whether the reconstruction error can be bounded *a priori* in the presence of this constraint.

We are now able to state our main technical results: A trace norm bound for the Choi state reconstructed in Algorithm 1, and the sample complexity required to achieve a desired distance. We only sketch proofs and defer technical details and intermediate lemmas to Appendixes B and C. Our first theorem establishes that under the assumption of exponentially decaying CMI, Algorithm 1 can achieve an arbitrarily small error in the trace norm of the Choi state if the reconstruction width is sufficiently large and the reduced Choi states are estimated with sufficiently small error.

Theorem 1. (Trace norm bound for the Choi state). Let Λ be an n -qubit channel with exponentially decaying CMI with parameters a and ξ according to Definition 1. Denote by $J(\Lambda)$ the corresponding Choi state and let $\hat{J}$ be the output of Algorithm 1. Assume that the estimates $\hat{J}_i$ of the reduced Choi state on subsystems $X_i = \{i-2w, \dots, i\}$ computed in Algorithm 1 satisfy

$$\frac{1}{2} \|\hat{J}_i - \text{tr}_{[n] \setminus X_i}[J(\Lambda)]\|_1 \leq \frac{a}{2} e^{-w/(2\xi)} \quad (6)$$

for all $i \in \{2w+1, \dots, n\}$. There exist constants $b, n_0 > 0$ such that if $n \geq n_0$ and

$$w \geq 4\xi \ln \left[\frac{bn}{\varepsilon} \ln \left(\frac{bn}{\varepsilon} \right) \right], \quad (7)$$

then

$$\|\hat{J} - J(\Lambda)\|_1 \leq \varepsilon \quad (8)$$

for any target accuracy $\varepsilon \in (0, 1)$.

The proof follows from Theorem 9 in Appendix B applied to the special case of Choi states, where each site contains both “in” and “out” parts of the doubled Hilbert space. To establish Theorem 9, we show that if the distance

between the local marginals of two states is small, then their global trace distance is controlled by two quantities: the local mismatches and the amount of long-range correlations as quantified by the CMI. This step combines the quantum Pinsker inequality [64] relating the trace distance to the so-called entropy concavity gap with continuity bounds for entropic quantities, yielding Lemma 6. We next show using the data processing inequality that the CMI of the state generated by Algorithm 1 is bounded by the CMI of the target state up to an error that depends on the local mismatch of the reconstructed state (Lemma 7). Finally, we control the local mismatch by comparing the algorithm's recovery map to a (rotated) Petz map, whose reconstruction error is governed by the approximate Markov property of the true state. This yields a recursion showing that the accumulated local mismatch is driven by the local tomography error combined with the exponentially small CMI (Proposition 8).

Using Theorem 1, we may bound the sample complexity for estimating the Choi state with a small error in trace distance. Concretely, we employ a process shadow tomography protocol of Ref. [17] and bound the number of samples required to satisfy Eq. (6). The condition can be satisfied for all subsystems X_i simultaneously with high probability, resulting in the following corollary.

Corollary 2. (End-to-end sample complexity). Let Λ be an n -qubit channel with exponentially decaying CMI with parameters a and ξ according to Definition 1. Denote by $J(\Lambda)$ the corresponding Choi state and let $\hat{J}$ be the output of Algorithm 1 with M records $(U, V, \mathbf{b})$ as an input. Given $\varepsilon, \delta \in (0, 1)$, there exists a choice of w such that $\hat{J}$ satisfies $\|\hat{J} - J(\Lambda)\|_1 \leq \varepsilon$ with probability at least $1 - \delta$ provided

$$M \geq c_1 \left(\frac{n}{\varepsilon}\right)^{c_2 \xi + c_3} f(n, \varepsilon, \delta, \xi), \quad (9)$$

where c_1, c_2 , and c_3 are constants and $f(n, \varepsilon, \delta, \xi)$ is a function that depends linearly on ξ and polylogarithmically on n, ε , and δ .

The proof is given in Appendix C 2, which includes detailed expressions for the constants $c_{1,2,3}$ and the function f .

The explicit dependence on the CMI decay length ξ in Eq. (9) delineates the regime of applicability of our method. For any fixed ξ , the sample complexity is polynomial in n and $1/\varepsilon$, but the degree of the polynomial grows linearly with ξ . If the CMI of the Choi state decays slowly or over long distances, the required reconstruction width w grows accordingly. Since the optimizations determining the recovery maps involve windows of $2w + 1$ sites, both the cost of the convex optimization and the sample complexity grow exponentially with w , rendering the protocol impractical. We note that our results are stated in terms of the trace norm error of the Choi state. By

contrast, the diamond norm $\|\Lambda - \Lambda'\|_\diamond$ has an operational meaning as the worst-case distinguishability of two channels. In general, the two are related by an inequality of the form $\|\Lambda - \Lambda'\|_\diamond \leq d^n \|J(\Lambda) - J(\Lambda')\|_1$, whose dimension-dependent prefactor can be loose. For structured noise models this conversion can be significantly tighter. In particular, for Pauli channels (i.e., channels whose process matrices are diagonal in the Pauli basis), we can control the diamond norm with trace norm error, which removes the dimension-dependent prefactor. We provide details in Appendix D.

IV. APPLICATIONS

A. Learning open-system dynamics

We now apply our reconstruction scheme to a concrete model of open quantum system dynamics. We consider an open Heisenberg spin chain with local anisotropic noise. Writing X_i, Y_i, Z_i for the Pauli operators on the i th system qubit, the reduced system dynamics are governed by the Lindblad master equation [65]

$$\begin{aligned} \dot{\rho}_t = & -i[H, \rho_t] \\ & + \sum_{i=1}^n \sum_{\alpha \in \{x,y,z\}} \left(L_i^{(\alpha)} \rho_t L_i^{(\alpha)\dagger} - \frac{1}{2} \{L_i^{(\alpha)\dagger} L_i^{(\alpha)}, \rho_t\} \right), \end{aligned} \quad (10)$$

with Hamiltonian

$$H = \frac{1}{2} \sum_{i=1}^{n-1} (X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1}), \quad (11)$$

and local noise with jump operators $L_i^{(\alpha)} = \sqrt{\gamma_\alpha} \sigma_i^{(\alpha)}$, where $(\sigma_i^{(x)}, \sigma_i^{(y)}, \sigma_i^{(z)}) = (X_i, Y_i, Z_i)$, with rates $\gamma_x = \gamma_y = 1/2$ and $\gamma_z = 1$. These dynamics generate a family of CPTP maps Λ_t such that $\Lambda_t(\rho_0) = \rho_t$, and their associated Choi states $J(\Lambda_t)$.

We emphasize that although we use Lindblad spin-chain dynamics as a controlled and physically motivated testbed, our reconstruction protocol does not assume a Lindbladian generator or Markovianity: it recovers the channel Λ_t directly from randomized measurements rather than identifying the underlying Hamiltonian and jump operators as in Lindbladian tomography. This distinction is practically relevant because in realistic devices the effective noise can be time-dependent, non-Markovian, or only approximately describable by a local master equation, whereas our method only relies on the locality structure of the induced Choi state. We also note that a general characterization of how the Choi state conditional mutual information grows under generic many-body Lindblad evolutions is still lacking. We therefore treat decaying CMI as a plausible hypothesis, which is justified by the accuracy and stability

of the reconstruction in the short-time regime considered below. In addition, we provide direct numerical evidence in Appendix E (Fig. 4). There, we evaluate the Choi-state CMI for the Lindbladian example we consider here. We observe that the CMI is bounded by an exponentially decaying envelope in the size of the conditioning region, with a decay length that grows with the evolution time, which is consistent with the picture that correlations spread under the dynamics.

To simulate $J(\Lambda_t)$ we work on a doubled chain, where each super-site consists of a system qubit and a reference qubit. We initialize a product of Bell pairs $|\omega\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ between system and reference on each super-site. All Hamiltonian and Lindblad operators act as $\mathbb{I} \otimes A$ on system and reference, so that the resulting state on the doubled chain at time t is precisely the Choi state $J(\Lambda_t)$ of the open-system channel on the system qubits. We present the $J(\Lambda_t)$ as an MPO and simulate the open-system dynamics using a TEBD-style tensor-network time-evolution method [66,67]. All tensor-network contractions and MPO manipulations are implemented using QUIMB [68].

From the global Choi state MPO we extract reduced marginals on contiguous windows along the doubled chain. We interpret the synthetic reduced Choi state as $\hat{J}_i$ on sites $\{i - 2w, \dots, i\}$. In the numerics below we take $w = 1$. In the noiseless setting, $\hat{J}_i$ is taken to be the exact reduced marginal extracted from the MPO. In the noisy setting, we consider a phenomenological tomography error model where we perturb the corresponding local Pauli expansion coefficients of Choi state by additive Gaussian noise. More specifically, for each coefficient c_α corresponding to the Pauli operator P_α , we sample independent noise $\xi_\alpha \sim \mathcal{N}(0, \sigma^2)$ and set $\tilde{c}_\alpha = c_\alpha + \xi_\alpha$. Then, we convert these noisy coefficients into an estimate of the local Choi state $\hat{J}_i$, projecting onto physical states by enforcing hermiticity, setting negative eigenvalues to zero, and renormalizing to unit trace. We quantify the local estimation error by $\eta_i = \frac{1}{2} \|\hat{J}_i - J_i\|_1$ and define $\eta = \max_i \eta_i$ as the maximum window error, which we use as a summary for the local noise level.

Given the collection of window estimates $\{\hat{J}_i\}$, we perform the sequential left-to-right reconstruction of Algorithm 1. Denoting by K_i the current reconstructed Choi state on sites $\{1, \dots, i\}$, we compute at each step a locally optimized recovery map $\mathcal{R}_i^* \in \mathcal{C}(B_i \rightarrow B_i C_i)$ by fitting the predicted 3-site window marginal $(\mathcal{I}_{A_i} \otimes \mathcal{R}_i)[\text{Tr}_{L_i}(K_i)]$ to the target window marginal $\hat{J}_{i+1}$, as in Eq. (5). In our numerical implementation, we optimize over the Choi matrix of $\mathcal{R}_i$. Instead of considering the trace norm as objective in Eq. (5), we use a Frobenius-norm objective as a practical surrogate, since it yields a smooth, inexpensive objective amenable to efficient solvers, leading to faster and stable optimization compared with trace distance in practice [69]. Since $\|X\|_1 \leq \sqrt{d_{\text{win}}} \|X\|_F$ on

each $(2w + 1)$ -site window, the resulting trace norm residual is within a factor 2^{2w+1} of the trace distance optimum. The resulting map $\mathcal{R}_i^*$ is then used to update the global estimate via $K_{i+1} = (\mathcal{I}_{A_i} \otimes \mathcal{R}_i^*)[K_i]$. The convex optimization is carried out using CVXPY [70]. Further details of the numerical implementation are given in Appendix F.

We use the diagonal entries $\chi_{\alpha\alpha}$ of the process matrix as a convenient diagnostic of how weight spreads under the channel. Figure 2 summarizes the reconstruction performance for this open-system dynamics. Panels (a) and (b) show that in the absence of shot noise, the reconstructed channel matches the exact channel in the Pauli-weight-resolved sums of diagonal Pauli-basis process matrix entries. The relative error Δ_R grows smoothly with the evolution time, from below 10^{-3} at $t = 10^{-2}$ to a few times 10^{-3} at $t = 10^{-1}$, reflecting the intrinsic accuracy of the reconstruction at fixed width $w = 1$. This growth indicates the buildup of correlations in the Choi state under the dynamics, in line with the growth of the CMI decay length observed in Appendix E. Panels (c) and (d) probe how the reconstruction error depends on the error of local window

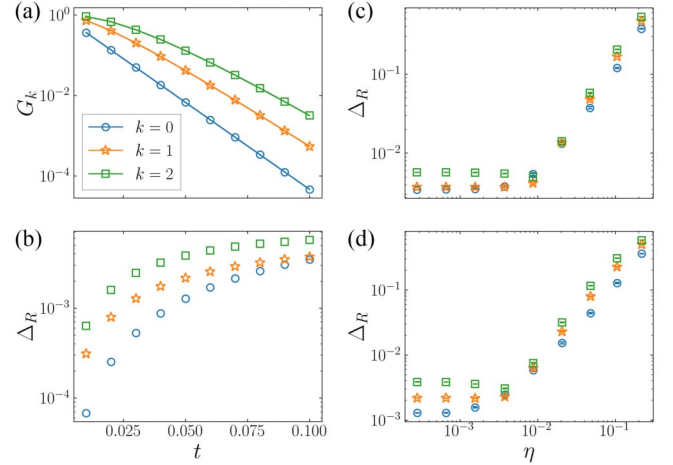


FIG. 2. Reconstruction of Pauli-weight-resolved diagonal process matrix elements in an open Heisenberg chain with local anisotropic noise ($n = 50$ qubits). (a) For Pauli weights $k = 0, 1, 2$, the cumulative quantity $G_k = \sum_{|\alpha| \leq k} |\chi_{\alpha\alpha}|$, where $|\alpha|$ denotes the Pauli weight of α , of the exact channel (solid lines) agrees well with that of the reconstructed channel obtained from our optimization procedure (markers). (b) The corresponding relative reconstruction error $\Delta_R = \sum_{|\alpha| \leq k} |\chi_{\alpha\alpha}^R - \chi_{\alpha\alpha}^E| / \sum_{|\alpha| \leq k} |\chi_{\alpha\alpha}^E|$, where the superscripts E and R refer to the exact and the reconstructed channel, respectively. Panels (a),(b) assume perfect local tomography without noise. (c),(d) Relative reconstruction error Δ_R in the presence of local noise, which is quantified by the error η in the reduced Choi states, at fixed evolution times $t = 10^{-1}$ (c) and $t = 5 \times 10^{-2}$ (d). Each data point is the mean over 10 independent noise realizations. The error bars denote the standard error of the mean and are often smaller than the marker size.

estimations. We introduce local estimation errors by perturbing the Pauli expansion coefficients of each local marginal as described above, and summarize the resulting local noise level by η . We then run the sequential reconstruction. For each noise level, we repeat the sampling of the noise and the full reconstruction of the $n = 50$ chain for 10 independent realizations, and report the mean of Δ_R together with its standard error. Figure 2(c) shows the resulting global error at fixed evolution time $t = 10^{-1}$, while (d) shows the same dependence at $t = 5 \times 10^{-2}$. In both cases, the global reconstruction error decreases as η is reduced, before saturating at small η at a plateau that coincides with the intrinsic, noise-free reconstruction error of panel (b) at the corresponding evolution time. Consistent with this, the plateau is lower at the earlier time. The global error is thus controlled by the local tomographic noise until the latter drops below the intrinsic accuracy of the stitching procedure. These results indicate that the local-to-global stitching procedure is stable to moderate local estimation errors for the $n = 50$ chain studied here.

