Markovianizing Cost of Tripartite Quantum States
Eyuri Wakakuwa, Akihito Soeda and Mio Murao

Abstract—We introduce and analyze a task that we call Markovianization, in which a tripartite quantum state is transformed to a quantum Markov chain by a randomizing operation on one of the three subsystems. We consider cases where the initial state is the tensor product of \(n\) copies of a tripartite state \(\rho^{ABC}\), and is transformed to a quantum Markov chain conditioned by \(B^n\) with a small error, using a random unitary operation on \(A^n\). In an asymptotic limit of infinite copies and vanishingly small error, we analyze the Markovianizing cost, that is, the minimum cost of randomness per copy required for Markovianization. For tripartite pure states, we derive a single-letter formula for the Markovianizing costs. Counterintuitively, the Markovianizing cost is not a continuous function of states, and can be arbitrarily large even if the state is close to a quantum Markov chain. Our results have an application in analyzing the cost of resources for simulating a bipartite unitary gate by local operations and classical communication.

I. INTRODUCTION

Tripartite quantum states for which the quantum conditional mutual information (QCMI) is zero are called short quantum Markov chains, or Markov states for short [1]. They play important roles, e.g., in analyzing the cost of quantum state redistribution [2], [3], investigating effects of the initial system-environment correlation on the dynamics of quantum states [4], and computing the free energy of quantum many-body systems [5].

In analogy to the quantum mutual information (QMI) of a bipartite state quantifying a distance to the closest product states, it would be natural to expect a similar relation between QCMI of a tripartite state and Markov states. However, this conjecture has been falsified [6] (see also [7], [8]). The recent results show that the relation between QCMI and Markov states is not so straightforward [6]–[10], particularly when compared to the relation between QMI and product states.

From an operational point of view, QMI quantifies the minimum cost of randomness required for destroying the correlation between two quantum systems in an asymptotic limit of infinite copies [11]. This fact and its variants including single-shot cases are called decoupling theorems, and have played a significant role in the development of quantum information theory for a decade [12]–[17]. In a simple analogy, one may ask the following question: Is QCMI equal to the minimum cost of randomness required for transforming a tripartite state to a Markov state?

In this paper, we address this question, and answer in the negative. We derive a single-letter formula for the “Markovianizing cost” of pure states, that is, the minimum cost of randomness per copy required for Markovianizing tripartite pure states in the asymptotic limit of infinite copies. The obtained formula is not equal to QCMI, or not even a continuous function of states. Moreover, the Markovianizing cost of a state can be arbitrarily large, regardless of how close the state is to a Markov state. In the proof, we improve a random coding method using the Haar distributed random unitary ensemble, which is widely used in the proof of the decoupling theorems, by incorporating the mathematical structure of Markov states.

There are two ways for defining the property of tripartite quantum states being “approximately Markov”: one by the condition that the state is close to a Markov state, on which our definition of Markovianization in this paper is based; and the other by the condition that the state is approximately recoverable [10], i.e., there exists a quantum operation \(\mathcal{E} : B \rightarrow BC\) such that \(\rho^{ABC} \approx \mathcal{E}(\rho^{AB})\). Ref. [10] proved that the latter condition has a direct connection with QCMI, namely, small QCMI implies recoverability with a small error.

In [18], we introduce another formulation of the Markovianizing cost by employing the concept of recoverability, and prove that the cost function is equal to the one obtained in this paper for pure states. We then apply the results in analyzing the cost of entanglement and classical communication for simulating a bipartite unitary gate by local operations and classical communication [19]. As a consequence, we prove in [20] that there is a trade-off relation between the entanglement cost and the number of rounds of communication for a two-party distributed quantum information processing.

The structure of this paper is as follows. In Section II we review mathematical theorems regarding the structure of quantum Markov chains, which are extensively used in this paper. In Section III we introduce the formal definition of Markovianization, and describe the main results. Outlines of proofs of the main results are presented in Section IV. In Section V we describe properties of the Markovianizing cost. In Section VI we calculate the Markovianizing cost of particular classes of tripartite pure states to illustrate its properties. Conclusions are given in Section VII. See Appendices for detailed proofs.

