

Matrix Product Density Operators: when do they have a local parent Hamiltonian?

Chi-Fang Chen

*Institute for Quantum Information and Matter
California Institute of Technology, Pasadena, CA 91125, USA*

Kohtaro Kato

*Institute for Quantum Information and Matter
California Institute of Technology, Pasadena, CA 91125, USA and
Center for Quantum Information and Quantum Biology,
Institute for Open and Transdisciplinary Research Initiatives,
Osaka University, Osaka 560-8531, Japan*

Fernando G.S.L. Brandão

*AWS Center for Quantum Computing, Pasadena, CA and
Institute for Quantum Information and Matter
California Institute of Technology, Pasadena, CA 91125, USA*

We study whether one can write a Matrix Product Density Operator (MPDO) as the Gibbs state of a quasi-local parent Hamiltonian. We conjecture this is the case for generic MPDO and give supporting evidences. To investigate the locality of the parent Hamiltonian, we take the approach of checking whether the quantum conditional mutual information decays exponentially. The MPDO we consider are constructed from a chain of 1-input/2-output ('Y-shaped') completely-positive maps, i.e. the MPDO have a local purification. We derive an upper bound on the conditional mutual information for bistochastic channels and strictly positive channels, and show that it decays exponentially if the correctable algebra of the channel is trivial.

We also introduce a conjecture on a quantum data processing inequality that implies the exponential decay of the conditional mutual information for every Y-shaped channel with trivial correctable algebra. We additionally investigate a close but nonequivalent cousin: MPDO measured in a local basis. We provide sufficient conditions for the exponential decay of the conditional mutual information of the measured states, and numerically confirmed they are generically true for certain random MPDO.

Contents

I. Introduction	3
II. Preliminary	5
A. Matrix Product Density Operators	6
B. Strong data-processing inequality constants	8

C. The conditional mutual information and Gibbs states	9
D. The correctable algebra and recovery channels	10
III. Main results	11
A. Decay of CMI for bistochastic Y-shaped channels	11
1. The trace norm CMI	13
B. Decay of CMI for Y-shaped channels with a forgetful component	14
C. A completely contractive DPI of CMI that implies exponential decay of CMI	15
D. Sufficient conditions for exponential decay of classically-conditioned mutual information	17
IV. Proofs of main theorems	18
A. Theorem III.1: decay of CMI for bistochastic channel	18
B. Theorem III.2: decay of trace norm CMI for partially invariant channel	20
C. Theorem III.3: decay of B -measured CMI in MPDO	22
1. $2 \implies 3$: uniformly bounding every sequence quantum operations by contraction in Hilbert's projective metric	22
2. $1 \implies 2$: uniformly bounding coarse graining length by generalizing quantum Wielandt's inequality to family of CP self-maps	23
V. Proofs for propositions around main theorem and lemmas supporting main theorems	24
A. Remaining Proof for Theorem III.1	24
1. Proof of Lemma IV.1, DPI for 2-norm	24
2. Proof of Proposition IV.2, the continuity of CMI	25
3. Proof of Lemma IV.3, spectral characterization of the contraction ratio	26
B. Proofs for propositions around Conjecture III.1	26
1. Proof of Proposition III.3	26
2. Proof of Proposition III.4	28
3. Proof of Proposition III.2	28
4. Proof of Proposition III.5	29
C. Remaining proof for Theorem III.2	29
1. Proof of Lemma IV.4	29
2. Proof of Proposition III.1	30
D. Remaining proof of Theorem III.3: decay of B -measured conditional mutual information in MPDO	30
1. The Hilbert's projective metric and proof of Lemma IV.5	30
2. Proof of Lemma IV.6	32

3. Proof of Lemma IV.7	33
4. Proof of Lemma IV.8	33
5. Proof of Proposition III.6	35
VI. Conclusion and discussions	35
VII. Acknowledgement	36
References	36
A. Equivalence between correctable algebra and the Petz recovered map for bistochastic channels	39
B. Structure of periodic MPDO generated by Y-shaped bistochastic channels.	40

I. INTRODUCTION

Tensor networks provide useful ansatz for quantum many-body systems. In one-dimensional (1D) systems, the ground states of gapped local Hamiltonians can be efficiently approximated by Matrix Product States (MPS) [1–3]. For the converse, generic MPS (which are technically called *injective* MPS) always have a gapped, local and frustration-free *parent Hamiltonian* whose unique ground state is the MPS [2, 4]. This correspondence between MPS and its parent Hamiltonian establishes a deep connection with gapped quantum systems, leading to the complete classification of 1D gapped quantum phases [5]. For higher-dimensional systems, Projected Entangled Pair States (PEPS) are a natural generalization of MPS. PEPS have been used successfully to study gapped ground states [1, 6]. Although the structural characterization of PEPS has not been completely established, local parent Hamiltonians also exist for injective and semi-injective PEPS [7].

Matrix Product Density Operators (MPDO) are generalization of MPS to describe 1D *mixed* states. In Ref. [8], Hastings showed that any 1D Gibbs state of a local Hamiltonian (local Gibbs state in short) can be well-approximated by a MPDO with polynomial bond dimension. This result justifies MPDO as a successful ansatz to study Gibbs states [9].

As a generic MPS is the ground state of the parent Hamiltonian, one can expect that a generic MPDO could be written as the Gibbs state of a local parent Hamiltonian. In fact, Cirac et al. analyzed MPDO at certain “renormalization fixed-points”, and showed that these fixed-point MPDO are Gibbs states of nearest-neighbor commuting Hamiltonians [10]. Unfortunately, unlike MPS, a renormalization operation transforming a given MPDO to these fixed-point MPDO has not been well-defined yet. Therefore, a (generic) condition under which MPDO has a local parent Hamiltonian is yet to be established.

In this paper, we investigate conditions when a MPDO can be associated with a local parent Hamiltonian. Our technical approach is to study whether the *conditional mutual information (CMI)* decays exponentially. The CMI $I(A : C|B)_\rho$ is a function defined for a tripartite state ρ_{ABC} as

$$I(A : C|B)_\rho := S(AB)_\rho + S(BC)_\rho - S(B)_\rho - S(ABC)_\rho, \quad (1)$$

where $S(A)_\rho = -\text{Tr} \rho_A \log_2 \rho_A$ is the von Neumann entropy of the reduced state on A . Small values of CMI (with respect to certain tri-partitions of 1D system) turns out to be the necessary and sufficient condition for being well-approximated by a local Gibbs state [11] (see Sec. II C for more details). Motivated by this result, the overarching theme of this paper is the following conjecture, with a deliberate ambiguity left for defining ‘generic’:

Conjecture I.1. *For ‘generic’ MPDO, the CMI decays exponentially: there exists a constant $c > 0$ such that for any tri-partition ABC of the system where B separates A from C with distance ℓ ,*

$$I(A : C|B) = O(e^{-c\ell}). \quad (2)$$

If this conjecture is true, then parent Hamiltonians of generic MPDO are (quasi-)local. In addition, we investigate the CMI of the MPDO after local measurements on the conditioning system B . As the CMI does not always decrease under measurement, it must be studied as an independent problem. The exponential decay of the CMI after measurement implies the outcome distribution is nearly a classical Markov distribution.

Main result. We first study the quantum CMI of each tripartition of states generated by 1-input/2-output, “Y-shaped” in short, channels. In particular, for bistochastic Y-shaped channels, we obtain an analytical bound for the CMI with a decay rate constant (Theorem III.1). The decay rate constant is strictly smaller than one if the channel has trivial correctable algebra, as defined in operator-algebra quantum error correction theory. We then generalize the argument for a slightly larger class of channels and derive a weaker bound on the CMI with another decay rate constant (Theorem III.2). Therefore we show the exponential decay of the CMI (and a trace norm analog of CMI) for these MPDO. We also prove the exponential decay of CMI for Y-shaped channels with a forgetful component which, in particular, include every strictly positive Y-shaped channels (Proposition III.2).

For general Y-shaped channels, we show that the CMI $I(A : C|B)$ must strictly decrease if one applies a channel with trivial correctable algebra on C (Proposition III.3). To bridge this result to Conjecture I.1, we proposed a conjecture in the form of a data processing inequality for CMI with an explicit decay rate (Conjecture III.1), which would imply exponential decay of CMI (Proposition III.4) for MPDO generated by such Y-shaped channels. The conjecture is known to be true for classical systems, however no result is known for quantum systems.

We further study the CMI of MPDO after local measurements on the conditional system B in the computational basis. The family of MPDO we consider here are those generated by completely positive (CP) Y-shaped maps¹. We prove two sufficient conditions, one is stronger than the other, that guarantee the exponential decay of the CMI of the measured MPDO (Theorem III.3). We then provide a simple polynomial algorithm to check the stronger condition (Proposition III.6), and we numerically find the condition generically holds for MPDO generated by Y-shaped channels whose Stinespring unitaries are sampled from the Haar measure.

Proof ideas. The proof for decay of CMI for bistochastic channel (Theorem III.1) relies on decomposing the state into an uncorrelated state on AB and C plus a deviation which is traceless on C . We show the Hilbert-Schmidt norm of the deviation contracts under a bistochastic channel. In Theorem III.2, we instead consider contraction of the deviation under the trace norm, using the associated tools for trace norm. For Y-shaped channels with a forgetful component (Proposition III.2), we use relative entropy convexity to show the CMI contracts at each step. The strict decay of the CMI for general channels (Proposition III.3) follows from techniques in operator-algebra quantum error correction and properties of the Petz recovery map.

For the measured MPDO (Theorem III.3) we are conditioning on a classical system \bar{B} , which implies the CMI $I(A : C | \bar{B})$ equals to the average of the mutual information for each outcome state $\rho_{AC, \bar{B}=b}$. The outcome states are constructed by sequential CP self-maps, and these are contractions in the *Hilbert's projective metric*. The CMI decays exponentially if the contraction ratio of every CP-map is strictly less than 1, which holds if the maps are all strictly positive (Condition 2). We also provide a stronger condition, Condition 1, which guarantees strict positivity after certain coarse-graining.

Structure of the paper. We introduce basic concepts on MPDO and few backgrounds in Section II. In Section III, we state our results and a conjecture on the data-processing inequality. The proofs are presented in Section IV with several lemmas, whose detailed proof is left at Section V.

II. PRELIMINARY

In this section, we introduce basic notations and quantum information theoretical concepts that will be used in this paper. A quantum state (density operator) ρ is a bounded operator on a finite-dimensional Hilbert space \mathcal{H} satisfying positivity $\rho \geq 0$ and unit trace $\text{Tr}\rho = 1$. We denote the set of quantum states on \mathcal{H} by $\mathcal{S}(\mathcal{H})$. Quantum systems are often denoted by capital letters A, B, C, \dots and we abuse the same notation for the associated Hilbert spaces and the bounded operator spaces. We denote the completely

¹ For technical reasons, the proof works for CP maps without the trace preserving (TP) constraint, whereas the unmeasured case it has to be a channel (CPTP-map).

mixed state on a Hilbert space B by τ_B . The reduced state of ρ associated to the system A is denoted by ρ_A . $\|\cdot\|$ is the operator norm, $\|\cdot\|_p$ is the p -Schatten norm, $\|\cdot\|_{p-q}$ is the $p-q$ superoperator norm, and $\|\cdot\|_{cb}$ is the completely bounded 1-1 norm.

A. Matrix Product Density Operators

A general open boundary (uniform) MPDO $\rho \in \mathcal{S}(\mathcal{H}_n)$ is a quantum state ρ written as

$$\rho = \sum_{i_1, j_1, \dots, i_n, j_n} \langle L | A^{i_1, j_1} A^{i_2, j_2} \dots A^{i_n, j_n} | R \rangle \times |i_1 i_2 \dots i_n\rangle \langle j_1 j_2 \dots j_n|, \quad (3)$$

where each $\{A^{ij}\}_{i,j}$ is a set of D -dimensional matrices and $|L\rangle, |R\rangle$ are D -dimensional vectors with a constant D . Here, $A^{ij}, |L\rangle$ and $|R\rangle$ are chosen so that positivity $\rho \geq 0$ is satisfied for arbitrary n . This is a non-trivial condition and in general it is computationally hard to determine whether a given MPO is positive or not [12]. In this paper, we will specialize to specific sub-classes of MPDO to guarantee the positivity of the state.

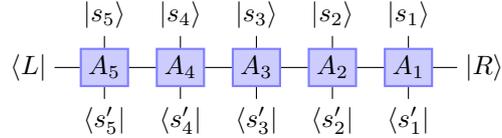


FIG. 1: The tensor network representation for general MPDO. It differs from MPS that both bra and ket indices are distinguished, describing an operator instead of a vector.

Throughout this paper, we consider MPDO constructed by concatenating Completely-Positive (CP) maps, i.e. MPDO with local purification. Let \mathcal{H}, \mathcal{K} be finite-dimensional Hilbert spaces and $\mathcal{N} : \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{K} \otimes \mathcal{H})$ be a 1-input/2-output CP-map where one of the two outputs system has the same dimension as the input. We interchangeably refer such a map as a *Y-shaped* map. For some initial state σ and $\ell \in \mathbb{N}$, we obtain a (possibly unnormalized) state $\rho_\ell^\sigma(\mathcal{N}) \in \mathcal{S}(\mathcal{K}^{\otimes \ell} \otimes \mathcal{H})$ defined by

$$\rho_\ell^\sigma(\mathcal{N}) := (\text{id}_{\mathcal{K}^{\otimes (\ell-1)}} \otimes \mathcal{N}) \circ \dots \circ (\text{id}_{\mathcal{K}} \otimes \mathcal{N}) \circ \mathcal{N}[\sigma], \quad (4)$$

where $\text{id}_{\mathcal{K}}$ is the identity map on \mathcal{K} . We mostly abbreviate $\rho_\ell^\sigma(\mathcal{N})$ as ρ_ℓ^σ . This kind of states are classified as a finite-dimensional C^* -finitely correlated states [4]². By using the Kraus representation $\mathcal{N}(X) = \sum_i T_i X T_i^\dagger$, one can easily verify ρ_ℓ^σ is a MPDO.

The normalization $\text{Tr} \rho_\ell^\sigma = 1$ is guaranteed when σ is a quantum state and \mathcal{N} is a quantum channel, i.e., CP and Trace-Preserving (CPTP) map. Since 1D local Gibbs states always have exponentially decaying

² Eq. (4) contains states without the consistency constraint imposed on the input state $\text{Tr}_B \rho_{BC}^\sigma = \sigma$, which is required for C^* -infinite finitely correlated states.

two-point correlation [13], we are interested in Eq. (4) with finite correlation length. The MPDO has a finite correlation length if $\text{Tr}_{\mathcal{K}} \circ \mathcal{N}$ has unique maximum eigenvalue 1, and especially the correlation length is exactly zero when $\text{Tr}_{\mathcal{K}} \circ \mathcal{N}$ is a constant channel (Fig. 2). The choice of σ is rather arbitrary in finite systems. For convenience we mainly choose σ as one side of the maximally entangled state, and denote the corresponding state on $\mathcal{H} \otimes \mathcal{K}^{\otimes \ell} \otimes \mathcal{H}$ by ρ_{ℓ} . We often denote A and C to be two systems \mathcal{H} at the end, and B_i ($i = 1, \dots, \ell$) to be the rest systems with \mathcal{K} . Note that we can recover arbitrary ρ_{ℓ}^{σ} from ρ_{ℓ} by applying a suitable positive operator on A , which is the other side of the maximally entangled state $\sigma_{A\bar{A}}$, and then trace out A .