B. Learning noisy circuits

To illustrate the applicability of our method to shallow noisy circuits, we consider channels of the form [see Fig. 3(a)]

$$\Lambda_{U,\epsilon,\gamma}(\rho) = \left(\bigotimes_{j=1}^n \mathcal{D}_\gamma^{(j)} \right) \circ U_\epsilon \circ U(\rho),$$

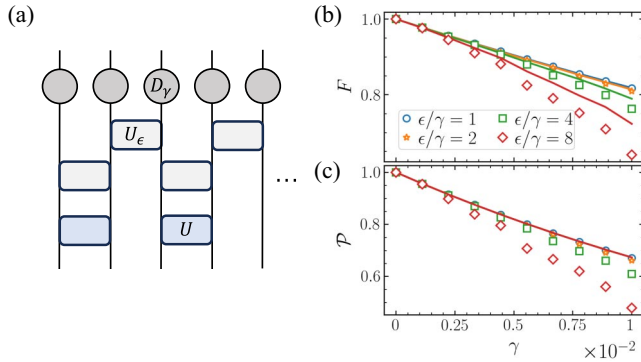


FIG. 3. Global reconstruction of a noisy shallow circuit with $n = 20$ qubits. (a) Structure of the channel $\Lambda_{U,\epsilon,\gamma} = (\bigotimes_j \mathcal{D}_\gamma^{(j)}) \circ U_\epsilon \circ U$, where U is a single layer of nearest-neighbor two-qubit gates, U_ϵ denotes a coherent perturbation, and $\mathcal{D}_\gamma$ is a single-qubit dephasing channel. (b) Process fidelity with respect to the ideal unitary circuit U , showing $F(\Lambda_{U,\epsilon,\gamma}, \Lambda_U)$ (solid lines) for the exact noisy channel and $F(\hat{\Lambda}_{U,\epsilon,\gamma}, \Lambda_U)$ (markers) for the optimization-based reconstruction, as a function of γ . (c) Purity of the Choi state $\mathcal{P}(J(\Lambda_{U,\epsilon,\gamma}))$ (solid lines) together with the purity of the reconstructed channel $\mathcal{P}(J(\hat{\Lambda}_{U,\epsilon,\gamma}))$ (markers), as a function of γ .

where U is a single layer of nearest-neighbor two-qubit gates acting on even bonds, $U = \bigotimes_{i \text{ even}} U_{i,i+1}$, and U_ϵ denotes a coherent-noise layer acting on even and odd bonds, $U_\epsilon = \bigotimes_{i \text{ odd}} U_{i,i+1}(\epsilon) \circ \bigotimes_{i \text{ even}} U_{i,i+1}(\epsilon)$, which models an undesired coherent coupling. Here $\mathcal{D}_\gamma^{(j)}$ is a single-qubit dephasing channel in the Z basis with strength γ on site j , e.g., $\mathcal{D}_\gamma^{(j)}(\rho) = (1 - \gamma)\rho + \gamma Z_j \rho Z_j$. For $\gamma = \epsilon = 0$, the channel reduces to a depth-one circuit of disjoint two-qubit unitaries on even bonds. Viewing the Choi state $J(\Lambda)$ as an n -site chain by grouping the input-output degrees of freedom of each qubit into a single site, $J(\Lambda)$ then factorizes into independent two-site blocks. Consequently, for any contiguous tripartition $A - B - C$ with $|B| \geq 1$, we have $I_{J(\Lambda)}(A : C|B) = 0$. In the weak-noise, shallow-depth regime, we therefore expect the Choi states generated by $\Lambda_{U,\epsilon,\gamma}$ to exhibit an analogous approximate Markov property.

We test this expectation in Appendix E (Fig. 5). We find that the maximum CMI decays rapidly with the width b of the conditioning region and reaches the numerical floor for $b \geq 3$, beyond the causal range of the circuit. At fixed b , the CMI scales approximately as $\epsilon^{1.8}$ with the coherent perturbation strength and rapidly saturates as the sizes of the outer regions A and C are increased. These results show that the coherent crosstalk generates only weak, short-ranged conditional correlations, supporting the exponentially decaying CMI assumption and helping to explain the good reconstruction accuracy already obtained with $w = 1$ in Fig. 3. We emphasize, however, that this expectation is restricted to sufficiently shallow depths. In noisy random circuits the CMI can spread superlinearly and become long-ranged at larger depths [71]. Our numerical results for noisy random Clifford circuits in Appendix E quantify this limitation within the parameter regime studied. As shown in Fig. 6, the Choi-state CMI remains well described by an exponential decay with the width of the conditioning region for all circuit depths and noise strengths considered. The decay persists when the outer regions A and C are extended to cover the entire chain (Fig. 7). However, the fitted decay length increases with both depth and noise strength, implying that progressively larger reconstruction windows are required as the noise rate and circuit depth increase.

When $\epsilon > 0$, the coherent crosstalk, in combination with the local dephasing, induces spatially correlated noise in the Choi state and increases the CMI. For $\epsilon = 0$ the Choi state factorizes for any γ and the CMI vanishes (see Appendix E). For each choice of (U, ϵ, γ) we compute the corresponding Choi state $J(\Lambda_{U,\epsilon,\gamma})$ and generate synthetic local data according to the access model of Sec. II. Here we consider $U = \bigotimes_{i \text{ even}} e^{i\frac{\pi}{4}(X_i X_{i+1} + Y_i Y_{i+1})}$, which corresponds to an iSWAP gate. We take $U_\epsilon = \bigotimes_{i \text{ odd}} e^{i\epsilon H_i} \circ \bigotimes_{i \text{ even}} e^{i\epsilon H_i}$, where $H_i = \alpha_1 X_i X_{i+1} + \alpha_2 Y_i Y_{i+1} + \alpha_3 Z_i Z_{i+1}$. We set $\epsilon = r\gamma$ with $r \in \{1, 2, 4, 8\}$.

For each data point, the coefficients $\alpha_1, \alpha_2, \alpha_3$ of each U_ϵ are drawn independently from a uniform distribution on $[0, 1)$. We vary the noise amplitude γ from 0 to 10^{-2} .

Figure 3 shows reconstruction performance for $n = 20$ qubits. To benchmark the reconstruction, we compute the fidelity between the Choi state of the noisy channel and that of the ideal unitary channel. Since the Choi state of the unitary channel is pure, the fidelity reduces to a simple overlap [72] $F(\Lambda_{U,\epsilon,\gamma}, \Lambda_U) = \text{Tr}[J(\Lambda_{U,\epsilon,\gamma})J(\Lambda_U)]$. The fidelity bounds the trace distance according to the Fuchs-van de Graaf inequalities $1 - \sqrt{F(\Lambda_{U,\epsilon,\gamma}, \Lambda_U)} \leq \frac{1}{2}\|J(\Lambda_{U,\epsilon,\gamma}) - J(\Lambda_U)\|_1 \leq \sqrt{1 - F(\Lambda_{U,\epsilon,\gamma}, \Lambda_U)}$. Figure 3(b) reports this fidelity as a function of γ . The reconstructed channel closely tracks the exact channel, with deviations that grow with the strength of the coherent perturbation and remain small for $\epsilon \leq 4\gamma$.

We also report the Choi state purity $\mathcal{P}(J(\Lambda_{U,\epsilon,\gamma})) = \text{Tr}[J(\Lambda_{U,\epsilon,\gamma})^2]$ as a simple global diagnostic of unitarity of the quantum process [73]: the amount of coherent versus incoherent noise in the channel.

Figure 3(c) shows that the purity computed from the reconstructed Choi state reproduces the exact values well for moderate coherent perturbations, whereas for the strongest perturbation, $\epsilon = 8\gamma$, the reconstruction underestimates the purity at the largest noise rates. This trend is consistent with the growth of the Choi-state CMI with ϵ quantified in Appendix E, where the reconstruction becomes less accurate as the CMI increases. For fixed γ , varying ϵ only modifies the coherent unitary part of the channel, which induces a unitary change of basis of the Choi state. Since the purity is invariant under unitary conjugation, $\mathcal{P}(J(\Lambda_{U,\epsilon,\gamma}))$ is independent of ϵ and depends only on γ . Together, these results demonstrate that our local-to-global learning scheme can reliably reconstruct physically relevant global functionals of noisy shallow circuit layers from local data.

V. DISCUSSION AND OUTLOOK

We introduced a local-to-global reconstruction framework for quantum channels in one dimension under the physically motivated assumption of exponentially decaying CMI of the Choi state. By combining scalable process classical shadows on small windows with a sequential recovery procedure, we obtain an explicit global estimate $\hat{J}$ and the corresponding process matrix $\hat{\chi}$. This representation supports efficient evaluation of both local and genuinely nonlocal quantities, including purity- and entropy-like functionals as well as selected process matrix elements. Numerically, we observed that the method can remain effective even when the rigorous worst-case regime is not strictly satisfied, enabling reconstructions for channels acting on up to 50 qubits from reduced windows of size 3 in our examples. While we are able to rigorously bound the 1-norm error in the Choi state, it remains an

open question whether tight bounds on the diamond norm of the channel can be obtained using a similar approach.

Several directions are particularly promising for improving practical performance and broadening the scope of the approach. A natural direction is to combine our local-to-global stitching framework with heuristic tensor-network learning algorithms [43,44]. These works parameterize the global Choi matrix by a tensor-network ansatz and fit its parameters via data-driven, generally non-convex optimization over an informationally complete measurement dataset. Inspired by this, we envision a hybrid strategy: one could learn a low-complexity tensor-network model restricted to each local window directly from the randomized measurement records, and then feed the learned window marginals into our recovery-map procedure to produce a globally consistent MPO estimate of the full Choi state. Such a hybrid approach could improve sample efficiency compared to worst-case window tomography while avoiding the need to train a single global tensor-network model directly.

Our access model assumes ideal local randomization gates and measurements. In experiments, state-preparation-and-measurement (SPAM) errors can bias the window estimates $\hat{J}_i$ and thereby propagate through the recovery procedure. An attractive feature of our approach is that SPAM errors are often largely local. Hence, if SPAM errors are well modeled by shallow local noise, they may be absorbed into the same locality structure exploited by the recovery maps. A careful empirical study benchmarking SPAM-mitigation strategies within our pipeline, especially in regimes where SPAM dominates over intrinsic channel noise, would help identify the most robust experimental pathway.

Our guarantees rely on exponential decay of CMI on the Choi state, which is a strong but physically motivated locality condition. We expect it to hold for channels generated by shallow circuits with local interactions and sufficiently weak noise, where information does not propagate far and long-range correlations remain suppressed. Conversely, for deep circuits, strongly correlated noise, or channels that generate long-range entanglement on the doubled system, the CMI tail may decay more slowly. In such regimes larger reconstruction widths w may be necessary, leading to larger bond dimensions and higher sample requirements. For practical deployments, it is therefore useful to develop diagnostics, e.g., estimating empirical CMIs on triples $A - B - C$ of windowed subsystems to probe whether $I(A : C|B)$ decreases as $|B|$ grows. We remark that such diagnostics use the same randomized-measurement records as the reconstruction, which allows one to estimate the CMI on windows of up to $2w + 1$ sites and to monitor its decay with $|B|$ (see Appendix E for benchmarks of this decay).

Two complementary consistency checks can be performed without introducing additional measurement

settings. One is cross-validation, in which the reconstructed state is used to predict the marginals of windows that were held out from the stitching procedure, and the other is monitoring the convergence of reconstructed global quantities as w is increased. A failure of either check signals that the required reconstruction width exceeds the accessible window size, in which case the output of the protocol should not be trusted. We note that these checks provide physical reassurance rather than a certificate. A mixed Choi state may contain system-spanning coherent components (e.g., GHZ-type coherences) that are invisible to every proper local marginal. However, this obstruction is not expected to be generic in the regime we discussed in our application section. Understanding typical decay lengths for relevant device noise models remains an important direction.

Extending the framework to higher-dimensional geometries is another important goal for near-term devices. The main challenge is that representing and manipulating 2D tensor-network objects (e.g., projected entangled-pair operators and states, PEPO/PEPS) typically requires approximate contraction schemes, and the computational cost can grow quickly with bond dimension and separator width. Developing principled patch-based or separator-based variants of our recovery approach in 2D, together with efficient approximate contraction and stability guarantees under CMI-like locality assumptions, is an exciting and challenging direction for future work.

ACKNOWLEDGMENTS

We thank J. Ignacio Cirac for insightful comments and feedback on the manuscript. We further thank Álvaro Alhambra, Flavio Baccari, Marianna Crupi, Giacomo Giudice, Daniel Malz, Daniel Stilck França, Lorenzo Piroli, and Yijian Zou for fruitful discussions. We acknowledge support from the German Federal Ministry of Education and Research (BMBF) through the funded project ALMANAQC, Grant No. 13N17236 within the research program “Quantum Systems.”

DATA AND CODE AVAILABILITY

The source code used to produce the numerical results of this work, comprising the implementation of the reconstruction algorithm, the tensor-network simulations of Sec. IV, and the evaluation of the Choi-state CMI of Appendix E is openly available on GitHub and Zenodo [74].

APPENDIX A: DEFINITIONS AND NOTATION

Given a Hilbert space $\mathcal{H}$, we denote by $\mathcal{B}(\mathcal{H})$ the space of bounded linear operators acting on $\mathcal{H}$. We define $\mathcal{D}(\mathcal{H}) := \{\rho \in \mathcal{B}(\mathcal{H}) : \rho \geq 0, \text{tr}(\rho) = 1\}$ as the space of density operators on $\mathcal{H}$. We refer to a map $\mathcal{R} : \mathcal{B}(\mathcal{H}_1) \rightarrow \mathcal{B}(\mathcal{H}_2)$ as a superoperator from $\mathcal{H}_1$ to $\mathcal{H}_2$

and denote the space of all such superoperators by $\mathcal{S}(\mathcal{H}_1, \mathcal{H}_2)$.