Notations. A Hilbert space associated with a quantum system \(A\) is denoted by \(\mathcal{H}^A\), and its dimension is denoted by \(d_A\). For \(\rho \in \mathcal{S}(\mathcal{H}^A)\), we denote \(\text{supp}[\rho] \subseteq \mathcal{H}^A\) by \(\mathcal{H}_\rho^A\). A system composed of two subsystems \(A\) and \(B\) is denoted by...
AB. When \( M \) and \( N \) are linear operators on \( \mathcal{H}^A \) and \( \mathcal{H}^B \), respectively, we denote \( M \otimes N \) as \( M^A \otimes N^B \) for clarity. We abbreviate \(|\psi\rangle^A \otimes |\varphi\rangle^B\) as \(|\psi\rangle^A |\varphi\rangle^B\). The identity operator on a Hilbert space is denoted by \( I \). We denote \((M^A \otimes I^B)|\psi\rangle^A \otimes |\varphi\rangle^B\) as \(M^A|\psi\rangle^A \otimes |\varphi\rangle^B\), and \((M^A \otimes I^B)\rho^{AB}(M^A \otimes I^B)^\dagger\) as \(M^A \rho^{AB} M^A \). We abbreviate \(M^A B(\rho^A \otimes I^B)M^{1A B}\) as \(M^{AB} \rho^{A} M^{1AB}\). When \( E \) is a quantum operation on \( A \), we denote \((E \otimes \text{id}^B)(\rho^{AB})\) as \((E^A \otimes \text{id}^B)(\rho^A)\) or \(E(\rho^A)\). For \( \rho^{AB} \), \( \rho^A \) represents \(\text{Tr}_B[\rho^{AB}]\). We denote \(|\psi\rangle\langle\psi|\) simply as \(\rho^A\).

A system composed of \( n \) identical systems of \( A \) is denoted by \( A^n \) or \( \hat{A} \), and the corresponding Hilbert space is denoted by \((\mathcal{H}^A)^\otimes n \) or \( \mathcal{H}^\hat{A} \). The Shannon entropy of a probability distribution is denoted as \( H(\{p_j\}_{j}) \), and von Neumann entropy of a state \( \rho^A \) is interchangeably denoted by \( S(\rho^A) \) and \( S(A)_{\rho^A} \).\( \log x \) represents the base 2 logarithm of \( x \).

**II. Preliminaries**

In this section, we present a decomposition of a Hilbert space called the Koashi-Imoto (KI) decomposition, which is introduced in [21] and is extensively used in the following part of this paper. We then summarize a result in [1], which states that the structure of Markov states is characterized by the KI decomposition.

**A. Koashi-Imoto Decomposition**

For any set of states on a quantum system, operations on that system are classified into two categories: one that do not change any state in the set, and the other that changes at least one state in the set. It is proved in [21] that there exists an effectively unique way of decomposing a Hilbert space into a direct-sum form, in such a way that all quantum operations that do not change a given set of states have a simple form with respect to the decomposition. We call this decomposition of the Hilbert space as the Koashi-Imoto decomposition, or the KI decomposition for short. As we verify in Remark in this section, the KI decomposition is equivalently represented in the form of a tensor product of three Hilbert spaces. Theorem 3 in [21], which proves the existence of the KI decomposition, is described in this tensor-product form as follows.

**Theorem 1** (see Theorem 9 in [1]) Consider a quantum system \( A \) described by a finite dimensional Hilbert space \( \mathcal{H}^A \). Associated to any set of states \( \mathcal{S} := \{ \rho_{j,k} \}_{j,k} \) on \( A \), there exist Hilbert spaces \( \mathcal{H}^{a_j}, \mathcal{H}^{a_k}, \mathcal{H}^{a_j a_k}, \) an orthonormal basis \( \{ |j\rangle \}_{j \in J} \) of \( \mathcal{H}^{a_j} \) and a linear isometry \( \Gamma \) from \( \mathcal{H}^A := \text{supp}(\sum_k \rho_{k}) \subseteq \mathcal{H}^A \) to \( \mathcal{H}^{a_j} \otimes \mathcal{H}^{a_k} \otimes \mathcal{H}^{a_j a_k} \), such that the following three properties hold. (For later convenience, we exchange labels \( L \) and \( R \) in the original formulation.)