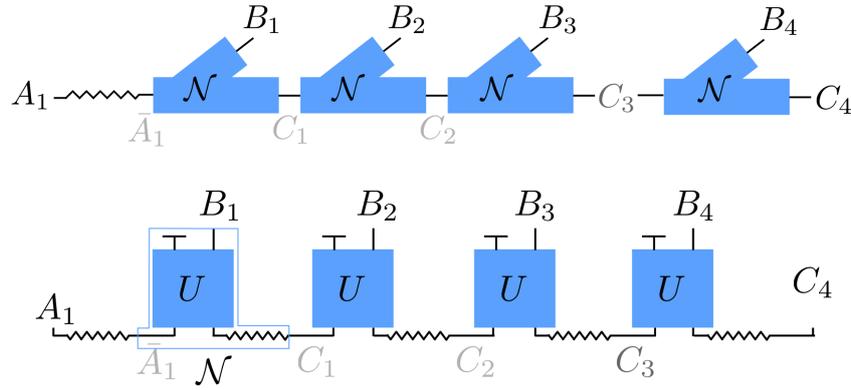


FIG. 2: (Top) The chain of Y-shaped channel \mathcal{N} . The zigzag line represents a maximally entangled pair. When we concatenate another channel to system $AB_1B_2B_3C_3$, subsystem C_3 is mapped to system B_4C_4 . (Bottom) Y-shaped channel with zero correlation length can be generated by unitaries or isometries and maximally entangled pairs, where tracing out any system B_k makes the left and right segment uncorrelated.

When we further perform a projective measurement on B to Eq. (4) in the computational basis $\{|s_1 s_2 \dots s_l\rangle_B\}$, the output subsystem \bar{B} becomes entirely classical (Fig. 3). The corresponding Y-shaped map \mathcal{N} can then be decomposed into $\mathcal{N}[\cdot] = \sum_{s_k} |s_k\rangle\langle s_k| \otimes \mathcal{M}_{s_k}[\cdot]$, where $\mathcal{M}_s[\cdot]$ is a CP self-map defined as

$$\mathcal{M}_{s_k}[\rho] := \langle s_k |_{B_k} (\mathcal{N}[\rho]) |s_k\rangle_{B_k}. \quad (5)$$

Note that if \mathcal{N} is TP, then $\sum_s \mathcal{M}_s$ is a CPTP-map and thus $\{\mathcal{M}_s\}_s$ forms a quantum instrument.

We can rewrite the measured MPDO by

$$\rho_{A\bar{B}C} = \sum_b p(b) |b\rangle\langle b|_{\bar{B}} \otimes \rho_{AC,b}, \quad (6)$$

where each $\rho_{AC,b}$ is an output state with a particular measurement outcome $b = \{s_1, \dots, s_l\}$

$$\rho_{AC,b} = \frac{M_b[\sigma_{A\bar{A}}]}{\text{Tr}(M_b[\sigma_{A\bar{A}}])} := \frac{\mathcal{M}_{s_\ell} \cdots \mathcal{M}_{s_1}[\sigma_{A\bar{A}}]}{\text{Tr}(\mathcal{M}_{s_\ell} \cdots \mathcal{M}_{s_1}[\sigma_{A\bar{A}}])}, \quad (7)$$

with probability

$$p(b) = \text{Tr}(\mathcal{M}_{s_\ell} \cdots \mathcal{M}_{s_1}[\sigma_{A\bar{A}}]). \quad (8)$$

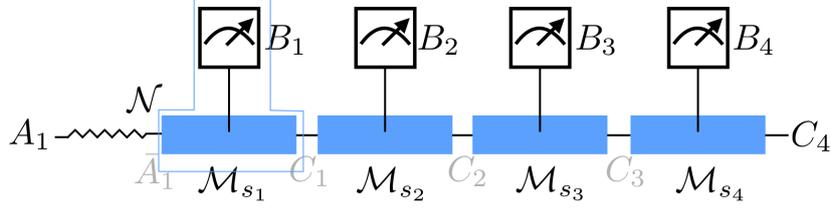


FIG. 3: The chain of Y-shaped CP maps reduced to sequences of CP self-maps upon measurement outcomes on B_1, \dots, B_4 . It mimics a classical hidden Markov chain, with the extra bit of the quantumness confined in system A and C .

B. Strong data-processing inequality constants

The quantum relative entropy $D(\rho\|\sigma) := \text{Tr}\rho \log_2(\rho - \sigma)$ is a distance-like measure between two quantum states ρ, σ . One crucial feature of the relative entropy is that it obeys the *data-processing inequality (DPI)*: for any states ρ, σ and any CPTP-map \mathcal{E} , it follows that [14, 15]

$$D(\mathcal{E}[\rho]\|\mathcal{E}[\sigma]) \leq D(\rho\|\sigma). \quad (9)$$

The above DPI implies the monotonicity of the mutual information $I(A : C)_\rho := D(\rho_{AC}\|\rho_A \otimes \rho_C)$ and the conditional mutual information $I(A : C|B)_\rho$ under a local CPTP-map $\mathcal{E}_C : C \rightarrow C'$, that is,

$$I(A : C')_{\mathcal{E}_C(\rho_{AC})} \leq I(A : C)_{\rho_{AC}}, \quad (10)$$

$$I(A : C'|B)_{\mathcal{E}_C(\rho_{ABC})} \leq I(A : C|B)_{\rho_{ABC}}. \quad (11)$$

The equality holds if and only if there is a CPTP-map $\mathcal{R}_{C'} : C' \rightarrow C$ such that $\mathcal{R} \circ \mathcal{E}[\rho] = \rho$ [16, 17].

Strong DPI constants, or the *contraction coefficients* are multiplicative factors in the above monotonicity inequalities. For the mutual information and CMI, they are defined as

$$\eta_{AC}(\mathcal{E}_C) := \sup_{\rho_{AC}} \frac{I(A : C')_{\mathcal{E}_C(\rho_{AC})}}{I(A : C)_{\rho_{AC}}}, \quad (12)$$

$$\eta_{ABC}(\mathcal{E}_C) := \sup_{\rho_{ABC}} \frac{I(A : C'|B)_{\mathcal{E}_C(\rho_{ABC})}}{I(A : C|B)_{\rho_{ABC}}}, \quad (13)$$

where the values are bounded as $0 \leq \eta_{AC}(\mathcal{E}_C), \eta_{ABC}(\mathcal{E}_C) \leq 1$ by monotonicity. Note that $\eta_{ABC}(\mathcal{E}_C)$ reduces to $\eta_{AC}(\mathcal{E}_C)$ by taking B to be trivial. The strong DPI constants bound how much correlations are preserved after applying the channel on subsystem C .

In the classical case, the strong DPI constants for different quantities are equivalent.

Theorem II.1 (Tight contractive DPI for classical mutual information [18, Theorem 4]). *For any probability distribution p_{AC} and any classical channel $\mathcal{E} : C \rightarrow C'$,*

$$\eta_{AC}(\mathcal{E}) = \eta_{ABC}(\mathcal{E}) = s^*(\mathcal{E}), \quad (14)$$

where $s^*(\mathcal{E})$ is the strong DPI constant for the relative entropy on C

$$s^*(\mathcal{E}) := \sup_{\substack{p,q \\ p \neq q}} \frac{D(\mathcal{E}[q_C] \parallel \mathcal{E}[p_C])}{D(p_C \parallel q_C)}. \quad (15)$$

Remarkably, $\eta_{AC}(\mathcal{E})$ and $\eta_{ABC}(\mathcal{E})$ are independent of A, B in the classical case. The crucial point is that the classical mutual information and the CMI are functions of conditional probability distribution on C . These functions are then written by convex combinations in classical systems, and regardless of how large auxiliary classical system is involved, the Carathéodory theorem implies the auxiliary system can always be reduced to dimension $|C| + 1$ [19].

Unfortunately, no analog of conditional distribution nor the Carathéodory-like cardinality bound has been found for quantum systems (see e.g., Ref. [20] for a discussion on this problem). Therefore the classical approach failed to work in quantum regime. For this reason, for quantum systems it remains open whether Theorem II.1 also holds, or whether the strong DPI constants are independent of the size of the auxiliary systems or not.

C. The conditional mutual information and Gibbs states

The CMI and Gibbs states are intimately connected. In classical systems, the Hammersley-Clifford theorem [21] states that a (positive) probability distribution p_{XYZ} is a Gibbs distribution

$$p_{XYZ}(x, y, z) = \frac{e^{-h_{XY}(x,y) - h_{YZ}(y,z)}}{Z}, \quad (16)$$

where Z is the normalization constant, if and only if $I(X : Z|Y)_p = 0$. Moreover, CMI can be written as

$$I(X : Z|Y)_p = \min_{q: I(X:Z|Y)_q=0} D(p_{XYZ} \parallel q_{XYZ}), \quad (17)$$

and thus the small value of the CMI implies the state is close to a Gibbs state (16). These results are naturally extended to 1D classical spin chains.

Although Eq. (17) does not hold in general quantum systems, Ref. [11] shows the following bound:

Theorem II.2 (Theorem 1, [11]). *Let $\rho_{A_1 \dots A_n}$ be a state satisfying $I(A_1 \dots A_{k-1} : A_k \dots A_n | A_k) \leq \varepsilon$. Then there exists a local Hamiltonian $H = \sum_{i=1} h_{A_i, A_{i+1}}$ with $h_{A_i, A_{i+1}}$ only acts on $A_i A_{i+1}$, such that*

$$S\left(\rho \parallel \frac{e^{-H}}{\text{Tr} e^{-H}}\right) \leq n\varepsilon. \quad (18)$$

If $\varepsilon = 0$, we recover the quantum Hammersley-Clifford theorem [22, 23] and the Hamiltonian H is the sum of terms like $-\log \rho_{A_i, A_{i+1}}$.

When MPDO has exponentially decaying CMI for appropriate tripartitions, then we first coarse-grain sufficiently many (possibly logarithmic w.r.t. $1/\varepsilon$) neighbouring sites as one site and then apply Theorem II.2 to show that the MPDO is well-approximated by a local Gibbs state.

D. The correctable algebra and recovery channels

To characterize what information is left undisturbed by a channel, we use the theory of operator-algebra quantum-error correction [24, 25]. For a given channel $\mathcal{E} : A \rightarrow A'$, we define the *correctable algebra* $\mathcal{A}(\mathcal{E}) \subset \mathcal{B}(\mathcal{H}_A)$ as

$$\mathcal{A}(\mathcal{E}) := \text{Alg} \left\{ O_A \left[\left[O_A, E_a^\dagger E_b \right] = 0, \quad \forall a, b \right] \right\}. \quad (19)$$

$\mathcal{A}(\mathcal{E})$ is a finite-dimensional C^* -algebra containing all observables whose information is perfectly preserved under \mathcal{E} . One can always perfectly recover the information associated to an observable in the correctable algebra: there exists a CPTP-map $\mathcal{R} : A' \rightarrow A$ called *recovery map* such that

$$\mathcal{E}^\dagger \circ \mathcal{R}^\dagger(O_A) = O_A, \quad \forall O_A \in \mathcal{A}(\mathcal{E}). \quad (20)$$

This implies that for any $O_A \in \mathcal{A}(\mathcal{E})$, there always is a corresponding observable $\mathcal{R}^\dagger(O_A)$ on A' such that $\text{Tr}(O_A \rho_A) = \text{Tr}(\mathcal{R}^\dagger(O_A) \mathcal{E}(\rho_A))$ for all ρ_A . The condition in Eq. (19) is a generalization of the well-known Knill-Laflamme condition [26] in the standard quantum error correction theory, in which $\mathcal{A}(\mathcal{E}) = \mathcal{B}(\mathcal{H}_{code})$ for a protected code subspace \mathcal{H}_{code} .

A common candidate for the recovery channel is the Petz map:

Definition 1. *The Petz recovery map $\mathcal{P}_{\sigma, \mathcal{E}}$ with the reference state σ and the target channel \mathcal{E} is defined as*

$$\mathcal{P}_{\sigma, \mathcal{E}}[\cdot] := \sigma^{1/2} \mathcal{E}^\dagger \left[\mathcal{E}(\sigma)^{-1/2} (\cdot) \mathcal{E}(\sigma)^{-1/2} \right] \sigma^{1/2}. \quad (21)$$

As it only features moderately in this paper, we save the vast literature on universal recovery channel elsewhere (see e.g., [27]), and include the relevant ideas on invariant subspaces in Appendix A. This yields another characterization of the correctable algebra handy for our purposes.

Proposition II.1. *The correctable algebra of a channel \mathcal{E} equals to the fixed-point algebra of the Petz-recovery channel composed with channel $\mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}$ (and the dual $\mathcal{E}^\dagger \circ \mathcal{P}_{\tau, \mathcal{E}}^\dagger$, due to being self-adjoint).*

$$S_{\mathcal{E}^\dagger \circ \mathcal{P}_{\tau, \mathcal{E}}^\dagger} = S_{\mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}} = \mathcal{A}(\mathcal{E}) \quad (22)$$

III. MAIN RESULTS

Here we state our main results in a concrete way.

A. Decay of CMI for bistochastic Y-shaped channels

Our first main result is showing that any state constructed from a bistochastic³ Y-shaped channel (CPTP-map) is exponentially close to a product state on AB and C . This provides an upper bound of the CMI (and also the trace norm CMI in Sec. III A 1). See Sec. IV A for the proof.

Theorem III.1. *For a bistochastic (unital) Y-shaped channel $\mathcal{N}[\tau_{C_1}] = \tau_{B_1} \otimes \tau_{C_2}$, consider a family of states defined by*

$$\rho_{AB_1 \dots B_\ell C_\ell} := \left(\text{id}_{\mathcal{H}} \otimes \text{id}_{\mathcal{K}}^{\otimes (\ell-1)} \otimes \mathcal{N} \right) \circ \dots \circ \left(\text{id}_{\mathcal{H}} \otimes \text{id}_{\mathcal{K}} \otimes \mathcal{N} \right) \circ \mathcal{N}[\sigma_{A\bar{A}}], \quad (23)$$

where $\sigma_{A\bar{A}}$ is the maximally entangled state. Then, it holds that

$$\|\rho_{AB_1 \dots B_\ell C_\ell} - \rho_{AB_1 \dots B_\ell} \otimes \tau_C\|_1 = O(\eta^\ell). \quad (24)$$

This implies the CMI is bounded as

$$I(A : C | B_1 \dots B_\ell)_{\rho_\ell} = O(\ell \eta^\ell). \quad (25)$$

The contraction ratio η is given as

$$\eta := \limsup_{\rho_C \rightarrow \tau_C} \frac{D(\mathcal{N}[\rho_C] \| \tau_{B_1 C})}{D(\rho_C \| \tau_C)}, \quad (26)$$

where $\eta < 1$ if and only if \mathcal{N} has trivial correctable algebra. Furthermore, η coincides with the second eigenvalue of channel $\mathcal{Z} := \mathcal{P}_{\tau, \mathcal{N}} \circ \mathcal{N}$, where $\mathcal{P}_{\tau, \mathcal{N}}$ is the Petz recovery map [28] as in Eq.(21) with the completely mixed state τ as reference.

Remark. Here we only consider a particular tripartition ABC to show the decay of the CMI. However, under additional moderate assumptions, the exponential decay of CMI of this tripartition implies the CMI is also small for other tripartitions like $A' := AB_1 \dots B_n$, $B' := B_{n+1} \dots B_{n+m}$, $C' := B_{n+m+1} \dots B_\ell C_\ell$. To see this, we use the chain rule of CMI: $I(A' : C' | B')_\rho = I(A : C_\ell | B_1 \dots B_\ell)_\rho + I(A : B_{n+m+1} \dots B_\ell | B_1 \dots B_{n+m})_\rho + I(B_1 \dots B_n : B_{n+m+1} \dots B_\ell C_\ell | B_{n+1} \dots B_{n+m})_\rho$. The second term is upper bounded by $I(A : C_{n+m} | B_1 \dots B_{n+m})_\rho$ by the monotonicity of the CMI, and thus the first and the second terms obey the exponential decay as m grows. Bounding the third term may require an

³ Usually, bistochastic refers to a self-map channel in the literature. Here we have higher output dimension than input, but we abuse the terminology in this paper.

additional assumption. One way to bound the last term is by explicit calculation. For instance, it is 0 for the bistochastic case since $\rho_{BC} = \tau_{BC}$. Another way to bound it is assuming the zero correlation length condition (bottom, Fig. 2), under which the state can be generated by the inverse channel $\tilde{\mathcal{N}} : C \rightarrow BA$ as well as \mathcal{N} . This guarantees $I(B_1 \dots B_n : B_{n+m+1} \dots B_{\ell} C_{\ell} | B_{n+1} \dots B_{n+m})_{\rho}$ is upper bounded by $I(A_n : C_{n+m} | B_{n+1} \dots B_{n+m})_{\rho}$, which is the CMI for ρ_m up to the translation shift.