We consider a system composed of n sites, which we label using the set of integers $[n] = \{1, 2, \dots, n\}$. Each site is described by a local Hilbert space $\mathbb{C}^d$ and we denote the full Hilbert space of the system by $\mathcal{H}^{(n)} = (\mathbb{C}^d)^{\otimes n}$. We will often deal with subsystems, which we specify in terms of the indices $A \subseteq [n]$ of the included sites. We denote the corresponding Hilbert space by $\mathcal{H}^{(n)}(A)$. Similarly, for any $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$, the state $\rho(A) \in \mathcal{D}(\mathcal{H}^{(n)}(A))$ denotes the reduced density operator on subsystem A , i.e., $\rho(A) = \text{tr}_{[n] \setminus A}(\rho)$. For multiple subsystems, we introduce the shorthands $AB = A \cup B$, $ABC = A \cup B \cup C$, etc. In a slight abuse of notation, we will sometimes use a string of site labels to directly specify the subsystem. For example, the state $\rho(i_1, i_2, \dots, i_k) \in \mathcal{D}(\mathcal{H}^{(n)}(i_1, i_2, \dots, i_k))$ should be understood as $\rho(A) \in \mathcal{D}(\mathcal{H}^{(n)}(A))$, where $A = \{i_1, i_2, \dots, i_k\}$. To avoid confusion, we use lower case letters for site labels and upper case letters for subsystems.

Several subsystems occur frequently throughout the proofs. Given an integer $w \geq 1$, we let $A_i = \{i - 2w + 1, \dots, i - w\}$, $B_i = \{i - w + 1, \dots, i\}$, and $C_i = \{i + 1\}$. Moreover, we define the subsystems making up the left and right edges, $L_i = \{1, \dots, i - 2w\}$ and $R_i = \{i + 2, \dots, n\}$, as well as the respective unions $\bar{A}_i = L_i \cup A_i$ and $\bar{C}_i = C_i \cup R_i$. We additionally use the shorthand $X_i = \{i - 2w, \dots, i\} = A_{i-1} \cup B_{i-1} \cup C_{i-1}$. We employ this notation consistently throughout, but will nevertheless include the relevant definitions in all formal statements for convenience.

For functions of states such as entropies and the CMI, we use subscripts to indicate on which state the function is applied and brackets to specify the relevant subsystems. For example, $S_\rho(A) = -\text{tr}[\rho(A) \log \rho(A)]$ denotes the von Neumann entropy of the reduced state $\rho(A)$. Similarly, $S_\rho(A|B) = S_\rho(AB) - S_\rho(B)$ is the conditional entropy and $I_\rho(A : C|B) = S_\rho(AB) + S_\rho(BC) - S_\rho(ABC) - S_\rho(B)$ is the CMI. Again, we sometimes use the site labels directly. For instance, $I_\rho(1, \dots, i - w : i + 1, \dots, n | i - w + 1, \dots, i)$ is equivalent to $I_\rho(\bar{A}_i : \bar{C}_i | B_i)$.

Using the above notation, we formalize the notion of states with exponentially decaying CMI as follows.

Definition 2. (Exponentially decaying CMI). Consider three subsystems $A, B, C \subset [n]$ with the definite ordering $i_A < i_B < i_C$ for all $i_A \in A$, $i_B \in B$, $i_C \in C$. We say that a state $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$ has exponentially decaying CMI if there exist constants a and ξ such that for any such subsystems

$$I_\rho(A : C|B) \leq a e^{-|B|/\xi}. \quad (\text{A1})$$

The Markov entropy [75,76] defined as follows serves as a convenient technical tool.

Definition 3. (Markov entropy). Given a quantum state $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$, we define the multipartite Markov entropy of

shielding width $w \geq 1$ as

$$S_\rho^M := \sum_{i=w+1}^n S_\rho(i-w|B_i) + S_\rho(B_n), \quad (\text{A2})$$

where $B_i = \{i-w+1, \dots, i\}$.

Markov entropy has the important property that it equals the von Neumann entropy when ρ is an exact Markov state with decay length less than w , as evident from the following lemma.

Lemma 3. Given a quantum state $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$, the Markov entropy of shielding width w satisfies

$$S_\rho^M = S_\rho + \sum_{i=w+1}^{n-1} I_\rho(i-w: \bar{C}_i|B_i), \quad (\text{A3})$$

where $B_i = \{i-w+1, \dots, i\}$ and $\bar{C}_i = \{i+1, \dots, n\}$.

Proof. By the definition of the CMI,

$$\sum_{i=w+1}^{n-1} I_\rho(i-w: \bar{C}_i|B_i) = \sum_{i=w+1}^{n-1} [S_\rho(i-w|B_i) - S_\rho(i-w|B_i \cup \bar{C}_i)]. \quad (\text{A4})$$

The second term is a telescoping sum that simplifies to

$$\sum_{i=w+1}^{n-1} S_\rho(i-w|B_i \cup \bar{C}_i) = \sum_{i=w+1}^{n-1} [S_\rho(i-w, \dots, n) - S_\rho(i-w+1, \dots, n)] \quad (\text{A5})$$

$$= S_\rho - S_\rho(n-w, \dots, n). \quad (\text{A6})$$

By observing that $S_\rho(n-w, \dots, n) = S_\rho(n-w|B_n) + S_\rho(B_n)$, we obtain the desired result

$$\sum_{i=w+1}^{n-1} I_\rho(i-w: \bar{C}_i|B_i) = \sum_{i=w+1}^n S_\rho(i-w|B_i) + S_\rho(B_n) - S_\rho = S_\rho^M - S_\rho. \quad (\text{A7})$$

Our reconstruction algorithm is based on sequentially adding sites using recovery maps that only act on a small region. Concretely, when applied to a state defined on the sites $1, 2, \dots, i$, the CPTP map $\mathcal{R}_i$ adds the site $i+1$. We assume throughout that $\mathcal{R}_i$ only acts on the region of width w preceding i , i.e., the subsystem $B_i = \{i-w+1, \dots, i\}$. This leads us to formally define sequentially generated states for the purpose of this work as follows.

Definition 4. (Sequentially generated states). Consider the Hilbert space $\mathcal{H}^{(n)}$ and the reconstruction width $w \geq 1$. Let $\bar{A}_i = \{1, 2, \dots, i-w\}$ and $B_i = \{i-w+1, i-w+2, \dots, i\}$ for $w+1 \leq i \leq n$. We say that the states $\sigma_i \in \mathcal{D}(\mathcal{H}^{(n)}(\bar{A}_i B_i))$ are sequentially generated if they satisfy $\sigma_{i+1} = (\mathcal{I}_{\bar{A}_i} \otimes \mathcal{R}_i)[\sigma_i]$ for all $w+1 \leq i \leq n-1$, where $\mathcal{I}_{\bar{A}_i} \in \mathcal{S}(\mathcal{H}^{(n)}(\bar{A}_i), \mathcal{H}^{(n)}(\bar{A}_i))$ is the identity channel and $\mathcal{R}_i \in \mathcal{S}(\mathcal{H}^{(n)}(B_i), \mathcal{H}^{(n)}(B_i \cup \{i+1\}))$ is a CPTP map.

Our approach is to choose the recovery maps $\mathcal{R}_i$ such that the distance from estimates $\hat{\rho}_i$ of local reduced density matrices is minimized at each step. We formalize this notion as follows.

Definition 5. (Locally optimal recovery). Consider the Hilbert space $\mathcal{H}^{(n)}$ and a set of sequentially generated

states $\sigma_i \in \mathcal{D}(\mathcal{H}^{(n)}(1, 2, \dots, i))$ with reconstruction width w according to Definition 4. Given an operator $\hat{\rho}_{i+1}$ (not necessarily a valid quantum state) on $\mathcal{H}^{(n)}(i-2w+1, \dots, i+1)$, we say that the recovery is locally optimal with respect to $\hat{\rho}_{i+1}$ if the map $\mathcal{R}_i$ of Definition 4 minimizes the trace distance $\|\hat{\rho}_{i+1} - (\mathcal{I}_{A_i} \otimes \mathcal{R}_i)[\text{tr}_{L_i}(\sigma_i)]\|_1$, where $A_i = \{i-2w+1, \dots, i-w\}$ and $L_i = \{1, \dots, i-2w\}$.

APPENDIX B: PROOF OF THEOREM 1

We first establish two lemmas concerning the continuity of the CMI and the Markov entropy. Both lemmas are standard results expressed in a convenient fashion for our purposes.

Lemma 4. (Continuity of the CMI). Consider two quantum states ρ and σ in a composite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ with local dimensions $\dim \mathcal{H}_A = d_A$, $\dim \mathcal{H}_B = d_B$, and $\dim \mathcal{H}_C = d_C$. If $\frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon \leq 1$, then

$$|I_\rho(A: C|B) - I_\sigma(A: C|B)| \leq 8\varepsilon \log_2(d_A d_C / \varepsilon). \quad (\text{B1})$$

Proof. The conditional entropy $S_\rho(A|B) = S_\rho(AB) - S_\rho(B)$ satisfies the continuity bound [77]

$$|S_\rho(A|B) - S_\sigma(A|B)| \leq 2\varepsilon \log_2 d_A + (1 + \varepsilon)H(\varepsilon/(1 + \varepsilon)), \quad (\text{B2})$$

where $H(x) = -x \log_2 x - (1 - x) \log_2 (1 - x)$ is the binary entropy function. Here, we used the fact that $\frac{1}{2} \|\rho(AB) - \sigma(AB)\|_1 \leq \frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon$. Using the bound $(1 + x)H(x/(1 + x)) \leq -x \log_2 (x/4)$, we can simplify the expression to

$$|S_\rho(A|B) - S_\sigma(A|B)| \leq \varepsilon \log_2 (4d_A^2/\varepsilon). \quad (\text{B3})$$

Similarly, we obtain

$$|S_\rho(C|B) - S_\sigma(C|B)| \leq \varepsilon \log_2 (4d_C^2/\varepsilon), \quad (\text{B4})$$

$$|S_\rho(AC|B) - S_\sigma(AC|B)| \leq \varepsilon \log_2 (4d_A^2 d_C^2/\varepsilon). \quad (\text{B5})$$

To complete the proof, we express the CMI in terms of the conditional entropy as

$$I_\rho(A : C|B) = S_\rho(A|B) + S_\rho(C|B) - S_\rho(AC|B). \quad (\text{B6})$$

Using the triangle inequality and the fact that $d_A \geq 2$, $d_C \geq 2$, and $\varepsilon \leq 1$, we obtain the desired result

$$\begin{aligned} |I_\rho(A : C|B) - I_\sigma(A : C|B)| &\leq \varepsilon \log_2 (64d_A^4 d_C^4/\varepsilon^3) \\ &\leq 8\varepsilon \log_2 (d_A d_C/\varepsilon). \end{aligned} \quad (\text{B7})$$

Lemma 5. (Continuity of the Markov entropy). Consider two quantum states $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$ and $\sigma \in \mathcal{D}(\mathcal{H}^{(n)})$. Given a shielding width $w \geq 1$ and assuming that the reduced density matrices satisfy $\frac{1}{2} \|\rho(i - w, \dots, i) - \sigma(i - w, \dots, i)\|_1 \leq \varepsilon \leq 1$ for all $w + 1 \leq i \leq n$, the difference in Markov entropy is bounded by

$$|S_\rho^M - S_\sigma^M| \leq 4n\varepsilon \log_2 (d/\varepsilon). \quad (\text{B8})$$

Proof. We let $B_i = \{i - w + 1, i - w + 2, \dots, i\}$. By following the same steps as in the proof of Lemma 4, we find that

$$|S_\rho(i - w|B_i) - S_\sigma(i - w|B_i)| \leq \varepsilon \log_2 (4d^2/\varepsilon). \quad (\text{B9})$$

The von Neumann entropy satisfies a similar continuity bound [77]

$$|S_\rho(B_n) - S_\sigma(B_n)| \leq w \log_2 (d)\varepsilon + H(\varepsilon), \quad (\text{B10})$$

which we simplify using $H(x) \leq -x \log_2 (x/e)$ to obtain

$$|S_\rho(B_n) - S_\sigma(B_n)| \leq \varepsilon \log_2 (ed^w/\varepsilon) \quad (\text{B11})$$

By expanding $|S_\rho^M - S_\sigma^M|$ in terms of Definition 3 and applying the triangle inequality, we obtain

$$\begin{aligned} |S_\rho^M - S_\sigma^M| &\leq \varepsilon[(n - w) \log_2 (4d^2/\varepsilon) + \log_2 (ed^w/\varepsilon)] \\ &\leq 4n\varepsilon \log_2 (d/\varepsilon) + \log_2 (e\varepsilon^{w-1}/(4d)^w), \end{aligned} \quad (\text{B12})$$

where the final inequality holds for all $d \geq 2$ and $\varepsilon \leq 1$. The second term is negative for all $w \geq 1$ and can therefore be dropped. ■

In the next step, we establish that the trace distance between two quantum states is small if the trace distance between their local marginals is small in addition to a condition on the CMI.

Lemma 6. Consider two quantum states $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$ and $\sigma \in \mathcal{D}(\mathcal{H}^{(n)})$. Given a shielding width $w \geq 1$, we let $B_i = \{i - w + 1, \dots, i\}$ and $\bar{C}_i = \{i + 1, \dots, n\}$. Assuming that the reduced density matrices satisfy $\frac{1}{2} \|\rho(i - w, \dots, i) - \sigma(i - w, \dots, i)\|_1 \leq \varepsilon \leq 1$ for all $w + 1 \leq i \leq n$, the trace distance between ρ and σ is bounded by

$$\|\rho - \sigma\|_1^2 \leq 4 \ln 2 [R_\rho + R_\sigma + 8n\varepsilon \log_2 (d/\varepsilon)], \quad (\text{B13})$$

where

$$R_\rho := \sum_{i=w+1}^{n-1} I_\rho(i - w : \bar{C}_i | B_i) \quad (\text{B14})$$

and similarly for σ .