1) The states in \( \mathcal{S} \) are decomposed by \( \Gamma \) as

\[
\Gamma \rho_{j,k} \Gamma^\dagger = \sum_{j \in J} p_{j|k} |j\rangle \langle j^a_k| \otimes \omega^a_j \otimes \rho^a_{j,k} \tag{1}
\]

with some probability distribution \( \{ p_{j|k} \}_{j \in J} \) on \( J := \{ 1, \ldots, \text{dim} \mathcal{H}^{a_j} \} \), states \( \omega_j \in S(\mathcal{H}^{a_j}) \) and \( \rho_{j,k} \in S(\mathcal{H}^{a_j a_k}) \).

2) A quantum operation \( E \) on \( S(\mathcal{H}^{a_j a_k}) \) leaves all \( \rho_{j,k} \) invariant if and only if there exists an isometry \( U : \mathcal{H}^A \to \mathcal{H}^A \otimes \mathcal{H}^B \) such that a Stinespring dilation of \( E \) is given by \( E(\tau) = \text{Tr}_E[U \tau U^\dagger] \), and that \( U \) is decomposed by \( \Gamma \) as

\[
(\Gamma \otimes I^E) U \Gamma^\dagger = \sum_{j \in J} |j\rangle \langle j^a_k| \otimes U^a_j \otimes I^a_j. \tag{2}
\]

Here, \( I_j \) are the identity operators on \( \mathcal{H}_j^a := \text{supp}(\sum_k \rho_{j,k}) \) and \( U_j : \mathcal{H}_j^a \to \mathcal{H}_j^a \otimes \mathcal{H}^E \) are isometries that satisfy \( \text{Tr}_E[U_j \omega_j U_j^\dagger] = \omega_j \) for all \( j \), where \( \mathcal{H}_j^a := \text{supp} \omega_j \).

3) \( \Gamma \) satisfies

\[
\text{img} \Gamma = \bigoplus_{j \in J} \mathcal{H}_j^a \otimes \mathcal{H}_j^{a_k} \otimes \mathcal{H}_j^{a_j a_k}, \tag{3}
\]

where \( \mathcal{H}_j^a \) are one-dimensional subspaces of \( \mathcal{H}_j^a \) spanned by \( |j\rangle \).

4) \( \mathcal{H}_j^a, \mathcal{H}_j^{a_k} \) and \( \mathcal{H}_j^{a_j a_k} \) are minimal in the sense that

\[
\dim \mathcal{H}_j^a = \max _{j \in J} \dim \mathcal{H}_j^a, \quad \dim \mathcal{H}_j^{a_k} = \max _{j \in J} \dim \mathcal{H}_j^{a_k}
\]

and

\[
\forall j \in J, \exists k \text{ s.t. } p_{j|k} > 0.
\]

We call \( \Gamma \) as the KI isometry on system \( A \) with respect to \( \mathcal{S} \). The KI decomposition and the corresponding KI isometry are uniquely determined from \( \mathcal{S} \), up to trivial changes of the basis (Lemma 7 in [21]). The dimensions of \( \mathcal{H}_j^a, \mathcal{H}_j^{a_k} \) and \( \mathcal{H}_j^{a_j a_k} \) are at most \( d_A \). An algorithm for obtaining the KI decomposition is proposed in [21].

It is also proved in [21] that the sets of states \( \{ \rho_{j|k} \}_{k} \) in \( \mathcal{S} \) are irreducible in the following sense.

**Lemma 2** (Corollary of Lemma 6 in [21]) The set of states \( \{ \rho_{j|k} \}_{k} \) in \( \mathcal{S} \) satisfies the following properties.

1) If a linear operator \( N \) on \( \mathcal{H}_j^a := \text{supp}(\sum_k \rho_{j,k}) \) satisfies \( p_{j|k} N \rho_{j|k} = p_{j|k} \rho_{j|k} N \) for all \( k \), then \( N = c \Gamma_j^a \) for a complex number \( c \), where \( \Gamma_j^a \) is the identity operator on \( \mathcal{H}_j^a \).