In this paper, we pay less attention to the case of periodic MPDO where output C of channel got fed into the input A , as this gluing step might be tricky. In the bistochastic case this is tractable (Appendix B) and the state becomes even simpler as the maximally mixed state plus an exponentially decaying global operator. There, the parent Hamiltonian would be trivial in the thermodynamic limit.

A crucial property of bistochastic Y-shaped channels is they act as a contraction onto certain operator subspace. This implies Eq. (24) at large ℓ , and then the CMI bound follows from the continuity of the entropy. An example of bistochastic Y-shaped channel is given as follows.

Example: A bistochastic Y-shaped channel $\mathcal{E}_p : \mathcal{B}(\mathbb{C}^2) \rightarrow \mathcal{B}(\mathbb{C}^2 \otimes \mathbb{C}^2)$ defined by Kraus operators

$$K_1 = \begin{bmatrix} \frac{\sqrt{1-p}}{2} & \frac{\sqrt{1-p}}{2} \\ 0 & 0 \\ \frac{\sqrt{1-p}}{2} & -\frac{\sqrt{1-p}}{2} \\ 0 & 0 \end{bmatrix}, K_2 = \begin{bmatrix} \frac{\sqrt{p}}{2} & \frac{\sqrt{p}}{2} \\ 0 & 0 \\ -\frac{\sqrt{p}}{2} & \frac{\sqrt{p}}{2} \\ 0 & 0 \end{bmatrix}, K_3 = \begin{bmatrix} 0 & 0 \\ \frac{\sqrt{1-p}}{2} & \frac{\sqrt{1-p}}{2} \\ 0 & 0 \\ -\frac{\sqrt{1-p}}{2} & \frac{\sqrt{1-p}}{2} \end{bmatrix}, K_4 = \begin{bmatrix} 0 & 0 \\ \frac{\sqrt{p}}{2} & \frac{\sqrt{p}}{2} \\ 0 & 0 \\ \frac{\sqrt{p}}{2} & -\frac{\sqrt{p}}{2} \end{bmatrix}, \quad (27)$$

has exponentially decaying CMI for $p \in (0, 1/2) \cup (1/2, 1)$. Otherwise, $I(A : C|B)_{\rho_{\ell}} = 0$ for $p = 1/2$ and $I(A : C|B)_{\rho_{\ell}} = 1$ for $p = 0, 1$.

While being bistochastic is crucial to provide an decay rate defined by relative entropy in Theorem III.1, we can work with slightly general channels with a likely suboptimal decay rate. See Sec. IV B for the proof.

Theorem III.2. For a channel such that $\mathcal{N}[\nu_{C_1}] = \sigma_{B_1} \otimes \nu_{C_2}$ for some σ and ν , and the maximally mixed input state $\sigma_{A\bar{A}}$, the resulting chain

$$\rho_{AB_1 \dots B_{\ell} C_{\ell}} := \left(\text{id}_{\mathcal{H}} \otimes \text{id}_{\mathcal{K}}^{\otimes (\ell-1)} \otimes \mathcal{N} \right) \circ \dots \circ \left(\text{id}_{\mathcal{H}} \otimes \text{id}_{\mathcal{K}} \otimes \mathcal{N} \right) \circ \mathcal{N}[\sigma_{A\bar{A}}] \quad (28)$$

satisfies

$$\|\rho_{AB_1 \dots B_{\ell} C_{\ell}} - \rho_{AB_1 \dots B_{\ell}} \otimes \nu_C\|_1 = O(\eta^{\ell}). \quad (29)$$

This implies the CMI is bounded as

$$I(A : C | B_1 \dots B_{\ell})_{\rho_{\ell}} = O(\ell \eta^{\ell}). \quad (30)$$

The contraction ratio η is given as

$$\eta := \min(1, 16d_C\eta_{1,C}), \quad (31)$$

where $\eta_{1,C}$ is the trace norm contraction ratio defined on system C only

$$\eta_{1,C} := \sup_{\rho_C, \rho'_C} \frac{\|\mathcal{E}_C[\rho_C] - \mathcal{E}_C[\rho'_C]\|_1}{\|\rho_C - \rho'_C\|_1}. \quad (32)$$

The decay of CMI for more general tripartition follows from the same argument as the bistochastic case (Theorem III.1). The distinction, though, is that our bound for the contraction ratio suffers from extra factor of dimension d_C due to conversion to the completely bounded norm (Lemma V.3), so the bound is meaningful only for the noisy regime $\eta_{1,C} < (16d_c)^{-1}$.

1. The trace norm CMI

In some cases, the trace norm (or the trace distance) has more applicable tools than the relative entropy. We now introduce a trace-norm variant of CMI. We expect it to help understanding the qualitative behavior of the CMI.

Definition 2 (The Trace norm CMI).

$$I_1(A : C|B) := \|\rho_{ABC} - \rho_A \otimes \rho_{BC}\|_1 - \|\rho_{AB} - \rho_A \otimes \rho_B\|_1 \quad (33)$$

The definition is motivated by the form of CMI as $I(A : C|B) = I(A : BC) - I(A : B)$ and replace each mutual information by the trace distance (without the normalization for the simplicity). The mutual information and the trace distance bound each other (in a non-linear way) by the quantum Pinsker inequality [29, Theorem 11.9.1] and the continuity of entropy, and thus we expect this trace-norm variant of CMI partially captures common characteristics of the states as the CMI. Indeed, in the special cases analyzed in the following two theorems (Theorem III.1, Theorem III.2), we arrive at nearly same bound on decay rate for both CMI and the trace norm CMI, up to fixed overheads and logarithmic correction.

Proposition III.1. *The trace norm CMI is bounded as*

$$I_1(A : C|B_1 \cdots B_\ell)_{\rho_\ell} = O(\eta^\ell) \quad (34)$$

for both states defined in Theorem III.1 and Theorem III.2, with η defined respectively.

Unfortunately, we do not know to what extent the properties of CMI carry over, such as the connection to Markov states and recoverability. We leave further analysis on this function in future works.

B. Decay of CMI for Y-shaped channels with a forgetful component

Although we have shown the upper bound of CMI for bistochastic Y-shaped channels, they are only a small portion of the space of channels. Here, we show another suggestive result towards Conjecture I.1.

Proposition III.2. *Consider a Y-shaped channel $\mathcal{N} : C \rightarrow B_1 C$ such that there is a constant $0 \leq \eta < 1$, a forgetful channel F and another channel \mathcal{N}' such that*

$$\mathcal{N} = (1 - \eta)F + \eta\mathcal{N}', \quad (35)$$

where $F[\rho] = \sigma$ maps any state to the same state. Then for any tripartite state $\rho_{A'B'C}$, the CMI contracts after applying \mathcal{N} ,

$$I(A' : C|B'B_1)_{\mathcal{E}(\rho)} \leq \eta I(A' : C|B')_{\rho}, \quad (36)$$

$$I_1(A' : C|B'B_1)_{\mathcal{E}(\rho)} \leq \eta I_1(A' : C|B')_{\rho}. \quad (37)$$

Hence the resulting chain has exponentially decaying CMI and trace norm CMI for any tripartition

$$I(AB_1 \dots B_n : B_{n+m+1} \dots B_{\ell} C_{\ell} | B_{n+1} \dots B_{n+m})_{\rho_{\ell}} = O(\eta^m), \quad (38)$$

$$I_1(AB_1 \dots B_n : B_{n+m+1} \dots B_{\ell} C_{\ell} | B_{n+1} \dots B_{n+m})_{\rho_{\ell}} = O(\eta^m). \quad (39)$$

Remarkably, each application of \mathcal{N} contracts the CMI by η , from simple application of convexity of relative entropy and monotonicity of CMI (proof in Sec. VB3). Intuitively the contraction ratio is η because the forgetful channel necessarily removes any correlation with system AB . The step wise decay implies the decay of CMI for any tripartition with a long conditioned system B , i.e. the Y-shaped channel is applied many times.

Note that any strictly positive Y-shaped channel $\mathcal{N}(\rho) > 0$ always has a decomposition (35) with F being the completely depolarizing channel. Moreover, for any Y-shaped channel \mathcal{N}' and any $0 \leq \eta < 1$, we can always construct a perturbed channel \mathcal{N} in Eq. (35) satisfying $\|\mathcal{N} - \mathcal{N}'\|_{cb} \leq 2(1 - \eta)$.

In the sense that any channel can be perturbed to have a forgetful component, Proposition III.2 is true for most of quantum channels. We should note that the appropriate choice of ‘generic’ MPDO certainly depends on the constraint of the problem at hand, for example if we restrict the number of Kraus operators to be small, then strict positivity would not hold generically.

A channel with forgetful component necessarily has trivial correctable algebra, but not vice versa. We can see this fact through an explicit example.

Example: The classical channel given by the following Kraus operators has trivial correctable algebra, but has no forgetful component. Input states $|1\rangle\langle 1|, |2\rangle\langle 2|, |3\rangle\langle 3|$ are mapped to

$\frac{|2\rangle\langle 2|+|3\rangle\langle 3|}{2}$, $\frac{|1\rangle\langle 1|+|3\rangle\langle 3|}{2}$, $\frac{|1\rangle\langle 1|+|2\rangle\langle 2|}{2}$, whose intersection is empty and thus cannot have a shared forgetful component.

$$K_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, K_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, K_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \quad (40)$$

C. A completely contractive DPI of CMI that implies exponential decay of CMI

We have shown the channels we considered in Sec. III A, III B incur the exponential decay of CMI. For channels with a forgetful component (Proposition III.2), the exponential decay simply comes from a contraction occurring at each application of channel, and it is tempting to ask whether this is the general case. This motivates the following discussion on the data processing inequality.

First, we show that a single application of any channel with trivial correctable algebra always induces strict decay of CMI.

Proposition III.3. *Let $\mathcal{E} : C \rightarrow C'$ be a CPTP map which has trivial correctable algebra. Consider a tripartite system $A \otimes B \otimes C$. Then, for any state ρ_{ABC} with $I(A : C|B)_\rho > 0$, we have*

$$I(A : C'|B)_{\mathcal{E}(\rho)} < I(A : C|B)_\rho. \quad (41)$$

The proof is given in Sec. VB 1. By regarding \mathcal{E} as Y-shaped channel $\mathcal{N}_{C_{i-1} \rightarrow B_i C_i}$, we see that $I(A : C|B)_{\rho_k} < I(A : C|B)_{\rho_{k-1}}$ holds for each k (we abuse notation BC for two different lengths). Note that having trivial correctable algebra provides only a sufficient condition to have a strict decay of the CMI. There is a Y-shaped channel with non-trivial correctable algebra which obey Eq. (41) (see Ref. [30] for a complementary result).

Unfortunately, recursively applying Eq. (41) may not imply the exponential decay of CMI of ρ_ℓ as the decay might get slower as system B gets larger. We therefore propose the following conjecture to explicitly include a contraction ratio as a strong DPI constant.

Conjecture III.1 (Completely Contractive DPI for the CMI). *For any channel $\mathcal{E} : C \rightarrow C'$ with trivial correctable algebra $\mathcal{A}(\mathcal{E}) = \mathbb{C}I$, there exists a constant $\eta < 1$ such that for any tripartite system ABC and any state ρ_{ABC} , it holds that*

$$I(A : C'|B)_{\mathcal{E}(\rho)} \leq \eta I(A : C|B)_\rho. \quad (42)$$

Invoking standard monotonicity inequalities (Sec. VB 2), the contraction accumulates and thus Conjecture III.1 implies the desired exponential decay of CMI:

Proposition III.4. *If Conjecture III.1 is true, then for any $\rho_\ell(\mathcal{N})$ constructed by Y-shaped channel \mathcal{N} with trivial correctable algebra,*

$$I(A' : C' | B')_\rho = \mathcal{O}(e^{-(\ln \eta)^m}) \quad (43)$$

for each tripartition $A'B'C' = ABC$ where B' separates A' from C' with distance m .

Conjecture III.1 holds for the channels with a forgetful component (Proposition III.2), where $\eta_{ABC}(\mathcal{E})$ does not approach 1 as AB grows. In the classical case generally an unbounded auxiliary system can be reduced to have bounded dimension only depending on C , and for the strong DPI constants the auxiliary system can be ignored $\eta_{ABC} = \eta_{AC} = \eta_C$ (Theorem II.1). In the quantum case we do not know whether AB can be reduced to a bounded dimension. Hence, the conjecture is open in general quantum systems and even showing its simpler variant for the mutual information (setting B trivial) would be a breakthrough. Very recently the case when A is classical and B is trivial was proven with a contraction ratio analogous to the classical case [45].

A similar DPI can be proposed for the trace norm CMI.

Conjecture III.2 (completely contractive DPI for the trace norm CMI). *For any channel $\mathcal{E} : C \rightarrow C'$ with local contraction ratio*

$$\eta_{1,C} := \sup_{\rho_C, \rho'_C} \frac{\|\mathcal{E}_C[\rho_C] - \mathcal{E}_C[\rho'_C]\|_1}{\|\rho_C - \rho'_C\|_1} < 1. \quad (44)$$

then there exists a global constant $\eta < 1$ such that for any tripartite system ABC and any state ρ_{ABC} , it holds that

$$I_1(A : C' | B)_{\mathcal{E}(\rho)} \leq \eta I_1(A : C | B)_\rho. \quad (45)$$

We do not know if extra constants or factor of dimension d_C should be present between $\eta_{1,C}$ and η like the crude bound in Theorem III.2. The trace norm seems more tractable than mutual information in the bipartite case.

Proposition III.5 (Contraction of the trace norm mutual information). *For any channel $\mathcal{N} : C \rightarrow C'$, it holds that*

$$\|\mathcal{N}[\rho_{BC} - \rho_B \otimes \rho_C]\|_1 \leq 4\eta_{1,C}d_C\|\rho_{BC} - \rho_B \otimes \rho_C\|_1 \quad (46)$$

for any state ρ_{BC} .

Here the contraction ratio is controlled by a bounded factor, where the dimension factor comes from bounding the completely bounded superoperator 1 – 1 norm (Lemma V.3).

D. Sufficient conditions for exponential decay of classically-conditioned mutual information

In this section we consider the CMI of MPDO after performing the local measurement on the computational basis of B . Here we do not require the original map \mathcal{N} in Eq. (4) to be trace-preserving.

Theorem III.3 (Sufficient conditions for decay of CMI for measured MPDO). *For a B -measured MPDO as in Eq. (6), the following are implications in the order $1 \implies 2 \implies 3 \implies 4$:*

1. (**Condition 1**) Each \mathcal{M}_s has p Kraus operators $\{E_1^s, \dots, E_p^s\}$ such that E_p^s is invertible, and

$$\left\{ (E_p^s)^{-1} E_1^s, \dots, (E_p^s)^{-1} E_{p-1}^s \right\}$$

generate full matrix algebra $M_D(\mathbb{C})$ by addition and multiplication.

2. (**Condition 2**) There exists uniform coarse-graining length ξ such that any sequence $\mathcal{M}'_t := \mathcal{M}_{s_{k+l}} \cdots \mathcal{M}_{s_k}$, where $t = (s_k, \dots, s_{k+l})$, is strictly positive map for $\forall l \geq \xi$. In other words,

$$\det(\mathcal{M}'_t[\rho]) \neq 0, \quad \forall t, \forall \rho \quad (47)$$

whenever $|t| \geq \xi$.