Proof. We follow Ref. [78] to define the concavity gap of the von Neumann entropy [79],

$$\Delta := S(\tau) - \frac{1}{2}S(\rho) - \frac{1}{2}S(\sigma), \quad (\text{B15})$$

where $\tau := \frac{1}{2}(\rho + \sigma)$. The concavity gap can be expressed as

$$\Delta = \frac{1}{2}D(\rho||\tau) + \frac{1}{2}D(\sigma||\tau), \quad (\text{B16})$$

where $D(\omega||\eta) = \text{Tr}[\omega(\log_2 \omega - \log_2 \eta)]$ denotes the relative entropy of ω with respect to η . To make $D(\omega||\eta)$ finite, we must have $\text{supp}(\omega) \subseteq \text{supp}(\eta)$, which is automatically satisfied for $D(\rho||\tau)$ and $D(\sigma||\tau)$ because ρ and σ are positive semi-definite and thus $\text{supp}(\rho) \subseteq \text{supp}(\rho + \sigma)$ and $\text{supp}(\sigma) \subseteq \text{supp}(\rho + \sigma)$. By the quantum Pinsker inequality [64], for any two states ω and η ,

$$D(\omega||\eta) \geq \frac{1}{2 \ln 2} \|\omega - \eta\|_1^2 \quad (\text{B17})$$

such that

$$\|\rho - \sigma\|_1^2 \leq 8 \ln 2 \Delta. \quad (\text{B18})$$

To establish an upper bound on Δ , we relate it to the Markov entropy and CMI. By Lemma 3, $R_\omega = S_\omega^M - S_\omega$ such that

$$\Delta = \frac{1}{2}(S_\tau - S_\rho^M) + \frac{1}{2}(S_\tau - S_\sigma^M) + \frac{1}{2}(R_\rho + R_\sigma). \quad (\text{B19})$$

Moreover, since the CMI is nonnegative, Lemma 3 implies that $S_\omega^M \geq S_\omega$. It follows that

$$\Delta \leq \frac{1}{2}(S_\tau^M - S_\rho^M) + \frac{1}{2}(S_\tau^M - S_\sigma^M) + \frac{1}{2}(R_\rho + R_\sigma). \quad (\text{B20})$$

By noting that

$$\|\tau - \rho\|_1 = \|\tau - \sigma\|_1 = \frac{1}{2}\|\rho - \sigma\|_1, \quad (\text{B21})$$

we obtain from Lemma 5

$$|S_\tau^M - S_\rho^M| \leq 2n\epsilon \log_2(2d/\epsilon) \leq 4n\epsilon \log_2(d/\epsilon) \quad (\text{B22})$$

and similarly for $|S_\tau^M - S_\sigma^M|$. The lemma follows by substituting this bound into Eqs. (B20) and (B18). ■

Applied to our problem, we choose ρ to be the physical state with exponentially decaying CMI, which immediately implies that R_ρ is exponentially suppressed in the size w of the Markov shield. We further let $\sigma = \sigma_n$ to be sequentially generated according to Definition 4. The next lemma relates the properties of σ_n pertinent to Lemma 6 to the properties of previously generated states.

Lemma 7. Consider a state $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$ and a set of sequentially generated states $\sigma_i \in \mathcal{D}(\mathcal{H}^{(n)}(1, 2, \dots, i))$ with reconstruction width $w \geq 1$ according to Definition 4. Assuming that

$$\frac{1}{2}\|\rho(i-2w, \dots, i) - \sigma_i(i-2w, \dots, i)\|_1 \leq \epsilon \leq 1 \quad (\text{B23})$$

for all $2w+1 \leq i \leq n$, the quantities R_ρ and R_{σ_n} defined in Lemma 6 satisfy

$$R_{\sigma_n} \leq R_\rho + 16wn\epsilon \log_2(d/\epsilon). \quad (\text{B24})$$

Moreover, the final reconstructed state satisfies

$$\frac{1}{2}\|\rho(i-w, \dots, i) - \sigma_n(i-w, \dots, i)\|_1 \leq \epsilon. \quad (\text{B25})$$

for all $w+1 \leq i \leq n$.

Proof. Let $B_i = \{i-w+1, \dots, i\}$ and $\bar{C}_i = \{i+1, \dots, n\}$ and define $\bar{A}_i = \{1, 2, \dots, i-w\}$. Using Definition 4, we may express the sequentially generated state as $\sigma_n = \Phi_i[\sigma_i]$, where $\Phi_i = (\mathcal{I}_{\bar{A}_{n-1}} \otimes \mathcal{R}_{n-1}) \circ \dots \circ (\mathcal{I}_{\bar{A}_i} \otimes \mathcal{R}_i)$. Here, Φ_i is a CPTP map that extends subsystem B_i to $B_i \cup \bar{C}_i$. Hence, $I_{\sigma_n}(i-w: \bar{C}_i|B_i) = I_{\Phi_{i+w}[\sigma_{i+w}]}(i-w: B_{i+w} \cup \bar{C}_{i+w}|B_i)$. Since Φ_{i+w} does not act on B_i , we may apply the data processing inequality for the CMI [80,81] to obtain $I_{\sigma_n}(i-w: \bar{C}_i|B_i) \leq I_{\sigma_{i+w}}(i-w: B_{i+w}|B_i)$ and

$$R_{\sigma_n} \leq \sum_{i=w+1}^{n-1} I_{\sigma_{i+w}}(i-w: B_{i+w}|B_i). \quad (\text{B26})$$

By the continuity of the CMI, Lemma 4, together with the assumption in Eq. (B23),

$$I_{\sigma_{i+w}}(i-w: B_{i+w}|B_i) \leq I_\rho(i-w: B_{i+w}|B_i) + 8\epsilon \log_2(d^{w+1}/\epsilon). \quad (\text{B27})$$

Since $\rho(\{i-w\} \cup B_i \cup B_{i+w}) = \text{tr}_{\bar{C}_{i+w}}[\rho(\{i-w\} \cup B_i \cup \bar{C}_i)]$, the data processing inequality implies that $I_\rho(i-w: B_{i+w}|B_i) \leq I_\rho(i-w: \bar{C}_i|B_i)$. We thus obtain

$$R_{\sigma_n} \leq R_\rho + 8(n-w-1)\epsilon \log_2(d^{w+1}/\epsilon), \quad (\text{B28})$$

which can be simplified to the first statement in the lemma for $w \geq 1$.

For the second part of the lemma, we first note that

$$\frac{1}{2}\|\rho(n-2w, \dots, n) - \sigma_n(n-2w, \dots, n)\|_1 \leq \epsilon \quad (\text{B29})$$

by assumption. Since the trace distance is non-increasing under the partial trace, Eq. (B25) thus holds for all $n-w \leq i \leq n$. To establish the result for $i < n-w$, we observe that the map $\mathcal{R}_{i+w}$ only acts on sites with index greater than i . Because the maps are trace preserving, this implies $\sigma_n(i-w, \dots, i) = \sigma_{i+w}(i-w, \dots, i)$. Hence,

$$\frac{1}{2}\|\rho(i-w, \dots, i) - \sigma_n(i-w, \dots, i)\|_1 = \frac{1}{2}\|\rho(i-w, \dots, i) - \sigma_{i+w}(i-w, \dots, i)\|_1 \quad (\text{B30})$$

$$\leq \frac{1}{2}\|\rho(i-w, \dots, i+w) - \sigma_{i+w}(i-w, \dots, i+w)\|_1 \leq \epsilon. \quad (\text{B31})$$

■

The above lemma applies to all sequentially generated states. However, it is not immediately useful as it depends on the trace distances between reduced density matrices of ρ and the sequentially generated states, which are generally unknown. We will now show that for locally optimal recovery maps according to Definition 5, the trace distances can be bounded in terms of the CMI with respect to ρ and the local tomography uncertainty.

Proposition 8. Consider a quantum state $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$ with exponentially decaying CMI according to Definition 2. Let $\hat{\rho}_i$ be an estimate of the reduced density operator $\rho(i - 2w, \dots, i)$ satisfying

$$\frac{1}{2} \|\rho(i - 2w, \dots, i) - \hat{\rho}_i\|_1 \leq \eta \quad (\text{B32})$$

for all $2w + 1 \leq i \leq n$. Further, let $\sigma_i \in \mathcal{D}(\mathcal{H}^{(n)}(1, 2, \dots, i))$ denote the sequentially generated states that are locally optimal with respect to $\hat{\rho}_i$ (Definition 5). Then, for all $2w + 1 \leq i \leq n$,

$$\frac{1}{2} \|\rho(i - 2w, \dots, i) - \sigma_i(i - 2w, \dots, i)\|_1 \leq (2\eta + a e^{-w/2\xi})n. \quad (\text{B33})$$

Proof. Let $A_i = \{i - 2w + 1, \dots, i - w\}$, $B_i = \{i - w + 1, \dots, i\}$, and $C_i = \{i + 1\}$. Consider the trace distance

$$\|\hat{\rho}_{i+1} - \sigma_{i+1}(i - 2w, \dots, i)\|_1 = \|\hat{\rho}_{i+1} - \sigma_{i+1}(A_i B_i C_i)\|_1 = \|\hat{\rho}_{i+1} - (\mathcal{I}_{A_i} \otimes \mathcal{R}_i)[\sigma_i(A_i B_i)]\|_1, \quad (\text{B34})$$

where $\mathcal{R}_i \in \mathcal{S}(\mathcal{H}^{(n)}(B_i), \mathcal{H}^{(n)}(B_i C_i))$ is the locally optimal recovery map. By definition, $\mathcal{R}_i$ minimizes this trace distance and we may use any other recovery map to obtain an upper bound. In particular, we will consider the rotated Petz recovery map (switching briefly to index j to distinguish it from the imaginary unit in the exponent) [61]

$$\mathcal{R}_j^{\text{Petz}}[\cdot] = \frac{\pi}{2} \int_{-\infty}^{\infty} dt \frac{1}{\cosh(\pi t) + 1} \rho(B_j C_j)^{1/2 - it} \rho(B_j)^{-1/2 + it} [\cdot] \rho(B_j)^{-1/2 - it} \rho(B_j C_j)^{1/2 + it}. \quad (\text{B35})$$

The rotated Petz recovery map has the important property (see Ref. [61], Theorem 2.1)

$$\frac{1}{2} \|\rho(A_i B_i C_i) - \mathcal{R}_i^{\text{Petz}}[\rho(A_i B_i)]\|_1 \leq \sqrt{I_\rho(A_i : C_i | B_i)}. \quad (\text{B36})$$

By using the rotated Petz recovery map to bound Eq. (B34), we obtain

$$\|\hat{\rho}_{i+1} - \sigma_{i+1}(A_i B_i C_i)\|_1 \leq \|\hat{\rho}_{i+1} - (\mathcal{I}_{A_i} \otimes \mathcal{R}_i^{\text{Petz}})[\sigma_i(A_i B_i)]\|_1 \quad (\text{B37})$$

$$\leq \|\hat{\rho}_{i+1} - \rho(A_i B_i C_i)\|_1 + \|\rho(A_i B_i C_i) - (\mathcal{I}_{A_i} \otimes \mathcal{R}_i^{\text{Petz}})[\rho(A_i B_i)]\|_1 + \|(\mathcal{I}_{A_i} \otimes \mathcal{R}_i^{\text{Petz}})[\rho(A_i B_i) - \sigma_i(A_i B_i)]\|_1, \quad (\text{B38})$$

where we used the triangle inequality. The first term on the right-hand side is bounded by 2η by assumption while the second term is bounded by Eq. (B36). For the final term, we use the fact that the trace distance is non-increasing under CPTP maps, yielding

$$\frac{1}{2} \|\hat{\rho}_{i+1} - \sigma_{i+1}(A_i B_i C_i)\|_1 \leq \eta + \sqrt{I_\rho(A_i : C_i | B_i)} + \frac{1}{2} \|\rho(A_i B_i) - \sigma_i(A_i B_i)\|_1 \quad (\text{B39})$$

$$\leq \eta + \sqrt{I_\rho(A_i : C_i | B_i)} + \frac{1}{2} \|\rho(A_{i-1} B_{i-1} C_{i-1}) - \sigma_i(A_{i-1} B_{i-1} C_{i-1})\|_1. \quad (\text{B40})$$

In the second line, we used the relation $\{i - 2w\} \cup A_i \cup B_i = A_{i-1} \cup B_{i-1} \cup C_{i-1}$ along with the fact that the trace distance is non-increasing under the partial trace. Defining $\delta_{i+1} := \frac{1}{2} \|\rho(A_i B_i C_i) - \sigma_{i+1}(A_i B_i C_i)\|_1$, we thus find

$$\delta_{i+1} \leq \eta + \frac{1}{2} \|\hat{\rho}_{i+1} - \sigma_{i+1}(A_i B_i C_i)\|_1 \leq 2\eta + \sqrt{I_\rho(A_i : C_i | B_i)} + \delta_i. \quad (\text{B41})$$

The initial state σ_{2w+1} is obtained from $\hat{\rho}_{2w+1}$ by discarding negative eigenvalues and renormalizing to unit trace (Algorithm 1). Since the negative eigenvalues of the unit-trace estimate carry weight at most

$\frac{1}{2} \|\hat{\rho}_{2w+1} - \rho(1, \dots, 2w+1)\|_1 \leq \eta$, and the renormalization changes the operator by at most the same amount, the projection at most doubles the initial error, $\delta_{2w+1} \leq 2\eta$. The later estimates $\hat{\rho}_{i+1}$ serve only as reference operators in the optimization and need not be positive semidefinite. Iterating down to this boundary condition, we obtain

$$\delta_{i+1} \leq 2(i-2w+1)\eta + \sum_{j=2w+1}^i \sqrt{I_\rho(A_j : C_j | B_j)}. \quad (\text{B42})$$

Using the fact that the CMI of ρ decays exponentially, we finally arrive at

$$\max_{2w+1 \leq i \leq n} \delta_i \leq (2\eta + a e^{-w/2\xi})n. \quad (\text{B43})$$

Finally, we combine all results into our main theorem. *Theorem 9.* Consider a quantum state $\rho \in \mathcal{D}(\mathcal{H}^{(n)})$ with exponentially decaying CMI according to Definition 2. Let $\hat{\rho}_i$ be an estimate of the reduced density operator

$\rho(i-2w, \dots, i)$ satisfying

$$\frac{1}{2} \|\rho(i-2w, \dots, i) - \hat{\rho}_i\|_1 \leq \eta \quad (\text{B44})$$

for all $2w+1 \leq i \leq n$. Further, let $\sigma_i \in \mathcal{D}(\mathcal{H}^{(n)}(1, 2, \dots, i))$ denote the sequentially generated states that are locally optimal with respect to $\hat{\rho}_i$ (Definition 5). There exist constants b and n_0 such that

$$\|\rho - \sigma_n\|_1 \leq \varepsilon. \quad (\text{B45})$$

for any $\varepsilon \leq 1$ if $w \geq 4\xi \ln[(bn/\varepsilon) \ln(bn/\varepsilon)]$, $\eta \leq (a/2)e^{-w/2\xi}$, and $n \geq n_0$.