2) If a linear operator \( N : \mathcal{H}_j^a \to \mathcal{H}_j^a (j \neq j') \) satisfies \( p_{j|k} N \rho_{j|k} = p_{j'|k} \rho_{j'|k} N \) for all \( k \), then \( N = 0 \).

Let us now describe an extension of the KI decomposition to a bipartite quantum states, which is introduced in [1]. Associated to any bipartite state \( \Psi^{AA'}(\in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^{A'}) \), there exists a set of states on \( A \) to which system \( A \) can be steered through \( \Psi^{AA'} \), i.e., the set of states that can be prepared by performing a measurement on \( A' \) on the state \( \Psi^{AA'} \) and post-selecting one outcome. The KI decomposition of \( A \) with respect to the set is then associated to \( \Psi^{AA'} \). It happens that any quantum operation on \( A \) which leaves all states in the set invariant also leaves \( \Psi^{AA'} \) invariant, and vice versa. Hence the set of operations preserving \( \Psi^{AA'} \) is completely characterized by the corresponding KI decomposition. More precisely, we have the following statements.

**Definition 3** Consider quantum systems \( A \) and \( A' \) described by finite dimensional Hilbert spaces \( \mathcal{H}^A \) and \( \mathcal{H}^{A'} \), respectively.
The KI decomposition of system $A$ with respect to a bipartite state $\Psi^{AA'} \in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^{A'})$ is defined as the KI decomposition of $A$ with respect to the following set $\mathcal{E}^{AA'}_{\Psi}$ of states, i.e., using $\mathcal{E}(\mathcal{H}^A)$ to denote the set of linear operators on $\mathcal{H}^A$, 

$$\mathcal{E}^{AA'}_{\Psi} := \{ \rho \in \mathcal{S}(\mathcal{H}^A) | \exists M \in \mathcal{E}(\mathcal{H}^A) \text{ s.t. } \rho = \text{Tr}_{A'}[M^{AA'}\Psi^{AA'}M^{AA'}] \}.$$  

(4)

The KI isometry on system $A$ with respect to $\Psi^{AA'}$ is defined as that with respect to $\mathcal{E}^{AA'}_{\Psi}$.

**Lemma 4** (See the proof of Theorem 6 in [1] and Equality (14) therein.) Let $\Gamma$ be the KI isometry on $A$ with respect to $\Psi^{AA'}$, and define $\mathcal{H}^{AA'}_{\Psi} := \text{supp}[\Psi^{AA'}] \subseteq \mathcal{H}^A$. $\Gamma$ satisfies the following properties.

1) $\Gamma$ gives

$$\Psi^{AA'}_{\Gamma_1} := \Gamma^A \Psi^{AA'} \Gamma_1^A = \sum_{j \in J} p_j |j\rangle \langle j|^{aa'} \otimes \omega_j \otimes \phi_j^{AA'}$$

(5)

with some probability distribution $\{p_j\}_{j \in J}$, orthonormal basis $\{|j\rangle\}_{j \in J}$ of $\mathcal{H}^{aa'}$, states $\omega_j \in \mathcal{S}(\mathcal{H}^{bb})$ and $\phi_j \in \mathcal{S}(\mathcal{H}^{cc} \otimes \mathcal{H}^{bb})$.

2) A quantum operation $\mathcal{E}$ on $\mathcal{S}(\mathcal{H}_{\Psi})$ leaves $\Psi^{AA'}$ invariant only if there exists an isometry $U : \mathcal{H}_{\Psi}^A \rightarrow \mathcal{H}_{\Psi}^A \otimes \mathcal{H}^E$ such that a Stinespring dilation of $\mathcal{E}$ is given by $\mathcal{E}(\tau) = \text{Tr}_{E}[U \tau U^\dagger]$, and that $U$ is decomposed by $\Gamma$ as (2), in which case we define $\mathcal{H}^{aa}_{\Psi} := \text{supp}[\phi_j]_{j \in J}$.