3. The mutual information of A, C observing any outcome b is bounded by a uniform decay rate $c_1 > 0$

$$I(A : C)_b \leq c_0 e^{-c_1 \ell}. \quad (48)$$

4. The measured conditional mutual information decay exponentially with the length ℓ of system B

$$I(A : C | \bar{B}) \leq c_0 e^{-c_1 \ell}. \quad (49)$$

Condition 2 (strict positivity of all \mathcal{M}_s) is unlikely a necessary condition for the decay of CMI, but it is convenient to state. We include coarse-graining in the statement as it is sometimes needed to ensure strict positivity of all \mathcal{M}_s . For example, if the original Y-shaped CP map comes from tracing out a system much smaller than the bond dimension, then it may not be strictly positive.

A similar bound for the coarse-graining length ξ for iteration of a single channel was given by the quantum Wielant inequality [31]: for every primitive channel \mathcal{N} (with Kraus rank p and Hilbert space dimension D), $\xi = D^2(D^2 - d + 1)$ iteration guarantees full-rank output. Our Condition 1 is a multi-channel generalization: it is a sufficient condition for all possible sequences of $\mathcal{M}_{s_\ell} \circ \cdots \circ \mathcal{M}_{s_2} \circ \mathcal{M}_{s_1}$ generated by CP maps $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_n\}$ to become strictly positive for coarse graining length $l \geq \xi := D^2 - p + 1$. See Sec. IV C for the proof.

Although strict positivity (Condition 2) is NP-hard to check in general [32], we prove that Condition 1 can be verified in a polynomial time (see Sec. V D 5 for the details).

Proposition III.6. *Condition 1 can be checked in time polynomial in the bond dimension D and the number of Kraus operators p .*

We numerically verified Condition 1 for Y-shaped channels generated by Haar-random Stinespring unitary (Fig. 2), and confirmed it seems true at least for computationally tractable bond dimensions⁴. Both conditions imply Eq. (48), which further implies our original goal, Eq. (49).

IV. PROOFS OF MAIN THEOREMS

Here we present proofs of the main results by putting the key lemmas together, whose proofs we postpone in Sec. V.

A. Theorem III.1: decay of CMI for bistochastic channel

The bistochastic assumption greatly simplifies the structure of state. The state can be decomposed into an uncorrelated state $\rho_{AB} \otimes \tau_C$ plus a deviation G_{ABC} that is traceless on system C . At each application of channel in the generation of ρ_ℓ , there are two key implications of being bistochastic: 1) the uncorrelated state $\rho_{AB} \otimes \tau_C$ is mapped to an uncorrelated state where system C remains the maximally mixed state τ_C , 2) the deviation G_{ABC} contracts w.r.t the normalized Hilbert-Schmidt norm. The rest of the proofs are obtained by standard conversion between norms. The proofs of the lemmas are shown independently in Sec. V A.

Proof. Let us take an operator basis containing identity which is orthogonal in the Hilbert-Schmidt inner product. Then any operator on C can be decomposed into the maximally mixed state τ_C and the traceless part K_C . After applying bistochastic Y-shaped channel $\mathcal{N} : C_i \rightarrow B_{i+1}C_{i+1}$, the traceless part is mapped to

$$\mathcal{N}[K_C] = O_B \otimes \tau_C + \sum_j O_B^j \otimes K_C^j$$

where we abused the notation O, O^j to denote general operators and K, K^j to denote traceless operators. In the following we will use the curly brackets $\{O_B \otimes K_C\} := \sum_j O_B^j \otimes K_C^j$ as a shorthand notation for this type of linear combinations. Iteratively decomposing system C into the traceless and maximally mixed components after applying \mathcal{N} , we can write down the structure of the state $\rho_{AB_1 \dots B_\ell C_\ell}$ explicitly.

⁴ Here is the repository of the jupyter code[46].

Starting with the maximally entangled input state $\sigma_{A\bar{A}} = \tau_A \otimes \tau_{\bar{A}} + \{K_A \otimes K_{\bar{A}}\}$, we obtain that

$$\begin{aligned} \rho_{AB_1C_1} &= \mathcal{N}[\tau_{A\bar{A}} + \{K_A \otimes K_{\bar{A}}\}] = \tau_{AB_1} \otimes \tau_{C_1} + O_{AB_1} \otimes \tau_{C_1} + \{O_{AB_1} \otimes K_{C_1}\} \\ &= \rho_{AB_1} \otimes \tau_{C_1} + \{O_{AB_1} \otimes K_{C_1}\}, \end{aligned} \quad (50)$$

$$\begin{aligned} \rho_{AB_1B_2C_1} &= \mathcal{N}[\rho_{AB_1C_1}] = \rho_{B_1} \otimes \tau_{B_2} \otimes \tau_{C_2} + O_{AB_1} \otimes O_{B_2} \otimes \tau_{C_2} + \{O_{AB_1} \otimes O_{B_2} \otimes K_{C_1}\} \\ &= \rho_{AB_1B_2} \otimes \tau_{C_1} + \{O_{AB_1B_2} \otimes K_{C_1}\}, \end{aligned} \quad (51)$$

$$\rho_{AB_1 \dots B_\ell C_\ell} = \tau_{AB_1} \otimes \dots \otimes \tau_{C_\ell} \quad (52)$$

$$+ O_{AB_1} \otimes \tau_{B_2} \dots \tau_{C_\ell} \quad (53)$$

$$+ \dots \quad (54)$$

$$+ \{O_{AB_1} \otimes O_{B_2} \dots \otimes K_{C_\ell}\} \quad (55)$$

$$= \rho_{AB} \otimes \tau_{C_\ell} + \{O_{AB} \otimes K_{C_\ell}\}, \quad (56)$$

where we simply denote system $B_1 \dots B_\ell$ by B . The second term $\{O_{AB} \otimes K_{C_\ell}\}$ in Eq. (56) is exponentially suppressed in the normalized Hilbert-Schmidt norm. To see this, we use the following lemma for bistochastic channels.

Lemma IV.1 (DPI for normalized HS-norm). *For bistochastic channel $\mathcal{N}_C[\tau_C] = \tau_{C'}$, the normalized Hilbert-Schmidt norm of operators K_{EC} traceless on C contracts. More precisely, if $\text{Tr}_C(K_{EC}) = 0$ then*

$$\text{Tr}(d_{EC'} \mathcal{N}_C[K]_{EC}^2) \leq \eta \text{Tr}(d_{EC} K_{EC}^2) \quad (57)$$

with the coefficient η defined as

$$\eta := \limsup_{\rho_C \rightarrow \tau_C} \frac{D(\mathcal{N}[\rho] \parallel \tau_{C'})}{D(\rho_C \parallel \tau_C)}. \quad (58)$$

Let \mathbb{P}_K be the orthogonal projection onto the traceless operator subspace on C , $G_{ABC} := \{O_{AB} \otimes K_C\}$ and $G_{A\bar{A}} := \{K_A \otimes K_{\bar{A}}\}$. Then, we obtain an expression

$$G_{ABC} = \mathbb{P}_K \circ \mathcal{N} \circ \mathbb{P}_K \circ \mathcal{N} \dots \mathbb{P}_K \circ \mathcal{N}[G_{A\bar{A}}]. \quad (59)$$

By Lemma IV.1, the normalized HS norm of G_{ABC} contracts under $\mathbb{P}_K \circ \mathcal{N}$ since the projection does not increase the HS norm. We thus obtain

$$d_{ABC} \text{Tr}(G_{ABC})^2 = d_A d_{B_1 \dots B_\ell} d_C \text{Tr}(\mathbb{P}_K \circ \mathcal{N} \circ \mathbb{P}_K \circ \mathcal{N} \circ \dots \mathcal{N}(G_{A\bar{A}}))^2 \quad (60)$$

$$\leq d_A d_{B_1 \dots B_\ell} d_C \text{Tr}(\mathcal{N} \circ \mathbb{P}_K \circ \mathcal{N} \circ \dots \mathcal{N}(G_{A\bar{A}}))^2 \quad (61)$$

$$\leq \eta \cdot d_A d_{B_{B_1 \dots B_{\ell-1}}} d_C \text{Tr}(\mathbb{P}_K \circ \mathcal{N} \circ \dots \mathcal{N}(G_{A\bar{A}}))^2 \quad (62)$$

$$\vdots \quad (63)$$

$$\leq d_A d_C \eta^l \text{Tr}(G_{A\bar{A}})^2. \quad (64)$$

By Cauchy-Schawrtz inequality, we can bound the trace norm of G_{ABC} by the HS norm.

$$\|G_{ABC}\|_1 \leq d_{ABC} \text{Tr}(G_{ABC})^2 \leq \eta^\ell d_A^2 \text{Tr}(G_{A\bar{A}})^2. \quad (65)$$

Here we used $d_A = d_C$.

Since $\|G_{ABC}\|_1$ is exponentially suppressed, ρ_{ABC} is close to $\rho_{AB} \otimes \tau_C$ for large ℓ . The next step is to convert to CMIs, and clearly $\rho_{AB} \otimes \tau_C$ has zero CMI. From the continuity of entropy, the difference of the CMI of two states are bounded as follows.

Lemma IV.2 (Continuity of CMI). *If $\rho_{ABC} = \sigma_{ABC} + G_{ABC}$ and $d_{ABC} \text{Tr}(G_{ABC})^2 \leq \epsilon \leq 1/e$, then*

$$|I(A : C|B)_\rho - I(A : C|B)_\sigma| \leq 4 \log(d_{ABC})\epsilon - 4 \log(\epsilon)\epsilon. \quad (66)$$

By setting $\sigma_{ABC} = \rho_{AB} \otimes \tau_C$ and taking large enough ℓ such that the deviation G_{ABC} is small enough $\eta^\ell d_A^2 \text{Tr}(G_{A\bar{A}})^2 \leq 1/e$, we obtain the exponential decay of the CMI:

$$I(A : C|B)_\rho \leq 4(\log(d_{B_1})\ell + 2 \log(d_A))\eta^\ell d_A^2 \text{Tr}(G_{A\bar{A}})^2 - 4(\ell \log(\eta) + \log(d_A^2 \text{Tr}(G_{A\bar{A}})^2))\eta^\ell d_A^2 \text{Tr}(G_{A\bar{A}})^2 \quad (67)$$

$$\leq O(\ell\eta^\ell). \quad (68)$$

Lastly, we alternatively characterize η by a spectral property:

Lemma IV.3. *For a bistochastic channel \mathcal{E} , the second largest singular value (which coincides with eigenvalue due to being self-adjoint) of the Petz-recovery channel composed with the channel $\mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}$ is exactly the contraction ratio η .*

$$\eta = \lambda_2(\mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}) = \limsup_{\rho \rightarrow \tau} \frac{D(\mathcal{E}[\rho] \| \tau')}{D(\rho \| \tau)}. \quad (69)$$

Therefore $\eta < 1$ if and only if the largest singular value (or equivalently the eigenvalue) $\lambda_1 = 1$ is unique. By Proposition II.1, the unit eigenvalue subspace of λ_1 being one-dimensional is equivalent to the correctable algebra being trivial.

$$\mathcal{A}(\mathcal{E}) = \mathbb{C}I \iff \lambda_2(\mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}) = \eta < 1. \quad (70)$$

This completes the second statement of the proof. \square

B. Theorem III.2: decay of trace norm CMI for partially invariant channel

The proof structure is analogous to the previous Section IV A, with the key distinction being the norm and the associated tools. Under the trace norm we have results for a slightly general family of channels satisfying $\mathcal{N}[\nu_C] = \sigma_B \otimes \nu_C$. Though, unlike the HS norm, the trace norm suffers from tensoring with

an auxiliary system by a bounded factor of dimension of system C (Lemma V.3). The contraction ratio obtained this way is likely not the most stringent, but at least it implies strict contraction of the CMI when the channel is sufficiently noisy. The proofs for the key lemmas are shown in Sec. VC.

Proof. First, we decompose the state into a product state $\rho_{AB} \otimes \nu_C$ plus some deviation G_{ABC} traceless on system C . Suppose

$$\mathcal{N}[K_C] = \mathbb{P}_\nu \circ \mathcal{N}[K_C] + (\text{id} - \mathbb{P}_\nu) \circ \mathcal{N}[K_C] = O_B \otimes \nu_C + \{O_B \otimes K_C\}, \quad (71)$$

where $\mathbb{P}_\nu := \nu_C \text{Tr}_C[\cdot]$ is the forgetful channel on subsystem C . Again we used K_C to denote an operator traceless on system C , but we should note the difference from Sec. IV A that the projection $(\text{id} - \mathbb{P}_\nu)$ is not an orthogonal projection under HS norm. Writing the maximally entangled state as $\sigma_{A\bar{A}} = \tau_A \otimes \nu_{\bar{A}} + \{O_A \otimes K_{\bar{A}}\}$, we obtain that

$$\begin{aligned} \rho_{AB_1C_1} &= \mathcal{N}[\tau_A \otimes \nu_{\bar{A}} + \{O_A \otimes K_{\bar{A}}\}] = \tau_A \otimes \sigma_{B_1} \otimes \nu_{C_1} + O_{AB_1} \otimes \nu_{C_1} + \{O_{AB_1} \otimes K_{C_1}\} \\ &= \rho_{AB_1} \otimes \nu_1 + \{O_{AB_1} \otimes K_{C_1}\}, \end{aligned} \quad (72)$$

$$\rho_{AB_1B_2C_1} = \mathcal{N}[\rho_{AB_1C_1}] \quad (73)$$

$$\begin{aligned} &= \rho_{AB_1} \otimes \sigma_{B_2} \otimes \nu_{C_2} + O_{AB_1} \otimes O_{B_2} \otimes \nu_{C_2} + \{O_{AB_1} \otimes O_{B_2} \otimes K_{C_1}\} \\ &= \rho_{AB_1B_2} \otimes \nu_{C_1} + \{O_{AB_1B_2} \otimes K_{C_1}\}, \end{aligned} \quad (74)$$

$$\begin{aligned} \rho_{AB_1 \dots B_\ell C_\ell} &= \tau_A \otimes \sigma_{B_1} \otimes \sigma_{B_2} \dots \nu_{C_\ell} \\ &+ O_{AB_1} \otimes \sigma_{B_2} \dots \nu_{C_\ell} \\ &+ \dots \\ &+ \{O_{AB_1} \otimes O_{B_2} \dots \otimes K_{C_\ell}\} \\ &= \rho_{AB} \otimes \nu_{C_\ell} + \{O_{AB} \otimes K_{C_\ell}\} \end{aligned} \quad (75)$$

$$= \rho_{AB} \otimes \nu_{C_\ell} + (\text{id} - \mathbb{P}_\nu) \circ \mathcal{N} \dots (\text{id} - \mathbb{P}_\nu) \circ \mathcal{N}(G_{A\bar{A}}). \quad (76)$$

Let $G_{ABC} := (\text{id} - \mathbb{P}_\nu) \circ \mathcal{N} \dots \circ (\text{id} - \mathbb{P}_\nu) \circ \mathcal{N}(G_{A\bar{A}})$, which is the output of iteration of the map $(\text{id} - \mathbb{P}_\nu) \circ \mathcal{N} \circ (\text{id} - \mathbb{P}_\nu)$, where we multiply a copy of $(\text{id} - \mathbb{P}_\nu)$ as its square is equal to itself. The completely bounded norm of this map governs the contraction ratio

$$\|G_{ABC}\|_1 \leq \|(\text{id} - \mathbb{P}_\nu) \circ \mathcal{N} \circ (\text{id} - \mathbb{P}_\nu)\|_{\text{cb}}^\ell \|G_{A\bar{A}}\|_1. \quad (77)$$

Unfortunately, the completely bounded norm of difference of channels is not less than one in general. However, we do obtain a meaningful contraction bound for $\eta_{1,C}$ is small enough.