Proof. We will show that the theorem holds for $n_0 = \max\{d/2a, e/b, 2a/b^2, \xi/12\}$.

Given the assumption $\eta \leq (a/2)e^{-w/2\xi}$, Proposition 8 gives the bound

$$\frac{1}{2} \|\rho(i-2w, \dots, i) - \sigma_i(i-2w, \dots, i)\|_1 \leq 2an e^{-w/2\xi} \leq \frac{2an}{(bn/\varepsilon)^2 \ln^2(bn/\varepsilon)}. \quad (\text{B46})$$

Since $\varepsilon \leq 1$ and $n \geq n_0$, the right-hand side is bounded from above by 1, which allows us to use the bounds in Lemma 7. Explicitly,

$$R_{\sigma_n} \leq R_\rho + 32awn^2[\log(d/2an) + w/2\xi]e^{-w/2\xi} \leq R_\rho + (16aw^2n^2/\xi)e^{-w/2\xi}, \quad (\text{B47})$$

$$\frac{1}{2} \|\rho(i-w, \dots, i) - \sigma_n(i-w, \dots, i)\|_1 \leq 2an e^{-w/2\xi}. \quad (\text{B48})$$

By substituting these bounds into Lemma 6, we obtain

$$\|\rho - \sigma\|_1^2 \leq 8 \ln 2 [R_\rho + (8aw^2n^2/\xi)e^{-w/2\xi} + 8an^2(\log_2(d/2an) + w/2\xi)e^{-w/2\xi}] \quad (\text{B49})$$

$$\leq 8 \ln 2 [R_\rho + (12aw^2n^2/\xi)e^{-w/2\xi}], \quad (\text{B50})$$

where we used $n \geq d/2a$ and $w \geq 1$. The definition of R_ρ (Lemma 6) in conjunction with the exponentially decaying CMI assumption (Definition 2) yields $R_\rho \leq an e^{-w/\xi}$. Since $n \geq \xi/12$, the second term in the square brackets of Eq. (B50) bounds R_ρ from above. Therefore,

$$\|\rho - \sigma\|_1^2 \leq 192 \ln 2 (aw^2n^2/\xi)e^{-w/2\xi}. \quad (\text{B51})$$

The function $x^2 e^{-x}$ decreases monotonically for all $x \geq 2$. This implies that we may substitute a lower bound on w into Eq. (B51) provided that $w \geq 4\xi$, which is indeed satisfied since $n \geq e/b$. We thus obtain

$$\|\rho - \sigma\|_1^2 \leq 3072 \ln 2 (a\xi/b^2)e^2 \left[1 + \frac{\ln \ln(bn/\varepsilon)}{\ln(bn/\varepsilon)} \right]^2. \quad (\text{B52})$$

The term in square brackets is bounded from above by $1 + 1/e \leq \sqrt{2}$. Setting $b^2 = 6144 \ln 2 a \xi$ thus establishes the theorem. ■

APPENDIX C: SAMPLE COMPLEXITY

1. Process tomography with classical shadows

The sample complexity of our algorithm is governed by the complexity of estimating the reduced channels $J_i = \text{tr}_{[n] \setminus X_i}[J(\Lambda)]$ for the subsystems $X_i = \{i - 2w, \dots, i\}$, $i \in \{2w + 1, \dots, n\}$. For concreteness, we employ the ShadowQPT protocol of Ref. [17]. Applied to our formalism, the gates U_i are chosen such that they prepare with equal probability one of the 6 eigenstates of the Pauli operators X , Y , and Z . Similarly, the V_i are drawn from a random distribution such that the measurement is performed in one of the Pauli bases with equal probability. Instead of specifying U_i and V_i , we may equivalently describe the input state $|\mu_i^{\text{in}}, b_i^{\text{in}}\rangle$ and the measured state $|\mu_i^{\text{out}}, b_i^{\text{out}}\rangle$, where $\mu_i^{\text{in}, \text{out}} \in \{X, Y, Z\}$ specifies the Pauli basis and $b_i^{\text{in}, \text{out}} \in \{-1, +1\}$ the eigenvalue associated with the prepared or measured state, respectively.

For a given shot, it is straightforward to restrict the record to a subsystem of interest X_i by only retaining $(\mu_j^{\text{in}}, b_j^{\text{in}}, \mu_j^{\text{out}}, b_j^{\text{out}})$ for $j \in X_i$. Denoting such a record for a given subsystem by r , we define the classical shadow

$$\hat{\zeta}_i(r) := \bigotimes_{j \in X_i} [\tau(\mu_j^{\text{in}}, b_j^{\text{in}})^T \otimes \tau(\mu_j^{\text{out}}, b_j^{\text{out}})], \quad (\text{C1})$$

where

$$\tau(\mu, b) := 3|\mu, b\rangle\langle\mu, b| - \mathbb{I} \quad (\text{C2})$$

and $(\cdot)^T$ denotes transpose in the computational basis. The classical shadow is an unbiased estimator for the reduced Choi state J_i : $\mathbb{E}[\hat{\zeta}_i(r)] = J_i$ [17].

Given M i.i.d. shots described by records $r_1, \dots, r_M$, we estimate the reduced Choi state by the empirical mean

$$\hat{J}_i := \frac{1}{M} \sum_{k=1}^M \hat{\zeta}_i(r_k), \quad (\text{C3})$$

so that $\mathbb{E}[\hat{J}_i] = J_i$. The distance of $\hat{J}_i$ from J_i can be bounded using the following theorem.

Theorem 10. (Sample complexity of ShadowQPT, Theorem II.1 of Ref. [17]). Let Λ be an n -qubit channel and let $J(\Lambda)$ be its normalized Choi state. Let $X_i \subseteq [n]$ be a subsystem comprising at most k qubits, where i is an index that enumerates all such subsystems. We denote by $J_i = \text{tr}_{[n] \setminus X_i}[J(\Lambda)]$ the corresponding reduced Choi state and by $\hat{J}_i$ the estimator obtained according to Eq. (C3) with M samples. There exists a constant $C > 0$ such that for any $\eta, \delta \in (0, 1)$,

$$M \geq C \frac{144^k}{\eta^2} \ln \left[\frac{(24n)^{2k}}{\delta} \right] \quad (\text{C4})$$

implies

$$\frac{1}{2} \|\hat{J}_i - J_i\|_1 \leq \eta \quad (\text{C5})$$

for all i with probability at least $1 - \delta$.

2. Proof of Corollary 2

Proof. Corollary 2 follows by combining Theorem 9 with Theorem 10. Theorem 9 requires estimating the reduced Choi states on $k = 2w + 1$ sites with an error $\eta \leq (a/2)e^{-w/2\xi}$. By substituting these values into Eq. (C4), we obtain

$$M \geq C' e^{(4 \ln 12 + 1/\xi)w} \left[\ln \left(\frac{1}{\delta} \right) + (4w + 2) \ln(24n) \right], \quad (\text{C6})$$

where $C' = 4 \times 144C/a^2$. Next, we insert the condition $w \geq 4\xi \ln[(bn/\varepsilon) \ln(bn/\varepsilon)]$ from Theorem 9, which yields

$$M \geq C' \left[\frac{bn}{\varepsilon} \ln \left(\frac{bn}{\varepsilon} \right) \right]^{16 \ln 12 \xi + 4} \left[\ln \left(\frac{1}{\delta} \right) + 16\xi \ln \left(\frac{bn}{\varepsilon} \ln \left(\frac{bn}{\varepsilon} \right) \right) \ln(24n) + 2 \ln(24n) \right]. \quad (\text{C7})$$

Using the (crude) bound $\ln(bn/\varepsilon) \leq bn/\varepsilon$, we see that Eq. (C7) is clearly of the form of Eq. (9) in Corollary 2. ■

APPENDIX D: DIAMOND NORM AND CHOI STATE BOUNDS

Our guarantees in Theorem 1 are stated in terms of the trace distance between the global Choi states and corresponding process matrix. In many applications, however,

the metric for comparing quantum channels is the diamond norm. Here we relate our Choi/process matrix bounds to diamond norm error.

Let A be the input system with $\dim(A) = D = d^n$ and define $\Delta := \Lambda - \Lambda'$. The diamond norm is [82]

$$\|\Delta\|_\diamond = \max_{\rho_{AR}} \|(\Delta \otimes \mathbb{I}_R)(\rho_{AR})\|_1,$$

where R is an ancilla with $\dim(R) = \dim(A) = D$ and the maximization is over density operators $\rho_{AR} \geq 0$ with $\text{Tr}(\rho_{AR}) = 1$.

Let $J(\Lambda)$ denote the Choi state. A standard inequality relates the diamond norm to the trace distance of Choi states:

$$\|\Lambda - \Lambda'\|_{\diamond} \leq D \|J(\Lambda) - J(\Lambda')\|_1, \quad (\text{D1})$$

where an exponential prefactor $D = d^n$ appears in the inequality.

We make the additional assumption that Λ and Λ' are Pauli channels, i.e., their process matrices are diagonal in the Pauli basis with diagonal entries $\chi_{a,a}$ and $\hat{\chi}_{a,a}$. Define $\delta_a := \chi_{a,a} - \hat{\chi}_{a,a}$. Then

$$\Delta(\rho) = \sum_a \delta_a P_a \rho P_a^\dagger.$$

Consequently, for any ρ_{AR} ,

$$(\Delta \otimes \mathbb{I}_R)(\rho_{AR}) = \sum_a \delta_a (P_a \otimes I_R) \rho_{AR} (P_a \otimes I_R).$$

Using the triangle inequality and unitary invariance of the trace distance,

$$\begin{aligned} \|(\Delta \otimes \mathbb{I}_R)(\rho_{AR})\|_1 &\leq \sum_a |\delta_a| \|(P_a \otimes I_R) \rho_{AR} (P_a \otimes I_R)\|_1 \\ &= \sum_a |\delta_a| \|\rho_{AR}\|_1 = \sum_a |\delta_a|, \end{aligned}$$

since $\|\rho_{AR}\|_1 = \text{Tr}(\rho_{AR}) = 1$. Maximizing over ρ_{AR} yields

$$\|\Lambda - \Lambda'\|_{\diamond} = \|\Delta\|_{\diamond} \leq \sum_a |\delta_a|.$$

Finally, because χ and $\hat{\chi}$ are diagonal,

$$\|\chi - \hat{\chi}\|_1 = \sum_a |\chi_{a,a} - \hat{\chi}_{a,a}| = \sum_a |\delta_a|,$$

and therefore

$$\|\Lambda - \Lambda'\|_{\diamond} \leq \|\chi - \hat{\chi}\|_1. \quad (\text{D2})$$

By the Choi state to process matrix transformation, we can also guarantee in terms of the Choi states:

$$\|\Lambda - \Lambda'\|_{\diamond} \leq c \|J(\Lambda) - J(\Lambda')\|_1, \quad (\text{D3})$$

up to a constant factor c . Hence, for Pauli channels we show that the exponential prefactor in Eq. (D1) can be removed.

APPENDIX E: NUMERICAL EVIDENCE FOR EXPONENTIALLY DECAYING CMI

In this appendix, we provide numerical evidence for the exponential decay of the CMI of Choi states in the sense of Definition 1. We organize the examples into three classes. First, we study the interacting Heisenberg chain of Sec. IV A corresponding to the open-system application. Second, we consider shallow circuits with Pauli noise and coherent crosstalk corresponding to the noisy-circuit application of Sec. IV B. Third, we consider random Clifford circuits subject to Pauli noise, with circuit depths up to 10. Within the parameter regimes examined, these examples support the assumption of exponentially decaying CMI.

Throughout this appendix, Λ denotes the channel under consideration and $J(\Lambda)$ its Choi state, Eq. (2), which we regard as a state on a chain of n sites, where site i comprises the “in” and the “out” qubit of the physical qubit i . We consider tripartitions of the chain into adjacent intervals A , B , and C of $a = |A|$, $b = |B|$, and $c = |C|$ sites, with B located between A and C , and we evaluate the CMI $I_{J(\Lambda)}(A : C|B) = S(AB) + S(BC) - S(B) - S(ABC)$, where $S(R)$ denotes the von Neumann entropy, measured in bits, of the reduced Choi state on the region R . We fit the data to the form $a_0 e^{-b/\xi}$ and refer to ξ as the decay length. The prefactor a_0 plays the role of the constant a in Definition 1, which we rename to avoid confusion with the size a of the region A . We note that the CMI considered here refers to the Choi state $J(\Lambda)$ of the noisy channel, which is the average over all noise realizations before evaluating the CMI. This differs from evaluating the CMI for each noise realization with fixed spacetime locations of depolarization and then averaging over all noise realizations. This averaged CMI has been found to spread superlinearly in noisy random circuits [71].

1. Lindbladian dynamics with a coherent Hamiltonian

To probe the CMI decay in the presence of a coherent interaction, we implement a tensor-network evaluation for the model studied in Sec. IV A, with $n = 50$ and local anisotropic noise with rates $\gamma_x = \gamma_y = 1/2$, $\gamma_z = 1$. We compute $I_{J(\Lambda)}(A_s : C_s|B_s)$ for windows with $|A| = |C| = 1$ and conditioning regions of size $|B| = b \in \{1, \dots, 4\}$, sweeping the window position s through the bulk of the chain and reporting the maximum, $I_J^{\max}(1, b, 1)$. The entropies are obtained by exact diagonalization of the reduced Choi states on windows of up to seven sites.