We call (5) as the KI decomposition of $\Psi^{AA'}$ on $A$. The following lemma, regarding the equivalence between the KI isometries of two bipartite states, immediately follows.

**Lemma 5** The following conditions are equivalent when $\mathcal{H}^A_{\Psi} = \mathcal{H}_{AA'}^A = \mathcal{H}^A$:

1) A quantum operation on $A$ leaves a state $\Psi_{1}^{AA'}$ invariant if and only if it leaves a state $\Psi_{2}^{AA'}$ invariant.

2) The KI isometries on $A$ with respect to $\Psi_{1}^{AA'}$ and $\Psi_{2}^{AA'}$ are the same.

We define the sub-KI isometries as follows.

**Definition 6** Consider a bipartite state $\Psi^{AA'}$, three Hilbert spaces $\mathcal{H}^{aa}$, $\mathcal{H}^{cb}$, $\mathcal{H}^{bc}$ and let $\Gamma$ be a linear isometry from $\mathcal{H}^A_{\Psi}$ to $\mathcal{H}^{aa} \otimes \mathcal{H}^{cb} \otimes \mathcal{H}^{bc}$. We call $\Gamma$ as a sub-KI isometry on system $A$ with respect to $\Psi^{AA'}$ if it satisfies Condition 1) in Lemma 4.

**B. Markov States**

A tripartite quantum state $\Upsilon^{ABC}$ is called a Markov state conditioned by $B$ if it satisfies $I(A : C | B)_{\Upsilon} = 0$. It is proved in [1] that the structure of Markov states is characterized by the KI decomposition as follows.

**Theorem 7** (See Theorem 6 in [1] and the proof thereof.) The following three conditions are equivalent:

1) $\Upsilon^{ABC}$ is a Markov state conditioned by $B$.

2) There exist three Hilbert spaces $\mathcal{H}^{bb}$, $\mathcal{H}^{bc}$, $\mathcal{H}^{bc}$ and a linear isometry $\Gamma$ from $\mathcal{H}^B_{\Psi} := \text{supp}[\Upsilon] B$ to $\mathcal{H}^{bb} \otimes \mathcal{H}^{bc} \otimes \mathcal{H}^{bc}$ such that $\Upsilon^{ABC}$ is decomposed by $\Gamma$ as

$$\Upsilon^{ABC} = \Gamma(\Upsilon^{AB}) = \Gamma' (\Upsilon^{BC}).$$

(6)

3) $\Upsilon^{ABC}$ is decomposed in the form of (6) with $\Gamma$ being the KI isometry on $B$ with respect to $\Upsilon^{BC}$.

4) There exist quantum operations $\mathcal{R}$ from $B$ to $BC$ and $\mathcal{R}'$ from $B$ to $AB$ such that

$$\Upsilon^{ABC} = \mathcal{R}(\Upsilon^{AB}) = \mathcal{R}' (\Upsilon^{BC}).$$

(7)

We call (6) as a Markov decomposition of a Markov state $\Upsilon^{ABC}$ (Figure 1).

**Remark:** The KI decomposition is first proved in [21] by a graphic approach, and by an algebraic proof in [1] afterward. A similar decomposition is derived in [22] and [23] in the context of “information preserving structure”. In these literatures, the decomposition is given in the form of the direct sum of Hilbert spaces $\bigoplus_j \mathcal{H}^j_A \otimes \mathcal{H}^j_B$. This is equivalent to the decomposition in the form of a tensor product of three Hilbert spaces described in this section, as verified by choosing $\mathcal{H}^{aa}$, $\mathcal{H}^{bc}$ and $\mathcal{H}^{bc}$ such that $\dim \mathcal{H}^{aa} = |J|$, $\dim \mathcal{H}^{bc} = \max_{j \in J} \mathcal{H}^j_{AB}$ and $\dim \mathcal{H}^{bc} = \max_{j \in J} \mathcal{H}^j_{AB}$. The corresponding KI isometry is defined as

$$\Gamma := \sum_{j \in J} |j\rangle^{aa} \otimes (\Gamma_{L,j} \otimes \Gamma_{R,j}) P_j$$