Lemma IV.4. *The completely bounded norm of the difference between channels is bounded by*

$$\|(\text{id} - \mathbb{P}_\nu) \circ \mathcal{N} \circ (\text{id} - \mathbb{P}_\nu)\|_{\text{cb}} \leq 2\|\mathcal{N} - \mathcal{N} \circ \mathbb{P}_\nu\|_{\text{cb}} \leq 16d_C \eta_{1,C}, \quad (78)$$

where $\eta_{1,C}$ is the trace norm contraction ratio

$$\eta_{1,C} := \sup_{\rho_C, \rho'_C} \frac{\|\mathcal{N}[\rho_C] - \mathcal{N}[\rho'_C]\|_1}{\|\rho_C - \rho'_C\|_1}. \quad (79)$$

The constant factor 16 is crude, and the factor of dimension d_C is due to converting 1-1 superoperator norm to the diamond norm. From this lemma Eq. (77) reduces to

$$\|G_{ABC}\|_1 \leq (16d_C\eta_{1,C})^\ell \|G_{A\bar{A}}\|_1. \quad (80)$$

By the triangle inequality we show that ρ_{ABC} become close to a product state in trace norm

$$\|\rho_{ABC} - \rho_{AB} \otimes \nu_C\|_1 \leq O\left((16\eta_C d_C)^\ell\right). \quad (81)$$

We conclude the proof by calling the continuity of CMI (Lemma IV.2) again. \square

C. Theorem III.3: decay of B -measured CMI in MPDO

First, 3 \implies 4 is immediate from that the measured CMI is equal to the expectation of mutual information between A and C over measurement outcomes b in \bar{B} ,

$$I(A : C | \bar{B})_\rho = \sum_b p_{\bar{B}}(b) I(A : C)_b, \quad (82)$$

where $I(A : C)_b$ is the mutual information of state $\rho_{AC| \bar{B}=b} \propto \text{Tr}_{\bar{B}}(|b\rangle\langle b| \rho_{ABC})$, defined on the Hilbert space of system A and C only, as we can individually discuss each classical outcome $\bar{B} = b$.

We show the rest two indications 1 \implies 2 and 2 \implies 3 in the following.

1. 2 \implies 3: uniformly bounding every sequence quantum operations by contraction in Hilbert's projective metric

Proof. We show Condition 2 implies the mutual information $I(A : C)_b$ decays exponentially with a uniform decay rate c for every outcome b . As showed in the above, this implies the decay of CMI without touching the probability distribution $p(b)$.

Recall that we have a set of CP-maps $\mathcal{M}_{s_k} : C \rightarrow C$ for each outcome s_k . These maps \mathcal{M}_{s_k} (or the coarse-grained maps $\mathcal{M}'_t := \mathcal{M}_{s_k+\xi} \circ \dots \circ \mathcal{M}_{s_k}$) are all contraction in the *Hilbert's projective metric* $h(a, b)$. A self-contained introduction and supporting theorems are included in Section VD 1. The contraction ratio associated to a CP-map \mathcal{M} is given by

$$\eta_{\mathcal{M}} := \sup_{a, b \in S_+} \frac{h(\mathcal{M}(a), \mathcal{M}(b))}{h(a, b)} \leq 1, \quad (83)$$

where the supremum is taken over S_+ , the set of unnormalized positive semi-definite operators. Condition 2 guarantee that $\{\mathcal{M}'_t\}$ has a strict contraction ratio at Eq. (83). After repeatedly applying \mathcal{M}'_t , any state is mapped to a fixed state (up to rescaling) with exponentially small error. Accounting the normalization and the conversion to trace norm, we obtain the following lemma.

Lemma IV.5. *If all $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_n\}$ are CP self-maps that maps any state to a full rank state, then for arbitrary sequence of $b = s_\ell, \dots, s_1 \in \{1, \dots, n\}^\ell$ and all ρ_1, ρ_2*

$$\left\| \frac{\mathcal{M}_b[\rho_1]}{\text{Tr}(\mathcal{M}_b[\rho_1])} - \frac{\mathcal{M}_b[\rho_2]}{\text{Tr}(\mathcal{M}_b[\rho_2])} \right\|_1 = O(e^{-c\ell}), \quad (84)$$

where the exponent c is independent of b .

This statement implies that the bipartite state $\rho_{AC,b}$ being close to a product state in the trace norm.

Lemma IV.6. *Suppose the CP map $\mathcal{M}_b : C' \rightarrow C$ is contractive in the sense that for all ρ_1, ρ_2 ,*

$$\left\| \frac{\mathcal{M}_b[\rho_1]}{\text{Tr}(\mathcal{M}_b[\rho_1])} - \frac{\mathcal{M}_b[\rho_2]}{\text{Tr}(\mathcal{M}_b[\rho_2])} \right\|_1 \leq \epsilon. \quad (85)$$

Then the state $\rho_{AC} := \mathcal{M}_b[\sigma_{AC'}]/\text{Tr}(\mathcal{M}_b[\sigma_A])$ is closed to the product state in the trace distance

$$2T_b := \|\rho_{AC,b} - \rho_{A,b} \otimes \rho_{C,b}\|_1 \leq 4d_{C'} \|\sigma_{C'}^{-1}\| \epsilon. \quad (86)$$

Plugging $\sigma_{C'} = \tau_{C'}$, and by the Alicki–Fannes–Winter inequality [33] (see also [29, Theorem 11.10.3]), we convert the $O(e^{-c\ell})$ bound on the trace norm to the mutual information with some constant overhead. \square

2. 1 \implies 2: *uniformly bounding coarse graining length by generalizing quantum Wielandt's inequality to family of CP self-maps*

First, we convert strict positivity to having full Kraus rank.

Lemma IV.7. *If $\text{span}\{K_1, \dots, K_p\} = \text{Mat}(D, D)$ then $\mathcal{N}[\rho] := \sum_i K_i \rho K_i^\dagger > 0, \forall \rho$.*

Our strategy to a uniformly bound on ξ now relies on demanding for each s_k , $\mathcal{M}_{s_k}[\cdot] = \sum_i E_i^{s_k} \cdot E_i^{s_k \dagger}$ must increase the dimension of span of Kraus operators until reaching full rank.

Lemma IV.8 (condition 1 implies the increment of span of Kraus operators). *Consider a CP self-map \mathcal{M} with $p > 2$ Kraus operators $\mathcal{M} \sim \{E_1, \dots, E_p\}$ such that E_p is invertible. Suppose that $((E_p)^{-1}E_1, \dots, (E_p)^{-1}E_{p-1})$ generate full matrix algebra $\text{Mat}_{\mathbb{C}}(D, D)$ by addition and multiplication. Then for any set of Kraus operators $\{T_j\}$ containing an invertible element and whose Kraus rank is not full, the dimension of span must increase after applying $\mathcal{M} \sim \{E_1, \dots, E_p\}$*

$$\dim(\text{span}\{E_i T_j\}) > \dim(\text{span}\{T_j\}). \quad (87)$$

We are now ready to prove 1 \implies 2:

Proof. Consider the linear span of Kraus operators from sequence $\mathcal{M}_{s_k} \cdots \mathcal{M}_{s_2} \mathcal{M}_{s_1}$

$$\mathcal{S}_{s_k \cdots s_1} := \text{span}\{E_{q_k}^{s_k} \cdots E_{q_2}^{s_2} E_{q_1}^{s_1}\}.$$

We will shorthand as \mathcal{S}_k , while keeping in mind its dependence on the sequence $s_k \cdots s_1$. Applying $\mathcal{M}_{s_{k+1}}$, i.e. left-multiplying its Kraus operators gives the new span of Kraus operator :

$$\mathcal{S}_{k+1} = \{E_1^{s_{k+1}} \mathcal{S}_k, \cdots, E_{p-1}^{s_{k+1}} \mathcal{S}_k\}.$$

By Lemma IV.8, $\mathcal{M}_{s_{k+1}}$ must increase the span of Kraus operators, i.e. any sequence $\mathcal{M}_{s_\ell} \cdots \mathcal{M}_{s_2} \mathcal{M}_{s_1}$ with $\ell \geq \xi := D^2 - p + 1$ must reach full Kraus rank and hence becomes strictly positive by proposition V.3. Requiring an invertible element for both \mathcal{M}_{s_k} and \mathcal{S}_k ensures that \mathcal{S}_{k+1} has non-decreasing span⁵. Therefore, $\{\mathcal{M}_t\}_{t \in \{1, \dots, n^\xi\}}$ would all have full-rank output (Condition 2) with a global upper bound on contraction ratio $\eta' = \sup_t(\eta_t) < 1$. It corresponds to the following coarse graining

$$b' = t_{\ell'}, \cdots, t_1, \quad (88)$$

$$t \in \{1, \cdots, n^\xi\} = \{1, \cdots, n\}^\xi, \quad (89)$$

where $\ell' = \frac{\ell}{\xi}$ is new length (we can assume ξ divides ℓ w.l.o.g.), and the exponential decay would be $(\eta')^{\ell'}$.

□

V. PROOFS FOR PROPOSITIONS AROUND MAIN THEOREM AND LEMMAS SUPPORTING MAIN THEOREMS

Here we compile the remaining proofs and lemmas, categorized by the theorem they are supporting: Theorem III.1 at Section V A; Theorem III.2 at Section V C; Theorem III.3 at Section V D.

A. Remaining Proof for Theorem III.1

1. Proof of Lemma IV.1, DPI for 2-norm

Proof. By definition, the contraction can be expressed by the completely bounded superoperator 2-2 norm of the channel.

$$(d_{EC'} \mathcal{N}_C[K]_{EC}^2) \leq \|\mathcal{I} \otimes (\mathcal{N} \circ \mathbb{P}_K)\|_{2-2}^2 \frac{d_{EC'}}{d_{EC}} \text{Tr}(d_{EC} K_{EC}^2) \quad (90)$$

$$= \|\mathcal{I} \otimes (\mathcal{N} \circ \mathbb{P}_K)\|_{2-2}^2 \frac{d_{C'}}{d_C} \text{Tr}(d_{EC} K_{EC}^2). \quad (91)$$

⁵ Note that the invertibility condition may not be necessary, but we do not know how to prove without this condition.

The superoperator 2 – 2 norm is equal to the largest-singular-value of the map (with input restricted to be traceless on system C) w.r.t. the 2-norm of operators. Tensoring with identity does not change the leading singular value, and hence the completely bounded 2 – 2 norm can be evaluated without auxiliary system as

$$\|\mathcal{I} \otimes (\mathcal{N} \circ \mathbb{P}_K)\|_{2-2}^2 \frac{d_{C'}}{d_C} = \|\mathcal{N} \circ \mathbb{P}_K\|_{2-2}^2 \frac{d_{C'}}{d_C} \quad (92)$$

$$= \sup_{K_C, \text{Tr}(K_C)=0} \frac{\text{Tr}(\mathcal{N}_C[K]_C^2)}{\text{Tr}(K_C^2)} \cdot \frac{d_{C'}}{d_C} \quad (93)$$

$$= \lim_{\rho_C \rightarrow \tau_C} \sup \frac{D(\mathcal{N}[\rho] \|\tau_{C'})}{D(\rho_C \|\tau_C)}, \quad (94)$$

In the last equality we convert the 2-2 norm to the perturbation of relative entropy

$$D(\rho \|\tau) = -S(\rho) + \log(d) = \frac{1}{2} \text{Tr}(dK^2) + O(K^3). \quad (95)$$

The $O(K^3)$ and higher order terms vanishes in the limit because both spaces C and C' have finite dimension, which guarantee the operator norm vanishes $\|K\| \rightarrow 0$ in the limit of $\rho_C \rightarrow \tau_C$. \square

2. Proof of Proposition IV.2, the continuity of CMI

The proof is based on the continuity of entropy w.r.t. the normalized H-S norm

Lemma V.1 (Continuity of entropy).

$$|S(\rho) - S(\sigma)| \leq \|\rho - \sigma\|_1 \log(d) - \|\rho - \sigma\|_1 \log(\|\rho - \sigma\|_1) \quad (96)$$

$$\leq d \|\rho - \sigma\|_2^2 (\log(d) - \log(d \|\rho - \sigma\|_2^2)). \quad (97)$$

Taking partial trace does not increase the normalized HS norm.

Proposition V.1.

$$d_A \text{Tr}(G_A^2) \leq d_{AB} \text{Tr}(G_{AB}^2). \quad (98)$$

Proof. Decompose orthogonally $G_{AB} = \tau_A \otimes G_B + \{K_A \otimes O_B\}$ such that which are orthogonal

$$d_{AB} \text{Tr}(G_{AB}^2) = d_{AB} \text{Tr}(\tau_A^2 G_B^2 + (K_A \otimes O_B)^2) \geq d_{AB} \text{Tr}(\tau_A^2 G_B^2) = d_B \text{Tr}(G_B^2). \quad (99)$$

\square

We can now prove the continuity of CMI, Proposition IV.2.

Proof. We can expand the CMI as

$$I(A : C|B)_\rho - I(A : C|B)_\sigma \quad (100)$$

$$= (S_\rho(ABC) - S_\sigma(ABC)) - (S_\rho(AB) - S_\sigma(AB)) - (S_\rho(BC) - S_\sigma(BC)) + (S_\rho(B) - S_\sigma(B)). \quad (101)$$

By the continuity of entropy (Lemma V.1), it holds that

$$|(S_\rho(AB) - S_\sigma(AB))| \leq d_{AB} \|G_{AB}\|_2^2 (\log(d) - \log(d_{AB} \|G_{AB}\|_2^2)) \quad (102)$$

$$\leq d_{ABC} \|G_{ABC}\|_2^2 (\log(d) - \log(d_{ABC} \|G_{ABC}\|_2^2)) \quad (103)$$

$$\leq \log(d_{ABC})\epsilon - \log(\epsilon)\epsilon. \quad (104)$$

In the second inequality we used Proposition V.1 and that function $-x \log(x)$ is monotonically increasing at $x \leq 1/e$. The same argument holds for all four terms. \square

3. Proof of Lemma IV.3, spectral characterization of the contraction ratio

Proof. We expand the Petz-recovered map

$$\mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}[\cdot] = \tau^{1/2} \mathcal{E}^\dagger[\mathcal{E}[\tau]^{-1/2} \mathcal{E}[\cdot] \mathcal{E}[\tau]^{-1/2}] \tau^{1/2} \quad (105)$$

$$= \frac{d'}{d} \mathcal{E}^\dagger[\mathcal{E}[\cdot]] \quad (106)$$

where \mathcal{N} being bistochastic substantially simplify the expression, making it self-adjoint and positive. Hence, the spectrum is positive and the second eigenvalue/singular value, expressed by removing the trace coincides with the relative entropy characterization Eq.(94)

$$\lambda_2(\mathcal{P}_{\tau, \mathcal{E}} \mathcal{E}) = \frac{d'}{d} \|\mathcal{E} \mathbb{P}_K\|_{2-2}^2 = \limsup_{\rho_C \rightarrow \tau_C} \frac{D(\mathcal{E}[\rho_C] \|\tau_{C'})}{D(\rho_C \|\tau_C)}. \quad (107)$$

\square

B. Proofs for propositions around Conjecture III.1

1. Proof of Proposition III.3

We prove a slightly more general lemma that include the case when the correctable algebra is not trivial, which immediately converts to Proposition III.3. Some necessary background are at Appendix A.

Lemma V.2. For all tripartite state ρ_{ABC} and channel acting on system C only $\mathcal{E} : C \rightarrow D$

$$I(A : BC)_\rho = I(A : BD)_{\mathcal{E}[\rho]} \iff I(A : B\gamma)_\rho = I(A : BC)_\rho, \quad (108)$$

where $\gamma := \mathcal{A}(\mathcal{E})$ is the correctable algebra of \mathcal{E} , and $I(A : B\gamma)_\rho := I(A : BC)_{E_\gamma(\rho)}$ with $E_\gamma : C \rightarrow \gamma$ the conditional expectation (see e.g., [34]) onto the subalgebra γ ⁶.