The results are shown in Fig. 4. At all times considered, the CMI is bounded by an exponential envelope, $I_J^{\max} \leq a_0 e^{-\kappa b}$, with decay lengths $\xi = 1/\kappa \approx 0.15\text{--}0.18$ as shown in Fig. 4(a). The CMI at fixed conditioning size grows monotonically with the evolution time, as shown in Fig. 4(b), and this growth parallels that of the intrinsic

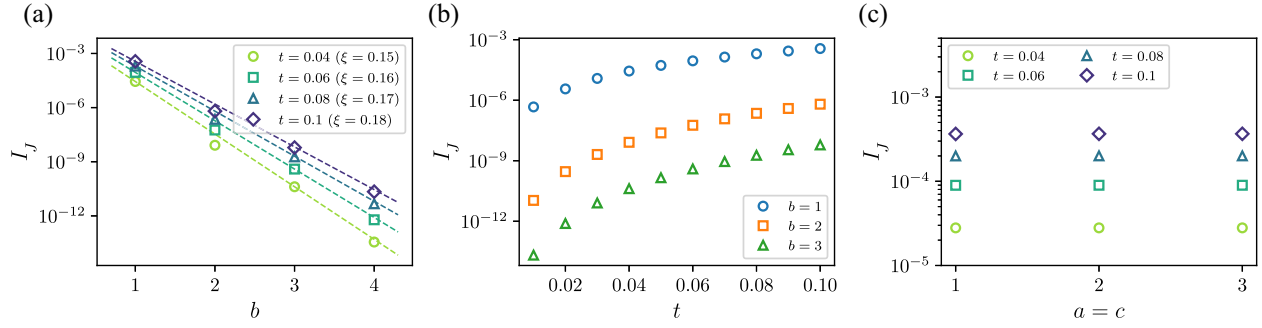


FIG. 4. Choi-state CMI for the open Heisenberg chain with local anisotropic noise ($n = 50$, $\gamma_x = \gamma_y = 1/2$, $\gamma_z = 1$), i.e., the model studied in Sec. IV A, including the coherent Heisenberg interaction of Eq. (11). (a) $I_J^{\max}(1, b, 1)$ as a function of the size b of the conditioning region at four evolution times. The data points for a fixed t are upper bounded by an exponential decay function $a_0 e^{-\kappa b}$, corresponding to decay lengths $\xi = 1/\kappa \approx 0.15$ – 0.18 sites that grow slowly with the evolution time. (b) $I_J^{\max}(1, b, 1)$ as a function of the evolution time t at fixed $b = 1, 2, 3$. The CMI grows monotonically with t and matches the growth of the reconstruction error in Fig. 2(b). (c) Dependence on the size $a = c$ at fixed $b = 1$, with the conditioning region fixed at the center of the chain as a grows. The CMI increases monotonically with a , but the total increase from $a = 1$ to $a = 3$ remains below 0.4% and is invisible on this scale.

reconstruction error in Fig. 2(b). Figure 4(c) shows that the CMI is almost independent of the sizes a and c , which is consistent with the uniformity requirement in Definition 1. These decay lengths are far below the reconstruction windows used in Sec. IV A, in agreement with the accuracy and stability of the reconstruction observed there.

2. Shallow circuits with coherent and Pauli noise

The noisy-circuit channel $\Lambda_{U, \epsilon, \gamma}$ of Sec. IV B with vanishing coherent perturbation, $\epsilon = 0$, belongs to the class of Clifford circuits with Pauli noise. In this case, the Choi state factorizes exactly into two-site blocks: the circuit is a single layer of disjoint two-qubit gates, and the on-site Pauli noise acts within the individual blocks, so $J(\Lambda)$ remains a tensor product of two-site blocks for any γ . Consequently, $I_{J(\Lambda)}(A: C|B) = 0$ for all contiguous tripartitions with $b \geq 1$. When $\epsilon > 0$, the coherent crosstalk

error U_ϵ takes the Choi state out of the stabilizer formalism, and we evaluate the CMI by tensor-network methods. For the same channel family as in the main text, we take $n = 20$, $\gamma = 10^{-2}$, $\epsilon = r\gamma$ with $r \in \{1, \dots, 10\}$. We consider $U_\epsilon = \bigotimes_{i=\text{odd}} U_i \circ \bigotimes_{i=\text{even}} U_i$, where $U_i = e^{i\epsilon(\alpha_1 X_i X_{i+1} + \alpha_2 Y_i Y_{i+1} + \alpha_3 Z_i Z_{i+1})}$ with the couplings $\alpha_{1,2,3}$ drawn uniformly at random on every bond. We construct the Choi state of the shallow circuit as a matrix-product operator, and the entropies of contiguous windows are obtained by exact diagonalization. Here, we define $I_J^{\max}(a, b, c)$ as the maximum of the CMI over the three central placements of the window (the centered window and its shifts by one site to either side), for contiguous regions with $|A| = a$, $|B| = b$, and $|C| = c$. We report the average of $I_J^{\max}$ over 10 independent disorder realizations of U_ϵ .

The results are shown in Fig. 5. In Fig. 5(a), we observe that the CMI decays rapidly within the causal cone of the

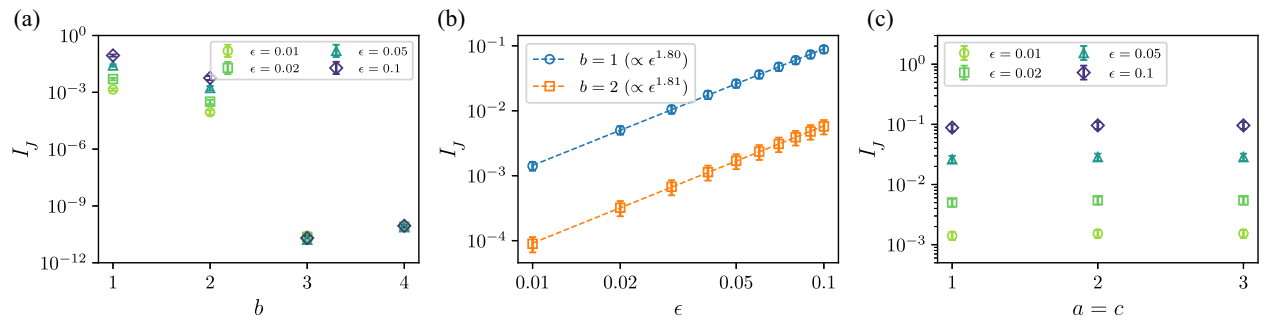


FIG. 5. Choi-state CMI of the noisy shallow circuit of Sec. IV B, including the coherent crosstalk layer U_ϵ ($n = 20$, $\gamma = 10^{-2}$, $\epsilon = r\gamma$ with $r \in \{1, \dots, 10\}$). Data points are averaged over 10 independent disorder realizations of U_ϵ . Error bars denote the standard error $\sigma/\sqrt{N}$ with $N = 10$ realizations, where σ is the standard deviation over realizations. (a) $I_J^{\max}(1, b, 1)$ as a function of the size b of the conditioning region. The CMI decays rapidly inside the causal cone of the circuit and drops to the numerical floor for $b \geq 3$. (b) Scaling with the strength ϵ of the coherent perturbation at fixed b , with fitted power laws $I_J \propto \epsilon^{1.80}$ ($b = 1$) and $\propto \epsilon^{1.81}$ ($b = 2$) (dashed lines). (c) Dependence on the size $a = c$ at fixed $b = 1$. The CMI increases monotonically by less than 10% from $a = 1$ to $a = 2$ and saturates thereafter.

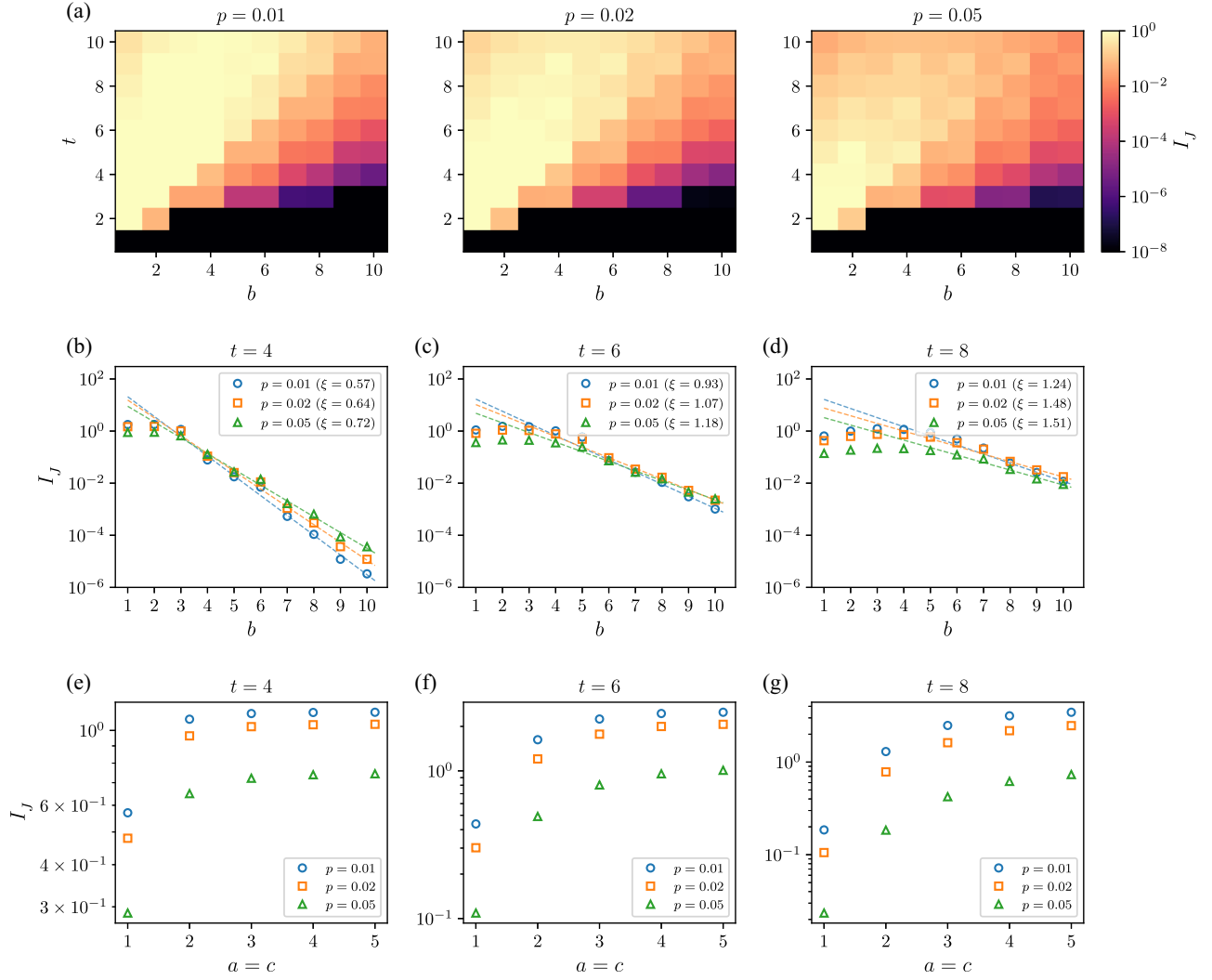


FIG. 6. Exact Choi-state CMI of random brickwork Clifford circuits with single-site depolarizing noise of strength p applied after every layer. Here we take $n = 40$, $|A| = |C| = 1$ in panels (a)–(d), with the sliding-window maximum over positions. Each data point is averaged over 10^3 independent Clifford circuit realizations. (a) CMI as a function of the circuit depth t and the size b of the conditioning region for $p = 0.01, 0.02$, and 0.05 . Panels (b)–(d) show the CMI decay with respect to the conditioning size b at different depths $t = 4, 6, 8$. Within each panel, the curves correspond to $p = 0.01, 0.02, 0.05$. The dashed lines are exponential fits to the region $b \geq t$ with decay length $\xi = 1/\kappa$, where κ is the decay rate. Panels (e)–(g) show the dependence on the size $a = |A| = |C|$ at fixed $b = 2$ for depths $t = 4, 6, 8$. For each realization, the conditioning region B is centered in the chain. The CMI increases monotonically with a and saturates on the scale of the light cone.

circuit and drops to the numerical floor beyond it when $b \geq 3$, where the outer regions are causally disconnected. In Fig. 5(b), we also observe that the CMI grows as an approximate power law in the coherent perturbation strength, $I_J \propto \epsilon^{1.80}$ at $b = 1$ and $I_J \propto \epsilon^{1.81}$ at $b = 2$. In Fig. 5(c), we increase the sizes a and c and find that the CMI depends only weakly on $a = c$, consistent with the requirement of Definition 1. Even at the largest coherent perturbation considered, the CMI at $b = 1$ remains of order 10^{-1} bits and is negligible for $b \geq 3$, which explains why the reconstruction with $w = 1$ in Sec. IV B performs well over most of the parameter range of Fig. 3, with reduced accuracy only at the strongest coherent perturbation.

3. Random Clifford circuits with Pauli noise

Consider a discrete noisy circuit

$$\Lambda = \mathcal{E}_{\text{noise}} \circ \mathcal{U}_{\text{Cliff}}, \quad (\text{E1})$$

where $\mathcal{U}_{\text{Cliff}}$ is a Clifford layer and $\mathcal{E}_{\text{noise}}$ is a Pauli channel,

$$\mathcal{E}_{\text{noise}}(\rho) = \sum_e p_e P_e \rho P_e, \quad p_e \geq 0, \quad \sum_e p_e = 1, \quad (\text{E2})$$

where the P_e are Pauli strings and p_e their probabilities. For channels of this form, $J(\Lambda)$ is a classical mixture of

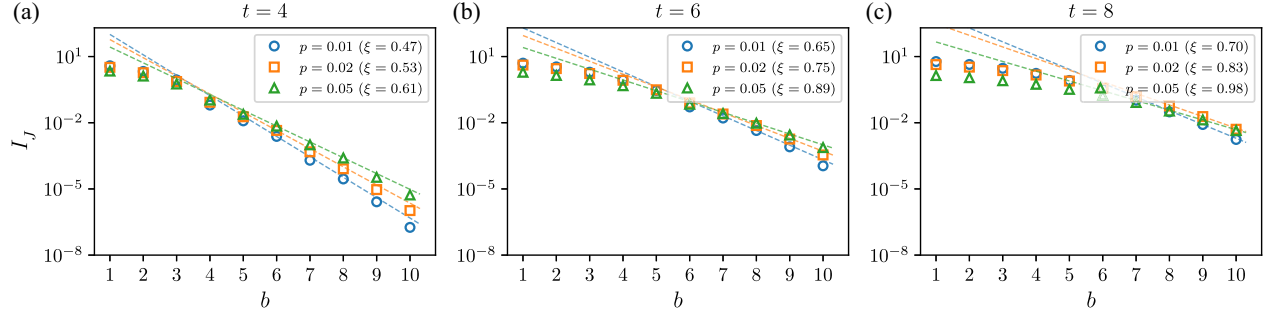


FIG. 7. Full-partition Choi-state CMI of random brickwork Clifford circuits with depolarizing noise of strength p , where the tripartition $A|B|C$ covers the entire chain of $n = 12$. For each conditioning size b , the regions A and C comprise all sites to the left and to the right of B , respectively, the position of B is swept over all $n - b - 1$ placements with $a, c \geq 1$, and we report the maximum CMI over the position of B . Each data point is averaged over 200 circuit realizations. Panels (a)–(c) show CMI I_J as a function of the conditioning size b at circuit depths $t = 4, 6, 8$. Inside each panel, the three data sets correspond to noise strengths $p = 0.01, 0.02, 0.05$. Dashed lines are exponential fits over $b \geq t$, with decay lengths ξ .

stabilizer states that share the same stabilizer group up to the signs of the generators, and the entropies of all reduced Choi states can be evaluated exactly by means of the stabilizer formalism [83,84]. The method is described at the end of Appendix F.