In words, if a channel does not decrease the mutual information of the state, then the correlation with system A must be perfectly stored in the correctable algebra $\mathcal{B}(\mathcal{H}_B) \otimes \gamma = \mathcal{A}(\mathcal{I}_B \otimes \mathcal{E}_C)$. Operationally, the LHS implies that from reduced state $E_\gamma(\rho_{ABC})$ there is some recovery channel $R : B\gamma \rightarrow BC$ that recovers the full state ρ_{ABC} .

Proof. We start with the recoverability theorem(see, e.g. [29, Corollary 12.5.1])

$$I(A : BC)_\rho = I(A : BD)_{\mathcal{E}[\rho]} \iff D(\rho_{ABC} \parallel \rho_A \otimes \rho_{BC}) = D(\mathcal{E}_C[\rho_{ABC}] \parallel \rho_A \otimes \mathcal{E}_C[\rho_{BC}]) \quad (109)$$

$$\implies \mathcal{P}_{\rho_A \otimes \rho_{BC}, \mathcal{E}} \circ \mathcal{E}[\rho_{ABC}] = \rho_{ABC} \quad (110)$$

where we are using the Petz map with reference state $\rho_A \otimes \rho_{BC}$, expanded explicitly as follows

$$\mathcal{P}_{\rho_A \otimes \rho_{BC}, \mathcal{E}}[\rho] = \sqrt{\rho_A \otimes \rho_{BC}} \mathcal{E}^\dagger[\sqrt{\mathcal{E}[\rho_A \otimes \rho_{BC}]}^{-1} \rho \sqrt{\mathcal{E}[\rho_A \otimes \rho_{BC}]}^{-1}] \sqrt{\rho_A \otimes \rho_{BC}} \quad (111)$$

$$= \sqrt{\rho_{BC}} \mathcal{E}^\dagger[\sqrt{\mathcal{E}[\rho_{BC}]}^{-1} \rho \sqrt{\mathcal{E}[\rho_{BC}]}^{-1}] \sqrt{\rho_{BC}} = \mathcal{I}_A \otimes \mathcal{P}_{\rho_{BC}, \mathcal{E}}[\rho] \quad (112)$$

The exact form of the Petz map does not mean so much for us here, and all we care is the recovery map $\mathcal{P}_{\rho_A \otimes \rho_{BC}, \mathcal{E}}$ only acts on subsystem BC due to factorization of the reference state $\rho_A \otimes \rho_{BC}$. Therefore we have the invariant subspace equation for $\mathcal{Z}_{BC} = \mathcal{P}_{\rho_{BC}, \mathcal{E}} \circ \mathcal{E}$ and ρ_{ABC} ,

$$\mathcal{I}_A \otimes \mathcal{Z}_{BC}[\rho_{ABC}] = \rho_{ABC}. \quad (113)$$

Suppose the fixed point algebra has factors α ,

$$S_{\mathcal{Z}_{BC}^\dagger} = \bigoplus_{\alpha} B(\mathcal{H}_\alpha) \otimes I_{\bar{\alpha}} \subset B(\mathcal{H}_{BC}). \quad (114)$$

where the α s are dependent on \mathcal{Z}_{BC} , and $\bar{\alpha}$ labels the factors of the commutant of $S_{\mathcal{Z}_{BC}^\dagger}$ and each $\bar{\alpha}$ has one-to-one correspondence with α . Then since $S_{\mathcal{I}_A \otimes \mathcal{Z}_{BC}^\dagger} = \mathcal{B}(\mathcal{H}_A) \otimes S_{\mathcal{Z}_{BC}^\dagger}$, by Theorem A.1, ρ_{ABC} must have the Markovian structure characterized by α s as follows:

$$\rho_{ABC} = \bigoplus_{\alpha} p_\alpha \rho_{A\alpha} \otimes \sigma_{\bar{\alpha}}. \quad (115)$$

To know how α s are embedding in system BC , certainly $S_{\mathcal{Z}_{BC}^\dagger}$ are by definition correctable, i.e. subalgebra of the correctable algebra of $\mathcal{I}_B \otimes \mathcal{E}$. Hence by Proposition A.2, denoting the factors of the correctable algebra $\gamma = \mathcal{A}(\mathcal{E})$ by β ,

$$S_{\mathcal{Z}_{BC}^\dagger} \subset \bigoplus_{\beta} B(\mathcal{H}_{B\beta}) \otimes I_{\bar{\beta}} \quad (116)$$

⁶ Sometimes E_γ is called the restriction to subalgebra (see e.g. [35]). If γ is trivial then this is taking partial trace over C .

Taking the dual, this is saying restricting to subalgebra $B\gamma$ keeps the components $\rho_{A\alpha}$ intact, and thus does not change the mutual information between BC and A ,

$$I(A : B\gamma)_\rho = I(A : BC), \quad (117)$$

completing the proof. \square

We can now prove Proposition III.3, where \mathcal{E} has trivial correctable algebra. Rewriting Lemma V.2,

$$I(A : C|B)_\rho = I(A : BC)_\rho - I(A : B)_\rho > I(A : BD)_{\mathcal{E}[\rho]} - I(A : B)_\rho = I(A : C'|B)_{\mathcal{E}[\rho]} \quad (118)$$

$$\iff I(A : B)_\rho < I(A : BC)_\rho \quad (119)$$

$$\iff I(A : C|B)_\rho > 0 \quad (120)$$

,where the strict inequalities are from taking the negation of equality.

2. Proof of Proposition III.4

The proof uses standard manipulation of CMI:

$$I(A : C_3|B_1B_2B_3) = I(A : B_1B_2B_3C_3) - I(A : B_1B_2B_3) \quad (121)$$

$$\leq I(A : B_1B_2B_3C_3) - I(A : B_1B_2) \quad (122)$$

$$\leq \eta(I(A : B_1B_2C_2) - I(A : B_1B_2)) \quad (123)$$

$$= \eta I(A : C_2|B_1B_2). \quad (124)$$

where in the first inequality we used the monotonicity of the mutual information. The second inequality we used Conjecture III.1 for $\mathcal{E} = \mathcal{N} : C_2 \rightarrow B_3C_3, B = B_1B_2, C = C_2$. For each tripartition separated by m sites, we obtain the exponential decay of CMI by iterating this argument m times. The proof is identical for the trace norm CMI.

3. Proof of Proposition III.2

Proof. It suffice to show \mathcal{N} satisfies the DPI conjecture I.1, using joint convexity of relative entropy

$$\begin{aligned} I(A : BC')_{\mathcal{N}_C[\rho]} &= D(\mathcal{N}[\rho_{ABC}] \| \rho_A \otimes \mathcal{N}[\rho_{BC}]) \\ &= D((1 - \eta)F[\rho_{ABC}] + \eta\mathcal{N}'[\rho_{ABC}] \| (1 - \eta)\rho_A \otimes F[\rho_{BC}] + \eta\rho_A \mathcal{N}'[\rho_{BC}]) \\ &\leq (1 - \eta)D(F[\rho_{ABC}] \| \rho_A \otimes F[\rho_{BC}]) + \eta D(\mathcal{N}'[\rho_{ABC}] \| \rho_A \otimes \mathcal{N}'[\rho_{BC}]) \\ &= (1 - \eta)D(\rho_{AB} \otimes \sigma \| \rho_A \otimes \rho_B \otimes \sigma) + \eta D(\mathcal{N}'[\rho_{ABC}] \| \rho_A \otimes \mathcal{N}'[\rho_{BC}]) \\ &\leq (1 - \eta)D(\rho_{AB} \| \rho_A \otimes \rho_B) + \eta D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) \end{aligned} \quad (125)$$

$$= I(A : B) + \eta(I(A : BC) - I(A : B)). \quad (126)$$

In the first inequality we used the joint convexity of relative entropy, and in the second inequality we used the DPI under \mathcal{N}' . We conclude the proof by moving $I(A : B)$ back to the LHS. The joint convexity holds for the trace-norm CMI as well and the lines are identical. Now we follow the lines in Sec. VB2 to get step-wise decay of CMI and then the decay of CMI for each tripartition. \square

4. Proof of Proposition III.5

Proof. We control the completely bounded superoperator $1 - 1$ norm of $\mathcal{I}_B \otimes \mathcal{N}$ by using the following lemma:

Lemma V.3 ([36, Section 3.11]). *For arbitrary map $\phi : \mathcal{M}_d \rightarrow \mathcal{M}_{d'}$, the completely bounded superoperator $1 - 1$ norm is at most the dimension d times the $1 - 1$ norm.*

$$\|\phi\|_{cb} := \sup_{k, X} \frac{\|(\phi \otimes id_k)[X]\|_1}{\|X\|_1} \leq d \sup_X \frac{\|\phi[X]\|_1}{\|X\|_1} = d\|\phi\|_{1-1}. \quad (127)$$

Then the proof of Prop. III.5 simply follows as

$$\|\mathcal{N}[\rho_{BC} - \rho_B \otimes \rho_C]\|_1 = \left\| (\mathcal{N} - \mathcal{N} \circ (\rho_C Tr_C))[\rho_{BC} - \rho_B \otimes \rho_C] \right\|_1 \quad (128)$$

$$\leq 4\eta d \|\rho_{BC} - \rho_B \otimes \rho_C\|_1, \quad (129)$$

where in the first line we insert a vanishing term $\rho_C Tr_C[\rho_{BC}] - \rho_B \otimes \rho_C = 0$, and in the second line we bounded the completely bounded trace norm using Lemma V.3. \square

C. Remaining proof for Theorem III.2

1. Proof of Lemma IV.4

Proof.

$$\|(\text{id} - \mathbb{P}_\nu) \circ \mathcal{N} \circ (\text{id} - \mathbb{P}_\nu)\|_{cb} \leq 2\|\mathcal{N} - \mathcal{N} \circ \mathbb{P}_\nu\|_{cb} \quad (130)$$

$$\leq 2d_C \|\mathcal{N} - \mathcal{N} \circ \mathbb{P}_\nu\|_{1-1} \quad (131)$$

$$\leq 8d_C \sup_{\rho \in S^+} \|(\mathcal{N} - \mathcal{N} \circ \mathbb{P}_\nu)[\rho]\|_1 \quad (132)$$

$$\leq 8d_C \sup_{\rho \in S^+} \|\mathcal{N}[\rho - \mathbb{P}_\nu \rho]\|_1 \quad (133)$$

$$\leq 16d_C \sup_{\rho, \rho'} \frac{\|\mathcal{N}[\rho] - \mathcal{N}[\rho']\|_1}{\|\rho - \rho'\|_1} \quad (134)$$

$$\leq 16d_C \eta_{1,C} \quad (135)$$

We used $\|\text{id} - \mathbb{P}_\nu\|_{cb} \leq 2$ in the first inequality. The second inequality follows from Lemma V.3. In the third inequality, we convert the optimization over operator $\|X\|_1$ into positive ρ by a factor of 4. This chain of inequalities is largely along the lines of [37, Theorem 45]. \square

2. Proof of Proposition III.1

We get the decay of trace norm CMI $I_1(A : C|B)$ from the continuity of the trace norm CMI.

Lemma V.4 (Continuity of trace norm CMI). *If $\rho_{ABC} = \sigma_{ABC} + G_{ABC}$, and the deviation is traceless on system C , $\text{Tr}_C(G_{ABC}) = 0$, and bounded as $\|G_{ABC}\|_1 \leq \epsilon$, then*

$$|I_1(A : C|B)_\rho - I_1(A : C|B)_\sigma| \leq 2\epsilon. \quad (136)$$

Proof. All we need is the triangle inequality, which makes the analysis much simpler than Lemma IV.2.

$$|I_1(A : C|B)_\rho - I_1(A : C|B)_\sigma| \quad (137)$$

$$= \left| \|\sigma_{ABC} + G_{ABC} - \sigma_A \otimes \sigma_{BC} - \sigma_A \otimes G_{BC}\|_1 - \|\sigma_{AB} - \sigma_A \otimes \sigma_B\|_1 \right| \quad (138)$$

$$- (\|\sigma_{ABC} - \sigma_A \otimes \sigma_{BC}\|_1 - \|\sigma_{AB} - \sigma_A \otimes \sigma_B\|_1) \quad (139)$$

$$\leq \|G_{ABC} - \sigma_A \otimes G_{BC}\|_1 \quad (140)$$

$$\leq \|G_{ABC} - \sigma_A \otimes G_{BC}\|_1 \leq 2\|G_{ABC}\|_1. \quad (141)$$

The C-traceless assumption reduced the expression that $\rho_{AB} = \sigma_{AB}$. □

To show Proposition III.1, choosing $\sigma_{ABC} = \rho_{AB} \otimes \tau_C$ for Theorem III.1, $\sigma_{ABC} = \rho_{AB} \otimes \nu_C$ for Theorem III.2 in the above we complete the proof.

D. Remaining proof of Theorem III.3: decay of B -measured conditional mutual information in MPDO

In subsection VD1, we provide the sufficient background for applying Hilbert's projective metric to show Condition 2 implies exponential decay of CMI; starting from subsection VD3 are the details for $1 \implies 2$.

1. The Hilbert's projective metric and proof of Lemma IV.5

The CP-self maps \mathcal{M}_{s_k} arising from measurement are not trace-preserving, hindering it difficult to approach from typical quantum information tools. It turns out the Hilbert's projective metric is suitable for this purpose, as it is designed to work for the set of all unnormalized states $S_+ := \{\rho \in \mathcal{B}(\mathcal{H}) | \rho \geq 0, \rho \neq 0\}$. While the general theory applies to convex cones, we will focus on the quantum case S_+ , following partly the ideas in [38].

Definition 3 (Hilbert's projective metric).

$$\forall a, b \in S, h(a, b) := \ln(\sup(a/b) \sup(b/a)) \quad (142)$$

$$\sup(a/b) := \inf\{\lambda \in \mathbb{R} | a \leq \lambda b\}. \quad (143)$$

Note that $\sup(a/b) \neq \sup(b/a)$ and the direction of inequality is important in (143). The metric is projective $h(a, b) = h(\alpha a, b)$, and it become a true metric when restricted to set of density operators, i.e., quotient out scalar multiples. This definition works for any *proper cone*, and only implicitly depend on the actual geometry of the cone. In this metric, every positive maps $\mathcal{M} : S_+ \rightarrow S_+$ is contracting⁷.

Theorem V.1 (Birkhoff-Hopf contraction theorem [38, Theorem 4]). *For all $\mathcal{M} : S_+ \rightarrow S_+$, the upper bound on contraction ratio is*

$$\eta_{\mathcal{M}} := \sup_{a, b \in S_+} \frac{h(\mathcal{M}(a), \mathcal{M}(b))}{h(a, b)} = \tanh\left(\frac{\Delta(\mathcal{M})}{4}\right), \quad (144)$$

where $\Delta(\mathcal{M})$ is the projective diameter

$$\Delta(\mathcal{M}) := \sup_{a, b \in S_+} h(\mathcal{M}(a), \mathcal{M}(b)). \quad (145)$$

Note that $\Delta(\mathcal{M}) < \infty$ would imply strict contraction $\eta_{\mathcal{M}} < 1$.

We eventually convert back to the norm via the following bound.

Proposition V.2 ([38, Eq. 38]). *For normalized density operators ρ_1, ρ_2 ,*

$$\frac{1}{2} \|\rho_1 - \rho_2\|_1 \leq \tanh\left(\frac{h(\rho_1, \rho_2)}{4}\right). \quad (146)$$

The above are the backgrounds we need to prove the following lemma.

Lemma V.5. *If all $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_n\}$ are CP maps that map any state to a full rank state, then for arbitrary sequence of $b = s_\ell, \dots, s_1 \in \{1, \dots, n\}^\ell$, it holds that*

$$\left\| \frac{\mathcal{M}_b[\rho_1]}{\text{Tr}(\mathcal{M}_b[\rho_1])} - \frac{\mathcal{M}_b[\rho_2]}{\text{Tr}(\mathcal{M}_b[\rho_2])} \right\|_1 = O(e^{-c\ell}), \quad (147)$$

where the exponent $c > 0$ is independent of b .

Proof. For each \mathcal{M}_s , the distance between any two input states is finite because the image is full rank $h(\mathcal{M}_s(a), \mathcal{M}_s(b)) < \infty$. Hence the projective diameter as a supremum over compact set is also finite, and each \mathcal{M}_s is strictly contracting. Maximizing over $s = 1, \dots, n$ provides a global contraction ratio bound $\eta < 1$.

$$\Delta(\mathcal{M}_s) := \sup_{a, b \in S_+} h(\mathcal{M}_s(a), \mathcal{M}_s(b)) < \infty \quad (148)$$

$$\forall s, \eta_{\mathcal{M}_s} = \tanh\left(\frac{\Delta(\mathcal{M}_s)}{4}\right) \leq \eta < 1. \quad (149)$$

⁷ In fact the unique metric for this to hold [39]

Then for all $a, b \in S_+$

$$h(\mathcal{M}_{s_\ell, \dots, s_1}(a), \mathcal{M}_{s_\ell, \dots, s_1}(b)) \leq h(\mathcal{M}_{s_1}(a), \mathcal{M}_{s_1}(b))e^{-c(\ell-1)} \leq \sup_s (\Delta(\mathcal{M}_s)) e^{-c(\ell-1)} = O(e^{-c\ell}), \quad (150)$$

where again we used the fact the projective diameter is finite. Finally we convert to the trace norm

$$\left\| \frac{\mathcal{M}_b[\rho_1]}{\text{Tr}(\mathcal{M}_b[\rho_1])} - \frac{\mathcal{M}_b[\rho_2]}{\text{Tr}(\mathcal{M}_b[\rho_2])} \right\|_1 \leq 2 \tanh\left(\frac{h(\mathcal{M}_b[\rho_1], \mathcal{M}_b[\rho_2])}{4}\right) \quad (151)$$

$$\leq O(e^{-c\ell}), \quad (152)$$

where $\tanh(x) \approx x$ for small x . \square

2. Proof of Lemma IV.6

Lemma V.6. *Suppose the CP map $\mathcal{M} : C' \rightarrow C$ is contractive in the sense that*

$$\left\| \frac{\mathcal{M}_b[\rho_1]}{\text{Tr}(\mathcal{M}_b[\rho_1])} - \frac{\mathcal{M}_b[\rho_2]}{\text{Tr}(\mathcal{M}_b[\rho_2])} \right\|_1 \leq \epsilon. \quad (153)$$

Then the state $\rho_{AC} := \mathcal{M}[\sigma_{AC'}]$ is closed to the product state

$$2T_b := \|\rho_{AC,b} - \rho_{A,b} \otimes \rho_{C,b}\|_1 \leq 4d_{C'} \|\sigma_{C'}^{-1}\|_\infty \epsilon. \quad (154)$$

Proof. Bounded factors depending on the hidden system C' may show up here and there, but it pose no threat when the error ϵ is exponentially small.

$$2T_b := \|\rho_{AC,b} - \rho_{A,b} \otimes \rho_{C,b}\|_1 \quad (155)$$

$$= \left\| \frac{\mathcal{M}_b[\sigma_{AC'}]}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} - \frac{\text{Tr}_C(\mathcal{M}_b[\sigma_{AC'}])}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \otimes \frac{\mathcal{M}_b[\sigma_{C'}]}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \right\|_1 \quad (156)$$

$$\leq \left\| \frac{\mathcal{M}_b[\cdot]}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} - \frac{\text{Tr}_C(\mathcal{M}_b[\cdot])}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \otimes \frac{\mathcal{M}_b[\sigma_{C'}]}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \right\|_\diamond \quad (157)$$

$$\leq \frac{d_A}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \sup_{\|X\|_1 \leq 1, X \subset \text{supp}(\sigma_{C'})} \left\| \mathcal{M}_b[X] - \text{Tr}(\mathcal{M}_b[X]) \frac{\mathcal{M}_b[\sigma_{C'}]}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \right\|_1 \quad (158)$$

$$\leq \frac{4d_A}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \sup_{\rho_1} \left\| \mathcal{M}_b[\rho_1] - \text{Tr}(\mathcal{M}_b[\rho_1]) \frac{\mathcal{M}_b[\sigma_{C'}]}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \right\|_1 \quad (159)$$

$$\leq 4d_{C'} \frac{\sup_{\rho_1 \subset \text{supp}(\sigma_{C'})} (\text{Tr}(\mathcal{M}_b[\rho_1]))}{\text{Tr}(\mathcal{M}_b[\sigma_{C'}])} \epsilon \quad (160)$$

$$\leq 4d_{C'} \|\sigma_{C'}^{-1}\|_\infty \epsilon, \quad (161)$$

where in the second and third inequality we used that diamond norm is bounded by d times the superoperator-norm (Lemma V.3) and the extra factor of 4 comes from turning X into density operator ρ_1 ; in the fourth inequality we used the assumption; in the last inequality we used that $\|\sigma_{C'}^{-1}\|_\infty \sigma_{C'} \geq I \geq \rho_1$. Note that we are working on the support of $\sigma_{C'}$ so that the inverse is well-defined. \square

2 \implies 3 by setting $\sigma_{AC'}$ to be the maximally entangled state on $A\bar{A}$, which yields $\|\rho_{AC,b} - \rho_{A,b} \otimes \rho_{C,b}\|_1 \leq 4d_A^2 \mathcal{O}(e^{-c\ell})$. The conversion to mutual information is straightforwardly follows from the AFW inequality,

$$I(A : C)_b \leq 2T_b \log(\min(d_A, d_C)) + (1 + T_b) \log(1 + T_b) - T_b \log(T_b) = O(e^{-c_1\ell}). \quad (162)$$

This immediately pass to CMI by taking expectation in Eq. (82) and thus completes the proof.

3. Proof of Lemma IV.7

Lemma V.7. *If $\text{span}\{K_1, \dots, K_p\} = \text{Mat}(D, D)$, then $\mathcal{N}[\rho] := \sum_i K_i \rho K_i^\dagger > 0, \forall \rho$.*

Proof. Suppose there exists $|u\rangle, |v\rangle$, s.t. $\sum_j \langle v | K_j | u \rangle \langle u | K_j^\dagger | v \rangle = 0$. Then we obtain

$$|\langle v | K_i | u \rangle|^2 = 0, \forall i \quad (163)$$

$$\implies \text{Tr}(K_i |u\rangle \langle v|) = 0, \forall i. \quad (164)$$

This is a contradiction because $\{K_i\}$ are full dimensional. \square

4. Proof of Lemma IV.8

The idea is rooted from a theorem of Burnside about algebra generated by matrices and the simultaneous invariant subspace.

Theorem V.2 (Burnside [40]). *Consider $m_1, \dots, m_{p-1} \in \text{Mat}_{\mathbb{C}}(D, D)$, acting on vector space \mathbb{C}^D . The following are equivalent.*

1. *The algebra generated by (m_1, \dots, m_{p-1}) is the full matrix algebra $\text{Mat}_{\mathbb{C}}(D, D)$*
2. *For all non-trivial subspace $V \subset \mathbb{C}^D$,*

$$m_q V \subset V, \forall q \implies V = \mathbb{C}^D, \quad (165)$$

i.e., the simultaneous invariant subspace of m_1 and m_2 is trivial or the whole space.

In our version, the vector space is $\text{Mat}_{\mathbb{C}}(D, D)$, where Kraus operators live :

Proposition V.3. *Consider $m_1, \dots, m_{p-1} \in \text{Mat}_{\mathbb{C}}(D, D)$, acting on vector space $\text{Mat}_{\mathbb{C}}(D, D)$ by left multiplication. The following are equivalent.*

1. *The algebra generated by (m_1, \dots, m_{p-1}) is the full matrix algebra $\text{Mat}_{\mathbb{C}}(D, D)$*

2. For all non-trivial subspace (as vector space) $W \subset \text{Mat}_{\mathbb{C}}(d, d)$ containing an invertible element B ,

$$m_q W \subset W, \forall q \implies W = \text{Mat}_{\mathbb{C}}(D, D).$$

Proof. (\implies) For all non-trivial subspace $W \in \text{Mat}_{\mathbb{C}}(D, D)$, if $m_q W \subset W, \forall q$, then same is true under left multiplication

$$(m_1, \dots, m_{p-1})W \subset W. \quad (166)$$

Then in particular W contains $(m_1, \dots, m_{p-1})B$, thus must be the full matrix algebra $W = \text{Mat}_{\mathbb{C}}(D, D)$.

(\Leftarrow) For a contradiction, suppose $(m_1, \dots, m_{p-1}) \neq \text{Mat}_{\mathbb{C}}(D, D)$. Then by theorem V.2, (m_1, \dots, m_{p-1}) is reducible with an proper invariant subspace $V \subsetneq \mathbb{C}^D$. Consider the basis for which the first entries are basis vectors spanning V , i.e. (m_1, \dots, m_{p-1}) has some zeros at the left down corner:

$$(m_1, \dots, m_{p-1}) = \begin{bmatrix} M_V & N \\ 0 & M_{V^c} \end{bmatrix}.$$

Consider the subspace $W' := (m_1, \dots, m_{p-1}) + \{\lambda I\}$ by adding the identity, which is invertible.

Then the resulting subspace W' is an invariant subspace

$$m_q W' = m_q(m_1, \dots, m_{p-1}) + m_q \subset W'. \quad (167)$$

$$(168)$$

We arrive at contradiction with $W' = (m_1, \dots, m_{p-1}) + \{\lambda I\} \neq \text{Mat}_{\mathbb{C}}(D, D)$. \square

We can now prove Lemma IV.8:

Lemma V.8 (Condition 1 implies the increment of span of Kraus operators). *Consider a CP self-map \mathcal{M} with $p > 2$ Kraus operators $\{E_1, \dots, E_p\}$ such that E_p is invertible, and $((E_p)^{-1}E_1, \dots, (E_p)^{-1}E_{p-1})$ generate full matrix algebra $\text{Mat}_{\mathbb{C}}(D, D)$ by addition and multiplication. Then for any set of Kraus operators $\{T_j\}$ not full rank and containing an invertible element, the dimension of span must increase after applying \mathcal{M}*

$$\dim(\text{span}\{E_i T_j\}) > \dim(\text{span}\{T_j\}). \quad (169)$$

Proof. Left multiplying E_i yield the span

$$\{E_1 T_j, \dots, E_p T_j\}.$$

Notice that $\dim(\{T_j\}) = \dim(E_p\{T_j\})$ because E_p is invertible, so if any of $E_1\{T_j\}, \dots, E_{p-1}\{T_j\}$ is not a subspace of $E_p\{T_j\}$, then the dimension would increase $\dim(\text{span}\{E_i T_j\}) > \dim(\text{span}\{T_j\})$; we only need to worry if

$$E_1\{T_j\}, \dots, E_{p-1}\{T_j\} \subset E_p\{T_j\}. \quad (170)$$

Invert E_p and by Proposition V.3, $\{T_j\}$ must reach full rank already.

$$\{T_j\} = \text{Mat}_{\mathbb{C}}(D, D). \quad (171)$$

□

5. Proof of Proposition III.6

It is simple to generate an algebra from set of matrices m_1, \dots, m_{p-1} . Start with $S = \text{span}\{m_1\}$, repeat the following steps:

1. Add left multiplied matrices to the set, resulting $S' = \{S, m_1 S, \dots, m_{p-1} S\}$.
2. Find a linear basis for S' .⁸ If $\dim(S') = \dim(S)$, then terminate, and $\text{span}\{S'\}$ is the algebra generated by m_1, \dots, m_{p-1} . If $\dim(\text{span}\{S'\}) = D^2$, then it is the full matrix algebra.

Checking linear independence uses polynomial runtime.

VI. CONCLUSION AND DISCUSSIONS

Motivated by the problem of showing the existence of the local parent Hamiltonians of MPDO, we have studied the CMI of MPDO. We have shown that MPDO constructed by bistochastic Y-Shaped channels with trivial correctable algebra have exponentially decaying CMI and thus have approximately local parent Hamiltonians. We have shown a similar bound for a slightly general class of channels under the restriction that certain decay constants are sufficiently small. We have also shown the exponential decay of CMI for Y-shaped channels with a forgetful component. We have introduced a trace norm variant of the CMI and have shown that for the above cases they obey no worse bounds than the CMI.

For more general Y-shaped channels, we have conjectured the completely contractive DPI (Conjecture III.1). We have shown that if this conjecture is true, every MPDO constructed by a Y-shaped channel with trivial correctable algebra has exponentially decaying CMI. For the measured MPDO, we have provided sufficient conditions implying the exponential decay of CMI. We have numerically confirmed

⁸ In the actual code, there always need to be an small error threshold to decide the linear independence.

(up to small bond dimension) that these conditions are generically true if the Y-shaped channel is generated by a Haar-random unitary.

Our results Theorem III.1 and Theorem III.2 only work for a restricted family of Y-shaped channels. The proof relies on the fact that the corresponding MPDO are approximately a product state. This is no longer true for general channels and the deviation from the product state do not obviously contract⁹. A possible solution to avoid the structure of many-body state is resorting to our Conjecture III.1. Note that having trivial correctable algebra is still only a sufficient condition, and a necessary and sufficient condition for exponentially decaying CMI is unclear yet.

Analysis of the CMI for the measured MPDO could be massively easier than the unmeasured case due to losing entanglement with B . However, our results on measured MPDO are still limited and only provide sufficient conditions. In contrast to the single channel quantum Wielandt's inequality [31] with sufficient and necessary conditions, Condition 1 in Theorem III.3 guarantees strict positivity for all sequences multiplicatively generated by a finite set of CP-maps $\{\mathcal{M}_s\}$. Quantum Wielandt's inequality has a classical analog in matrix theory, however this multiple-channel generalization has limited results even in the classical case (see e.g., [42] for a related result).

VII. ACKNOWLEDGEMENT

We thank Jean-Francois Quint for comments on multiplicative ergodic theory. We thank Mario Berta, Marco Tomamichel, Hao-Chung Cheng for discussions about the DPI for CMI. CFC is thankful for Physics TA Relief Fellowship and the Physics TA Fellowship at Caltech. KK acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1733907) and MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant Number JPMXS0120319794. FB acknowledges funding from NSF.

-
- [1] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki. Rigorous results on valence-bond ground states in antiferromagnets. *Phys. Rev. Lett.*, 59(7), 1987.
 - [2] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix product state representations. *Quantum Inf. Comput.*, 7(5-6), 2007.
 - [3] M. B. Hastings. An area law for one-dimensional quantum systems. *J. Stat. Mech.-Theory. E.*, 2007(08), 2007.
 - [4] M. Fannes, B. Nachtergaele, and R. F. Werner. Finitely correlated states on quantum spin chains. *Commun. Math. Phys.*, 144(3), 1992.

⁹ Bistochastic channels are rather special that it contracts all p -norm. See [41] to see how this fails for non-bistochastic channels.