We consider random brickwork Clifford circuits of increasing depth up to $t = 10$ layers, subject to local depolarizing noise. At each layer, two-qubit gates drawn uniformly from the two-qubit Clifford group act on alternating bonds, followed by a single-qubit depolarizing channel on every site,

$$\mathcal{E}_p(\rho) = (1 - p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z). \quad (\text{E3})$$

We sweep the window across the chain from left to right, take the maximum over all bulk window positions, and finally average over 10^3 circuit realizations. The results are shown in Fig. 6. Without noise, the Choi state of a depth- t circuit has a strict light cone and CMI vanishes beyond it. We observe that the depolarizing noise smooths this hard cutoff into an exponential tail. Panel (a) provides a heat map of the CMI as a function of the circuit depth t and the conditioning size b . Panels (b)–(d) show CMI decay with respect to the conditioning size b at fixed circuit depths $t = 4, 6, 8$. At each depth and noise strength studied here, the CMI decays exponentially with b for $b \geq t$. We also observe that the fitted decay length $\xi = 1/\kappa$ (dashed lines) grows with the circuit depth. For $p = 0.01$ it increases from $\xi \approx 0.57$ at $t = 4$ to $\xi \approx 1.24$ sites at $t = 8$, indicating the spreading of correlations. At fixed depth, ξ also grows with the noise strength within the weak noise regime, reaching $\xi \approx 1.51$ at $t = 8$ and $p = 0.05$. This is consistent with the observation that local decoherence can increase the range of conditional correlations in stabilizer systems, as captured by the spacetime Markov length of Ref. [85], and in noisy random circuits [71]. In our setting, the

increase suggests that the Pauli-error history mediates classical correlations between the regions A and C . Panels (e)–(g) show the dependence of the CMI on the sizes a and c . The saturation rate is related to the depth t and the decay length. Because the CMI is monotone under enlarging A and C , the observed rapid saturation provides evidence that increasing the outer regions further may have only a limited effect.

We remark that Definition 1 requires the decay to be uniform over all subsystem choices, including A and C regions that extend arbitrarily far. To probe this worst case, we additionally evaluate the CMI of the full-chain tripartition with $a + b + c = n$, where for each conditioning size b , the regions A and C comprise all remaining sites to the left and to the right of B , and we maximize over the position of B . Since the syndrome enumeration for the full chain scales as 2^{2n} , we perform this computation on a chain with $n = 12$, averaging over 200 random Clifford circuit realizations. The results are shown in Fig. 7, which exhibits exponential decay, with decay lengths $\xi \approx 0.47\text{--}0.98$ obtained from fits to the data with $b \geq t$. The exponential decay of the CMI therefore cannot be attributed to small A and C regions.

APPENDIX F: NUMERICAL DETAILS

1. Convex optimization

Convex optimization problems were formulated using the CVXPY modeling framework and solved with the SCS solver (Splitting Conic Solver), a first-order primal-dual method for large-scale conic optimization [86]. For each local step k , we determine a CPTP map $\mathcal{N}_{k \rightarrow k, k+1}$ by minimizing the Frobenius distance between the resulting reconstructed operator $\hat{J}_{B_{k-1}B_kB_{k+1}}(\mathcal{N}_{k \rightarrow k, k+1})$ and the target operator $J_{B_{k-1}B_kB_{k+1}}$,

$$L^* = \min_{\mathcal{N}_{k \rightarrow k, k+1}} \|\hat{J}_{B_{k-1}B_kB_{k+1}}(\mathcal{N}_{k \rightarrow k, k+1}) - J_{B_{k-1}B_kB_{k+1}}\|_F, \quad (\text{F1})$$

subject to the CPTP constraints

$$\text{Tr}_{(B_kB_{k+1})_{\text{out}}}[\mathcal{N}_{k \rightarrow k, k+1}] = \mathbb{I}_{(B_k)_{\text{in}}}, \quad \mathcal{N}_{k \rightarrow k, k+1} \geq 0. \quad (\text{F2})$$

SCS applies a first-order primal-dual splitting algorithm and iteratively updates the optimization variables until the residuals of the Karush-Kuhn-Tucker conditions [69] fall below a numerical tolerance. For further details about convex optimization, we refer to Ref. [86].

2. Tensor-network simulation of the open-system dynamics

For the open-system dynamics, we use the open quantum system TEBD algorithms implemented in OQUPy [87] and Quimb to simulate a 50-qubit spin chain. The system Lindbladian equation is shown in Eq. (10). We conduct the TEBD with time step $\Delta t = 10^{-3}$, using the first-order Suzuki-Trotter decomposition [88]. During the TEBD propagation, we compress the bond dimensions by singular value decomposition, and discard all singular values smaller than a relative threshold $\epsilon_{\text{rel}} = 10^{-7}$ with respect to the largest singular value.

3. Model of local tomography errors

In the noisy setting, we expand the exact marginal Choi state in the Pauli basis

$$J_{B_{k-1}B_kB_{k+1}} = \sum_{\mu_{k-1}, \mu_k, \mu_{k+1}=0}^3 \sum_{\nu_{k-1}, \nu_k, \nu_{k+1}=0}^3 c_{\mu, \nu} (\sigma_{\mu_{k-1}}^{\text{in}} \otimes \sigma_{\mu_k}^{\text{in}} \otimes \sigma_{\mu_{k+1}}^{\text{in}}) \otimes (\sigma_{\nu_{k-1}}^{\text{out}} \otimes \sigma_{\nu_k}^{\text{out}} \otimes \sigma_{\nu_{k+1}}^{\text{out}}), \quad (\text{F3})$$

$$c_{\mu, \nu} = \frac{1}{2^6} \text{Tr}[(\sigma_{\mu_{k-1}}^{\text{in}} \otimes \sigma_{\mu_k}^{\text{in}} \otimes \sigma_{\mu_{k+1}}^{\text{in}} \otimes \sigma_{\nu_{k-1}}^{\text{out}} \otimes \sigma_{\nu_k}^{\text{out}} \otimes \sigma_{\nu_{k+1}}^{\text{out}}) J_{B_{k-1}B_kB_{k+1}}], \quad (\text{F4})$$

where the factor 2^6 follows from the orthogonality relation $\text{Tr}(P_\alpha P_\beta) = 2^6 \delta_{\alpha\beta}$ for Pauli strings on six qubits. To model noise, we perturb each Pauli coefficient by additive i.i.d. Gaussian noise $\xi_\alpha \sim \mathcal{N}(0, \sigma^2)$, i.e., $\tilde{c}_\alpha = c_\alpha + \xi_\alpha$, and form $\tilde{J}_{B_{k-1}B_kB_{k+1}} = \sum_\alpha \tilde{c}_\alpha P_\alpha$. We then project $\tilde{J}_{B_{k-1}B_kB_{k+1}}$ onto the set of physical states by imposing hermiticity, setting negative eigenvalues to zero, and renormalizing to unit trace.

4. Exact evaluation of entropies for Clifford circuits with Pauli noise

For a Clifford circuit with Pauli noise as in Appendix E 3, the noiseless Choi state $|J_0\rangle = (\mathbb{I}_{\text{in}} \otimes U_{\text{Cliff}})|\Phi^{(n)}\rangle$ is a stabilizer state whose stabilizer group $\mathcal{S}_0$ has rank $2n$, and each Pauli error acts on $S \in \mathcal{S}_0$ as $P_e S P_e = \nu_e(S) S$ with $\nu_e(S) \in \{\pm 1\}$. The entropies entering the CMI can then be computed exactly as follows. For any region R of physical sites, let R_{Choi} denote the corresponding $2|R|$ Choi qubits and $J_R = \text{tr}_{[n] \setminus R}[J(\Lambda)]$ the corresponding reduced Choi state. We further define

$$\mathcal{S}_R = \{S \in \mathcal{S}_0 : \text{supp}(S) \subseteq R_{\text{Choi}}\} \quad (\text{F5})$$

as the subgroup of stabilizers fully supported inside R_{Choi} , of rank k_R with independent generators $\{g_j^R\}_{j=1}^{k_R}$. The Pauli noise restricted to these local generators defines a binary

syndrome $s_R(e) \in \mathbb{F}_2^{k_R}$ via $\nu_e(g_j^R) = (-1)^{s_R(e)_j}$. The entropy of the reduced Choi state J_R is then given by

$$S(J_R) = |R_{\text{Choi}}| - k_R + H(s_R), \quad (\text{F6})$$

where $H(s_R) = -\sum_s q_R(s) \log_2 q_R(s)$ is the Shannon entropy of the syndrome distribution

$$q_R(s) = \sum_{e: s_R(e)=s} p_e. \quad (\text{F7})$$

This formula follows from the stabilizer subsystem-entropy formula of Ref. [83] applied to each syndrome sector. An equivalent syndrome-entropy formula for stabilizer states under Pauli noise was derived in Ref. [84]. In the standard binary-symplectic representation of the Pauli group [89], $\nu_e(S)$ is given by the symplectic inner product of the binary vectors representing P_e and S , so the syndrome is a linear map over $\mathbb{F}_2$:

$$s_R = M_R e \pmod{2}, \quad (\text{F8})$$

where e collects the Pauli error bits of the noise model and M_R is the $k_R \times L$ binary matrix whose columns are the syndromes of elementary errors. Computing $H(s_R) = H(M_R e)$ is thus a purely classical problem. The local stabilizer rank k_R is obtained by binary Gaussian elimination at a cost of $O(n^3)$ per region.

For the brickwork circuits with depolarizing noise $\mathcal{E}_p$ of Appendix E 3, this leads to the following explicit procedure. For a fixed circuit realization U , commuting all noise insertions to the output writes the Choi state of the noisy channel as

$$\rho_U = \sum_E p(E) Q_E |\psi_0\rangle \langle \psi_0| Q_E^\dagger, \quad (\text{F9})$$

with $|\psi_0\rangle$ the Choi state of the noiseless circuit and Q_E Pauli operators [89].

The entropy of a window X , comprising $|X_{\text{Choi}}|$ Choi qubits, is evaluated in four steps. First, binary Gaussian elimination yields the local subgroup $\mathcal{S}_X$ and its rank k_X . Second, we compute the expectation value of every element $g \in \mathcal{S}_X$. Conjugating g backward through the Clifford circuit produces a single Pauli string at each noise layer, and at each of the $K(g)$ spacetime locations where this string is nontrivial the depolarizing channel flips the sign of g independently with probability $2p/3$, because two of the three jump operators anticommute with any nontrivial Pauli. The expectation value is therefore damped as

$$\langle g \rangle_{\rho_U} = \left(1 - \frac{4p}{3}\right)^{K(g)}, \quad (\text{F10})$$

where the spacetime support $K(g)$ is the sum over noise layers of the spatial support of the backward-conjugated operator. Equation (F10) is the Clifford specialization of the Pauli-path damping that underlies classical simulations of noisy circuits [90]. Third, the expectation values determine the syndrome distribution, being its $\mathbb{F}_2$ -Fourier coefficients: $\langle g \rangle_{\rho_U} = \sum_s q(s) (-1)^{s(g)}$, where $s(g) = \sum_j a_j s_j \bmod 2$ for $g = \prod_j (g_j)^{a_j}$ in terms of the generators g_j of $\mathcal{S}_X$, so q is recovered by the inverse transform of Walsh-Hadamard type. Fourth, substituting q into Eq. (F6) gives the entropy in fully explicit form,

$$S(\rho_{U,X}) = |X_{\text{Choi}}| - k_X - \sum_{s \in \mathbb{F}_2^{k_X}} q(s) \log_2 q(s),$$

$$q(s) = \frac{1}{2^{k_X}} \sum_{g \in \mathcal{S}_X} (-1)^{s(g)} \left(1 - \frac{4p}{3}\right)^{K(g)}. \quad (\text{F11})$$

The CMI is obtained from the entropies of the four regions AB , BC , B , and ABC . Since elements of $\mathcal{S}_X$ with disjoint spacetime supports have additive K , the enumeration factorizes over connected components and remains tractable for all windows considered.

dynamics and entanglement generation with hundreds of trapped ions, *Science* **352**, 1297 (2016).

- [2] S. Ebadi, T.T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W.W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M.D. Lukin, *Quantum phases of matter on a 256-atom programmable quantum simulator*, *Nature (London)* **595**, 227 (2021).
- [3] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, *Evidence for the utility of quantum computing before fault tolerance*, *Nature (London)* **618**, 500 (2023).
- [4] S.-A. Guo, Y.-K. Wu, J. Ye, L. Zhang, W.-Q. Lian, R. Yao, Y. Wang, R.-Y. Yan, Y.-J. Yi, Y.-L. Xu, B.-W. Li, Y.-H. Hou, Y.-Z. Xu, W.-X. Guo, C. Zhang, B.-X. Qi, Z.-C. Zhou, L. He, and L.-M. Duan, *A site-resolved two-dimensional quantum simulator with hundreds of trapped ions*, *Nature (London)* **630**, 613 (2024).
- [5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, *Quantum supremacy using a programmable superconducting processor*, *Nature (London)* **574**, 505 (2019).
- [6] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, *Quantum computational advantage using photons*, *Science* **370**, 1460 (2020).
- [7] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan *et al.*, *Strong quantum computational advantage using a superconducting quantum processor*, *arXiv:2106.14734*.
- [8] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins *et al.*, *Quantum computational advantage with a programmable photonic processor*, *Nature (London)* **606**, 75 (2022).
- [9] E. Magesan, J. M. Gambetta, and J. Emerson, *Scalable and robust randomized benchmarking of quantum processes*, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [10] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, *Characterization of addressability by simultaneous randomized benchmarking*, *Phys. Rev. Lett.* **109**, 240504 (2012).
- [11] D. C. McKay, A. W. Cross, C. J. Wood, and J. M. Gambetta, *Correlated randomized benchmarking*, *arXiv:2003.02354*.
- [12] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, *Detecting crosstalk errors in quantum information processors*, *Quantum* **4**, 321 (2020).
- [13] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, *General framework for randomized benchmarking*, *PRX Quantum* **3**, 020357 (2022).
- [14] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A. H. Werner, J. Eisert, and I. Roth, *Shadow estimation of gate-set properties from random sequences*, *Nat. Commun.* **14**, 5039 (2023).
- [15] H.-Y. Huang, S. Chen, and J. Preskill, *Learning to predict arbitrary quantum processes*, *PRX Quantum* **4**, 040337 (2023).