- [5] N. Schuch, D. Perez-García, and J. I. Cirac. Classifying quantum phases using matrix product states and projected entangled pair states. *Phys. Rev. B*, 84(16), 2011.
- [6] F. Verstraete and J. I. Cirac. Renormalization algorithms for quantum-many body systems in two and higher dimensions, arXiv:cond-mat/0407066, 2004.
- [7] A. Molnar, Y. Ge, N. Schuch, and J. I. Cirac. A generalization of the injectivity condition for projected entangled pair states. *J. Math. Phys.*, 59(2), 2018.
- [8] M. B. Hastings. Solving gapped Hamiltonians locally. *Phys. Rev. B*, 73(8), 2006.
- [9] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac. Matrix product density operators: Simulation of finite-temperature and dissipative systems. *Phys. Rev. Lett.*, 93(20), 2004.
- [10] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete. Matrix product density operators: Renormalization fixed points and boundary theories. *Ann. Phys.*, 378, 2017.
- [11] K. Kato and F. G. S. L. Brandão. Quantum approximate markov chains are thermal. *Commun. Math. Phys.*, 370(1):117–149, 2019.
- [12] M. Kliesch, D. Gross, and J. Eisert. Matrix-product operators and states: Np-hardness and undecidability. *Phys. Rev. Lett.*, 113:160503, 2014.
- [13] H. Araki. Gibbs states of a one dimensional quantum lattice. *Commun. Math. Phys.*, 14(2), 1969.
- [14] G. Lindblad. Completely positive maps and entropy inequalities. *Commun. Math. Phys.*, 40(2):147–151, 1975.
- [15] A. Uhlmann. Relative entropy and the wigner-yanase-dyson-lieb concavity in an interpolation theory. *Commun. Math. Phys.*, 54(1):21–32, 1977.
- [16] D. Petz. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Commun. Math. Phys.*, 105(1):123–131, 1986.
- [17] D. Petz. Sufficiency of channels over Von Neumann algebras. *Q. J. Math.*, 39(1), 1988.
- [18] V. Anantharam, A. Gohari, S. Kamath, and C. Nair. On hypercontractivity and a data processing inequality. *2014 IEEE International Symposium on Information Theory*, 3022-3026, 2014.
- [19] G. E. D.-Uchida. *Carathéodory theorem*. Springer US, Boston, MA, 2001.
- [20] S. Beigi and A. Gohari. On dimension bounds for quantum systems. In *2013 Iran Workshop on Communication and Information Theory*, 1–6, 2013.
- [21] J.M. Hammersley and P.E. Clifford. Markov field on finite graphs and lattices. <http://www.statslab.cam.ac.uk/grg/books/hammfest/hamm-cliff.pdf>, 1971.
- [22] P. Hayden, R. Jozsa, D. Petz, and A. Winter. Structure of States Which Satisfy Strong Subadditivity of Quantum Entropy with Equality. *Commun. Math. Phys.*, 246(2), 2004.
- [23] W. Brown and D. Poulin. Quantum Markov Networks and Commuting Hamiltonians. arXiv:1206.0755 2012.
- [24] C. Bény, A. Kempf, and D. W. Kribs. Generalization of quantum error correction via the Heisenberg picture. *Phys. Rev. Lett.*, 98(10), 2007.
- [25] C. Bény, A. Kempf, and D. W. Kribs. Quantum error correction of observables. *Phys. Rev. A*, 76(4), 2007.
- [26] E. Knill, R. Laflamme, and L. Viola. Theory of quantum error correction for general noise. *Phys. Rev. Lett.*, 84:2525–2528, 2000.
- [27] 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(10):2955–2978, Aug 2018.
- [28] D. Petz. sufficiency of channels over von neumann algebras. *The Quarterly Journal of Mathematics*, 39(1):97–108, 03 1988.
- [29] M. M. Wilde. From classical to quantum shannon theory, 2011.
- [30] K. Kato and F. G. S. L. Brandão. Toy model of boundary states with spurious topological entanglement entropy. *Phys. Rev. Research*, 2:032005, Jul 2020.
- [31] M. Sanz, D. Pérez-García, M. M. Wolf, and J. I. Cirac. A quantum version of wielandt’s inequality. *IEEE Transactions on Information Theory*, 56(9):4668–4673, 2010.
- [32] S. Gaubert and Z. Qu. Checking the strict positivity of Kraus maps is np-hard. arXiv:1402.1429, 2014.
- [33] A. Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints, *Commun. Math. Phys.* 347(1):291-313, 2016.
- [34] M. Takesaki. *Theory of operator algebras. I*, volume 124 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2002.
- [35] C.-F. Chen, G. Penington, and G. Salton. Entanglement wedge reconstruction using the petz map. *JHEP*, 2020(1), Jan 2020.
- [36] V. Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2003.
- [37] M. James. Quantum markov chain mixing and dissipative engineering. PhD thesis, 2015.
- [38] D. Reeb, M. J. Kastoryano, and M. M. Wolf. Hilbert’s projective metric in quantum information theory. *J. Math. Phys.*, 52(8):082201, 2011.
- [39] E. Kohlberg and J. W. Pratt. The Contraction Mapping Approach to the Perron-Frobenius Theory: Why Hilbert’s Metric? *Math. Oper. Res.*, 7(2):198–210, 1982.
- [40] V. Lomonosov and P. Rosenthal. The simplest proof of Burnside’s theorem on matrix algebras. *Linear Algebra Appl.*, 383:45 – 47, 2004. Special Issue in honor of Heydar Radjavi.
- [41] D. Pérez-García, M. M. Wolf, D. Petz, and M. B. Ruskai. Contractivity of positive and trace-preserving maps under L_P norms. *J. Math. Phys.*, 47(8):083506, 2006.
- [42] V. Y. Protasov and A.S. Voynov. Sets of nonnegative matrices without positive products. *Linear. Algebra. Appl.*, 437(3):749 – 765, 2012.
- [43] M. Koashi and N. Imoto. Operations that do not disturb partially known quantum states. *Phys. Rev. A*, 66:022318, Aug 2002.
- [44] G. Lindblad. A general no-cloning theorem. *Lett. Math. Phys.*, 47(2):189–196, 1999.
- [45] C. Hirche, C. Rouzé, and D. S. França On contraction coefficients, partial orders and approximation of capacities for quantum channels. arxiv:2011.05949, 2020
- [46] https://github.com/Shawnger/supplementary_MPDO

Appendix A: Equivalence between correctable algebra and the Petz recovered map for bistochastic channels

The structure of CPTP self-map, the invariant subspace, and the fixed-point algebra of the dual has been studied.

Theorem A.1 (Combination of [22, 43, 44]). *For every CPTP self-map $\mathcal{Z} : S(\mathbb{C}^d) \rightarrow S(\mathbb{C}^d)$, then the following holds:*

1. *The invariant subspace $S_{\mathcal{Z}^\dagger}$ of \mathcal{Z}^\dagger forms an subalgebra $M \subset B(\mathcal{H}_d)$, with factors α .*

$$S_{\mathcal{Z}^\dagger} = \bigoplus_{\alpha} B(\mathcal{H}_{\alpha}) \otimes I_{\bar{\alpha}} \quad (\text{A1})$$

2. *The invariance subspace of $S_{\mathcal{Z}}$ of \mathcal{Z} has form*

$$S_{\mathcal{Z}} = \bigoplus_{\alpha} p_{\alpha} \rho_{\alpha} \otimes \sigma_{\bar{\alpha}} \quad (\text{A2})$$

with $\sigma_{\bar{\alpha}}$ determined by \mathcal{Z} , and $\rho_{\alpha}, p_{\alpha}$ are free.

3. *The channel \mathcal{Z} restricted on the block diagonal entries has form (its acting on off-diagonal part is not as simple)*

$$\mathcal{Z} = \bigoplus_{\alpha} \mathcal{I}_{\alpha} \otimes \mathcal{Z}_{\bar{\alpha}} \quad (\text{A3})$$

and $\mathcal{Z}_{\bar{\alpha}}$ has unique fixed point σ_{α}

Though, when a channel has different input-output dimension, we alternatively consider the correctable algebra $\mathcal{A}(\mathcal{E})$. It turned out corresponds to the invariant subspace of $Z = \mathcal{P}_{\tau, \mathcal{E}} \mathcal{E}$, the original channel composed with the Petz recovery map with the maximally mixed state as the reference.

Proposition A.1 (Recap of Proposition II.1). *The correctable algebra of a channel \mathcal{E} equals to the fixed-point algebra of the Petz-recovery channel composed with channel $\mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}$ (and the dual $\mathcal{E}^\dagger \circ \mathcal{P}_{\tau, \mathcal{E}}^\dagger$, due to being self-adjoint).*

$$S_{\mathcal{E}^\dagger \circ \mathcal{P}_{\tau, \mathcal{E}}^\dagger} = S_{\mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}} = \mathcal{A}(\mathcal{E}) \quad (\text{A4})$$

Proof. By [35, Theorem 1], the Petz map with the maximally mixed reference state is a universal subalgebra recovery map, i.e. it recovers any subalgebra that can be recovered from any channel D .

$$D \circ \mathcal{E}(\rho)|_a = \rho_a, \forall \rho \implies \mathcal{P}_{\tau, \mathcal{E}} \circ \mathcal{E}(\rho)|_a = \rho_a, \forall \rho \quad (\text{A5})$$

The invariant subspace of $\mathcal{E}^\dagger \circ \mathcal{P}_{\tau, \mathcal{E}}^\dagger$ contains the correctable algebra, and the converse is by definition true.

$$\mathcal{A}(\mathcal{E}) \subset S_{\mathcal{E}^\dagger \circ \mathcal{P}_{\tau, \mathcal{E}}^\dagger} \subset \mathcal{A}(\mathcal{E}) \quad (\text{A6})$$

We conclude the proof by that the Petz-recovered channel is self-adjoint w.r.t. to the H.S. norm

$$\mathcal{E}^\dagger \circ \mathcal{P}_{\tau, \mathcal{E}}^\dagger = \mathcal{E}^\dagger \circ [\mathcal{E}[\tau]^{-1/2} \tau \mathcal{E}[\cdot] \mathcal{E}[\tau]^{-1/2}] = \tau \mathcal{E}^\dagger [\mathcal{E}[\tau]^{-1/2} \mathcal{E}[\cdot] \mathcal{E}[\tau]^{-1/2}] \quad (\text{A7})$$

$$= \mathcal{P}_{\tau, \mathcal{E}} \mathcal{E}. \quad (\text{A8})$$

□

This provides an alternative proof that the correctable algebra behaves nicely when tensored with auxiliary system B .

Proposition A.2. *Suppose the correctable algebra is*

$$\mathcal{A}(\mathcal{E}) := \gamma = \bigoplus_{\beta} B(\mathcal{H}_{\beta}) \otimes I_{\bar{\beta}} \subset B(\mathcal{H}_C) \quad (\text{A9})$$

then

$$\mathcal{A}(\mathcal{I}_B \otimes \mathcal{E}) = \bigoplus_{\beta} B(\mathcal{H}_{B\beta}) \otimes I_{\bar{\beta}} = B(\mathcal{H}_B) \otimes \gamma \quad (\text{A10})$$

Proof. By Proposition II.1, we get

$$\mathcal{A}(\mathcal{I}_B \otimes \mathcal{E}) = S_{\mathcal{I}_B \otimes \mathcal{E}^\dagger \circ \mathcal{P}_{\tau_{BC}, \mathcal{I}_B \otimes \mathcal{E}}^\dagger} = S_{\mathcal{I}_B \otimes (\mathcal{E}^\dagger \circ \mathcal{P}_{\tau_C, \mathcal{E}}^\dagger)} \quad (\text{A11})$$

where we can use structure theorem A.1 for $\mathcal{Z} := \mathcal{P}_{\tau_C, \mathcal{E}} \circ \mathcal{E}$

$$\mathcal{I}_B \otimes (\mathcal{E}^\dagger \circ \mathcal{P}_{\tau_C, \mathcal{E}}^\dagger) = \mathcal{I}_B \otimes \bigoplus_{\beta} \mathcal{I}_{\beta} \otimes \mathcal{Z}_{, \bar{\beta}} \quad (\text{A12})$$

$$= \bigoplus_{\beta} \mathcal{I}_{B\beta} \otimes \mathcal{Z}_{, \bar{\beta}} \quad (\text{A13})$$

The invariant subspace of $\mathcal{I}_B \otimes \mathcal{Z}^\dagger$ coincide with Eq.(A9) □

Appendix B: Structure of periodic MPDO generated by Y-shaped bistochastic channels.

Theorem B.1. *Consider the following MPDO constructed by closed loop of Y-shaped channels*

$$\rho_{B_1 \dots B_\ell} := \sum_{i, j} \text{Tr}(|j\rangle\langle i| \mathcal{N} \circ \dots \circ \mathcal{N}[|i\rangle\langle j|]). \quad (\text{B1})$$

Then the state is close to the maximally mixed state

$$\rho_{B_1 \dots B_\ell} = \tau_{B_1 \dots B_\ell} + O_{B_1 \dots B_\ell} \quad (\text{B2})$$

up to an exponentially small global operator

$$\|O_{B_1 \dots B_\ell}\|_1 = \mathcal{O}(\eta^\ell), \quad (\text{B3})$$

which implies the decay of the CMI for any tripartition $A'B'C'$

$$I(A' : C'|B') = \mathcal{O}(\ell\eta^\ell). \quad (\text{B4})$$

Where the contraction ratio η is given as

$$\eta := \limsup_{\rho_C \rightarrow \tau_C} \frac{D(\mathcal{N}[\rho_C] \parallel \tau_{C'})}{D(\rho_C \parallel \tau_C)}. \quad (\text{B5})$$

Proof. We first break the loop by rewriting

$$\rho_{B_1 \dots B_\ell} = \sum_{i,j} \text{Tr}(|j\rangle\langle i| \mathcal{N} \circ \dots \mathcal{N}[|i\rangle\langle j|]) \quad (\text{B6})$$

$$= \sum_{O_{ij}} \text{Tr}(O_{ij}^\dagger \mathcal{N} \circ \dots \mathcal{N}[O_{ij}]), \quad (\text{B7})$$

where O_{ij} are any complete basis of operators orthonormal w.r.t. the Hilbert Schimdt inner product. Then recall the decomposition of $\mathcal{N} \circ \dots \mathcal{N}$ as in the open boundary case

$$\mathcal{N} \circ \dots \mathcal{N}[O] = (1 - \mathbb{P}_K) \mathcal{N}[O] \otimes \tau_3 \dots \tau_\ell \quad (\text{B8})$$

$$+ (1 - \mathbb{P}_K) \mathcal{N} \mathbb{P}_K \mathcal{N}[O] \otimes \tau_4 \dots \tau_\ell \quad (\text{B9})$$

$$+ \dots \quad (\text{B10})$$

$$+ \mathbb{P}_K \mathcal{N} \dots \mathbb{P}_K \mathcal{N} \mathbb{P}_K \mathcal{N}[O]. \quad (\text{B11})$$

Plugging into Eq. (B7), and choose O_{ij} to split into the maxmially mixed I/\sqrt{d} and the basis for orthogonal complement $\{K_n\}$.

$$\sum_{O_{ij}} \text{Tr}(O_{ij}^\dagger \mathcal{N} \circ \dots \mathcal{N}[O_{ij}]) = \sum_n \text{Tr}(K_n^\dagger \mathcal{N} \circ \dots \mathcal{N}[K_n]) \quad (\text{B12})$$

$$+ d \text{Tr}(\tau \mathcal{N} \circ \dots \mathcal{N}[\tau]) \quad (\text{B13})$$

$$= \tau_{B_1 \dots B_\ell} + \sum_n \text{Tr}(K_n^\dagger \mathcal{N} \mathbb{P}_K \circ \dots \mathbb{P}_K \mathcal{N}[K_n]), \quad (\text{B14})$$

where the sum over traceless K_n contains a single term since other terms has τ_ℓ which is orthogonal to K_n . When \mathcal{N} has trivial correctable algebra, by Lemma IV.1 the global operator is exponentially decaying in the normalized H-S norm and hence the trace norm using Cauchy-Schwartz inequality. Finally by continuity of the CMI (Lemma IV.2), each tripartition must have CMI exponentially small w.r.t to total length ℓ . \square