[1] J. G. Bohnet, B. C. Sawyer, J. W. Britton, M. L. Wall, A. M. Rey, M. Foss-Feig, and J. J. Bollinger, *Quantum spin*

- [16] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, *Shadow process tomography of quantum channels*, *Phys. Rev. A* **107**, 042403 (2023).
- [17] R. Levy, D. Luo, and B. K. Clark, *Classical shadows for quantum process tomography on near-term quantum computers*, *Phys. Rev. Res.* **6**, 013029 (2024).
- [18] Y. Wang, G. Liu, Z. Liu, Y. Tang, X. Ma, and H. Dai, *Robust estimation of nonlinear properties of quantum processes*, *Phys. Rev. A* **110**, 032415 (2024).
- [19] K. Kato and F. G. Brandao, *Quantum approximate Markov chains are thermal*, *Commun. Math. Phys.* **370**, 117 (2019).
- [20] T. Kuwahara, *Clustering of conditional mutual information and quantum Markov structure at arbitrary temperatures*, *Phys. Rev. X* **15**, 041010 (2025).
- [21] C.-F. Chen and C. Rouzé, *Quantum gibbs states are locally Markovian*, [arXiv:2504.02208](https://arxiv.org/abs/2504.02208).
- [22] F. Verstraete, J. J. Garcia-Ripoll, and J. I. Cirac, *Matrix product density operators: Simulation of finite-temperature and dissipative systems*, *Phys. Rev. Lett.* **93**, 207204 (2004).
- [23] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, *Matrix product states and projected entangled pair states: Concepts, symmetries, theorems*, *Rev. Mod. Phys.* **93**, 045003 (2021).
- [24] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, *Matrix product operator representations*, *New J. Phys.* **12**, 025012 (2010).
- [25] A. Jamiolkowski, *Linear transformations which preserve trace and positive semidefiniteness of operators*, *Rep. Math. Phys.* **3**, 275 (1972).
- [26] I. L. Chuang and M. A. Nielsen, *Prescription for experimental determination of the dynamics of a quantum black box*, *J. Mod. Opt.* **44**, 2455 (1997).
- [27] J. F. Poyatos, J. I. Cirac, and P. Zoller, *Complete characterization of a quantum process: The two-bit quantum gate*, *Phys. Rev. Lett.* **78**, 390 (1997).
- [28] R. Harper, S. T. Flammia, and J. J. Wallman, *Efficient learning of quantum noise*, *Nat. Phys.* **16**, 1184 (2020).
- [29] J. Haah, R. Kothari, R. O'Donnell, and E. Tang, *Query-optimal estimation of unitary channels in diamond distance*, in *Proceedings of the 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)* (2023), pp. 363–390.
- [30] A. Oufkir, *Sample-optimal quantum process tomography with non-adaptive incoherent measurements*, in *Proceedings of the 2023 IEEE International Symposium on Information Theory (ISIT)* (2023).
- [31] A. A. Mele and L. Bittel, *Optimal learning of quantum channels in diamond distance*, [arXiv:2512.10214](https://arxiv.org/abs/2512.10214).
- [32] K. Chen, N. Yu, and Z. Zhang, *Quantum channel tomography and estimation by local test*, [arXiv:2512.13614](https://arxiv.org/abs/2512.13614).
- [33] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, *Gate set tomography*, *Quantum* **5**, 557 (2021).
- [34] H.-Y. Huang, Y. Liu, M. Broughton, I. Kim, A. Anshu, Z. Landau, and J. R. McClean, *Learning shallow quantum circuits*, in *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (2024), pp. 1343–1351.
- [35] M. Crupi, J. I. Cirac, and F. Baccari, *Efficient characterization of coherent and correlated noise in layers of gates*, *PRX Quantum* **6**, 040374 (2025).
- [36] D. Stilck França, L. A. Markovich, V. V. Dobrovitski, A. H. Werner, and J. Borregaard, *Efficient and robust estimation of many-qubit Hamiltonians*, *Nat. Commun.* **15**, 311 (2024).
- [37] T. Olsacher, T. Kraft, C. Kokail, B. Kraus, and P. Zoller, *Hamiltonian and Liouvillian learning in weakly-dissipative quantum many-body systems*, *Quantum Sci. Technol.* **10**, 015065 (2025).
- [38] G. Cemin, F. Carnazza, S. Andergassen, G. Martius, F. Carollo, and I. Lesanovsky, *Inferring interpretable dynamical generators of local quantum observables from projective measurements through machine learning*, *Phys. Rev. Appl.* **21**, L041001 (2024).
- [39] G. Cemin, M. Cech, E. Weiss, S. Soltan, D. Braun, I. Lesanovsky, and F. Carollo, *Machine learning of reduced quantum channels on noisy intermediate-scale quantum devices*, *Phys. Rev. A* **110**, 052418 (2024).
- [40] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, *Scalable reconstruction of density matrices*, *Phys. Rev. Lett.* **111**, 020401 (2013).
- [41] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, *Efficient tomography of a quantum many-body system*, *Nat. Phys.* **13**, 1158 (2017).
- [42] M. Holzäpfel, M. Cramer, N. Datta, and M. B. Plenio, *Petz recovery versus matrix reconstruction*, *J. Math. Phys.* **59**, 042201 (2018).
- [43] G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita, *Quantum process tomography with unsupervised learning and tensor networks*, *Nat. Commun.* **14**, 2858 (2023).
- [44] S. Mangini, M. Cattaneo, D. Cavalcanti, S. Filippov, M. A. C. Rossi, and G. García-Pérez, *Tensor network noise characterization for near-term quantum computers*, *Phys. Rev. Res.* **6**, 033217 (2024).
- [45] Y. Guo and S. Yang, *Quantum state tomography with locally purified density operators and local measurements*, *Commun. Phys.* **7**, 322 (2024).
- [46] M. Votto, M. Ljubotina, C. Lancien, J. I. Cirac, P. Zoller, M. Serbyn, L. Piroli, and B. Vermersch, *Learning mixed quantum states in large-scale experiments*, [arXiv:2507.12550](https://arxiv.org/abs/2507.12550).
- [47] M. Fanizza, N. Galke, J. Lumberras, C. Rouzé, and A. Winter, *Learning finitely correlated states: Stability of the spectral reconstruction*, [arXiv:2312.07516](https://arxiv.org/abs/2312.07516).
- [48] A. M. Alhambra, A. Capel, P. Gondolf, A. Ruiz-de Alarcón, and S. O. Scalet, *Conditional independence of 1D gibbs states with applications to efficient learning*, [arXiv:2402.18500](https://arxiv.org/abs/2402.18500).
- [49] S. O. Scalet, A. Capel, A. N. Chowdhury, H. Fawzi, O. Fawzi, I. H. Kim, and A. Tikku, *Classical estimation of the free energy and quantum gibbs sampling from the Markov entropy decomposition*, [arXiv:2504.17405](https://arxiv.org/abs/2504.17405).
- [50] L. Gao and N. Yu, *Sample optimal tomography of quantum Markov chains*, *IEEE Trans. Inf. Theory* **71**, 7016 (2025).
- [51] M.-D. Choi, *Completely positive linear maps on complex matrices*, *Linear Algebra Appl.* **10**, 285 (1975).
- [52] Y.-K. Liu, *Consistency of local density matrices is qma-complete*, in *International Workshop on Approximation Algorithms for Combinatorial Optimization* (Springer, New York, 2006), pp. 438–449.

- [53] Y.-K. Liu, M. Christandl, and F. Verstraete, *Quantum computational complexity of the N -representability problem: Qma complete*, *Phys. Rev. Lett.* **98**, 110503 (2007).
- [54] C. Schilling, *The quantum marginal problem*, in *Mathematical Results in Quantum Mechanics* (World Scientific, Singapore, 2014), pp. 165–176.
- [55] C.-Y. Hsieh, M. Lostaglio, and A. Acín, *Quantum channel marginal problem*, *Phys. Rev. Res.* **4**, 013249 (2022).
- [56] A. Broadbent and A. B. Grilo, *Zero-knowledge for QMA from locally simulatable proofs*, [arXiv:1911.07782](https://arxiv.org/abs/1911.07782).
- [57] D. Sutter, *Approximate Quantum Markov Chains*, SpringerBriefs in Mathematical Physics (Springer International Publishing, Cham, 2018), Vol. 28.
- [58] D. Petz, *Sufficient subalgebras and the relative entropy of states of a von Neumann algebra*, *Commun. Math. Phys.* **105**, 123 (1986).
- [59] D. Petz, *Monotonicity of quantum relative entropy revisited*, *Rev. Math. Phys.* **15**, 79 (2003).
- [60] O. Fawzi and R. Renner, *Quantum conditional mutual information and approximate markov chains*, *Commun. Math. Phys.* **340**, 575 (2015).
- [61] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter, *Universal recovery maps and approximate sufficiency of quantum relative entropy*, *Ann. Henri Poincaré* **19**, 2955 (2018).
- [62] Y. Hu and Y. Zou, *Petz map recovery for long-range entangled quantum many-body states*, *Phys. Rev. B* **110**, 195107 (2024).
- [63] S. Vardhan, A. Y. Wei, and Y. Zou, *Petz recovery from subsystems in conformal field theory*, *J. High Energy Phys.* **2024**, 016 (2024).
- [64] K. M. R. Audenaert and J. Eisert, *Continuity bounds on the quantum relative entropy*, *J. Math. Phys.* **46**, 102104 (2005).
- [65] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (OUP Oxford, Oxford, England, 2002).
- [66] G. Vidal, *Efficient simulation of one-dimensional quantum many-body systems*, *Phys. Rev. Lett.* **93**, 040502 (2004).
- [67] M. Zwolak and G. Vidal, *Mixed-state dynamics in one-dimensional quantum lattice systems: A time-dependent superoperator renormalization algorithm*, *Phys. Rev. Lett.* **93**, 207205 (2004).
- [68] J. Gray, *Quimb: A python package for quantum information and many-body calculations*, *J. Open Source Software* **3**, 819 (2018).
- [69] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, England, 2004).
- [70] S. Diamond and S. Boyd, *CVXPY: A python-embedded modeling language for convex optimization*, *J. Mach. Learn. Res.* **17**, 1 (2016), <https://www.jmlr.org/papers/v17/15-408.html>.
- [71] S.-U. Lee, C. Oh, Y. Wong, S. Chen, and L. Jiang, *Universal spreading of conditional mutual information in noisy random circuits*, *Phys. Rev. Lett.* **133**, 200402 (2024).
- [72] A. Gilchrist, N. K. Langford, and M. A. Nielsen, *Distance measures to compare real and ideal quantum processes*, *Phys. Rev. A* **71**, 062310 (2005).
- [73] A. Montanaro and R. de Wolf, *A survey of quantum property testing*, [arXiv:1310.2035](https://arxiv.org/abs/1310.2035).
- [74] Z. Liu and D. S. Wild, *Code and data for “Efficiently Learning Global Quantum Channels with Local Tomography”*, Zenodo, 2026, release v1.1, <https://github.com/zidu-liu/learning-global-channel>.
- [75] D. Poulin and M. B. Hastings, *Markov entropy decomposition: A variational dual for quantum belief propagation*, *Phys. Rev. Lett.* **106**, 080403 (2011).
- [76] A. J. Ferris and D. Poulin, *Algorithms for the Markov entropy decomposition*, [arXiv:1212.1442](https://arxiv.org/abs/1212.1442).
- [77] A. Winter, *Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints*, *Commun. Math. Phys.* **347**, 291 (2016).
- [78] I. H. Kim, *On the informational completeness of local observables*, [arXiv:1405.0137](https://arxiv.org/abs/1405.0137).
- [79] I. Kim and M. B. Ruskai, *Bounds on the concavity of quantum entropy*, *J. Math. Phys.* **55**, 092201 (2014).
- [80] G. Lindblad, *Completely positive maps and entropy inequalities*, *Commun. Math. Phys.* **40**, 147 (1975).
- [81] N. J. Beaudry and R. Renner, *An intuitive proof of the data processing inequality*, [arXiv:1107.0740](https://arxiv.org/abs/1107.0740).
- [82] J. Watrous, *The Theory of Quantum Information* (Cambridge university press, Cambridge, England, 2018).
- [83] D. Fattal, T. S. Cubitt, Y. Yamamoto, S. Bravyi, and I. L. Chuang, *Entanglement in the stabilizer formalism*, [arXiv:quant-ph/0406168](https://arxiv.org/abs/quant-ph/0406168).
- [84] K. Goodenough, A. Sajjad, E. Kaur, S. Guha, and D. Towsley, *Bipartite entanglement of noisy stabilizer states through the lens of stabilizer codes*, in *Proceedings of the 2024 IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2024), pp. 545–550.
- [85] A.-R. Negari, T. D. Ellison, and T. H. Hsieh, *Spacetime Markov length: A diagnostic for fault tolerance via mixed-state phases*, [arXiv:2412.00193](https://arxiv.org/abs/2412.00193).
- [86] B. O’donoghue, E. Chu, N. Parikh, and S. Boyd, *Conic optimization via operator splitting and homogeneous self-dual embedding*, *J. Optim. Theory Appl.* **169**, 1042 (2016).
- [87] G. E. Fux, P. Fowler-Wright, J. Beckles, E. P. Butler, P. R. Eastham, D. Gribben, J. Keeling, D. Kilda, P. Kirton, E. D. C. Lawrence, B. W. Lovett, E. O’Neill, A. Strathearn, and R. de Wit, *Oqupy: A Python package to efficiently simulate non-Markovian open quantum systems with process tensors*, *J. Chem. Phys.* **161**, 124108 (2024).
- [88] M. Suzuki, *Generalized Trotter’s formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems*, *Commun. Math. Phys.* **51**, 183 (1976).
- [89] D. Gottesman, *Stabilizer codes and quantum error correction*, Ph.D. thesis, California Institute of Technology, 1997.
- [90] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, *A polynomial-time classical algorithm for noisy random circuit sampling*, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (2023), pp. 945–957.