(8)

where $\Gamma_{L,j} : \mathcal{H}^j_A \rightarrow \mathcal{H}^{aa}$ and $\Gamma_{R,j} : \mathcal{H}^j_B \rightarrow \mathcal{H}^{bc}$ are linear isometries and $P_j$ is the projection onto $\mathcal{H}^j_A \otimes \mathcal{H}^j_B \in \mathcal{H}$. As stressed in [21], $\mathcal{H}^{aa}$ in [1] holds the “classical” part of information possessed by $\rho_k$, namely, $\sum_k q_k |k\rangle^{aa} \otimes \phi_k^{ab} \otimes \phi_k^{ac}$.
III. DEFINITIONS AND MAIN RESULTS

In this section, we introduce the formal definition of Markovianization, and state the main results on the Markovianizing cost of tripartite pure states. The outlines of proofs are given in Section IV. Rigorous proofs will be given in Appendix B and C.

Definition 8 A tripartite state $\rho_{ABC}$ is Markovianized with the randomness cost $R$ on $A$, conditioned by $B$, if the following statement holds. That is, for any $\epsilon > 0$, there exists $n_\epsilon$ such that for any $n \geq n_\epsilon$, we find a random unitary operation $V_n : \tau \mapsto 2^{-nR} \sum_{k=1}^{2^{nR}} V_k \tau V_k^\dagger$ on $n$ and a Markov state $\Upsilon_n^{A^n B^n C^n}$ conditioned by $B^n$ that satisfy

$$\|V_n^n(\rho^\otimes n) - \Upsilon_n^{A^n B^n C^n}\|_1 \leq \epsilon.$$  \hspace{1cm} (9)

The Markovianizing cost of $\rho_{ABC}$ is defined as $M_{A|B}(\rho_{ABC}) := \inf \{ R \mid \rho_{ABC} \text{ is Markovianized with the randomness cost } R \text{ on } A, \text{conditioned by } B \}$. The following theorem is the main contribution of this work. The outline of the proof is given in the next section.

Theorem 9 Let $|\Psi\rangle_{ABC}$ be a pure state, and let

$$\Psi_{KI}^{AC} = \sum_{j \in J} p_j |j\rangle \otimes \omega_{ji}^{AC}$$  \hspace{1cm} (10)

be the KI decomposition of $\Psi^{AC}$ on $A$. Then we have $M_{A|B}(\Psi_{ABC}) = H(\{p_j \}_{j \in J}) + 2 \sum_{j \in J} p_j S(\omega_{ji}^{AC})$.

Based on this theorem, it is possible to compute the Markovianizing cost of pure states once we obtain the KI decomposition of its bipartite reduced density matrix. However, the algorithm for obtaining the KI decomposition, which is proposed in [21], involves repeated application of decompositions of the Hilbert space into subspaces, and is difficult to execute in general.

Below we propose an algorithm by which we can compute the Markovianizing cost of pure states for a particular class of pure states, without obtaining an explicit form of the KI decomposition. The algorithm is based on the following theorem, which connects the Markovianizing cost of a pure state and the Petz recovery map corresponding to the state. Here, the Petz recovery map of a tripartite state $\Psi_{ABC}$ from $A \to AC$, an idea first introduced in [1], is defined by

$$R_{\Psi}^{A \to AC}(\tau) = (\Psi^{AC})^{1/2} \tau (\Psi^{AC})^{-1/2}$$

$$(\forall \tau \in S(\mathcal{H}^A)).$$

A proof of the theorem will be given in Appendix C.

Theorem 10 Let $|\Psi\rangle_{ABC}$ be a pure state, such that a CPTP map $E$ on $S(\mathcal{H}^A)$ defined by

$$E := \text{Tr}_C \circ R_{\Psi}^{A \to AC}$$  \hspace{1cm} (11)

is self-adjoint. Define another CPTP map $E_\infty$ by

$$E_\infty := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^N E^n,$$  \hspace{1cm} (12)

and consider the state

$$\Psi_{ABC}^\infty := E_\infty (|\Psi\rangle \langle |\Psi|_{ABC}).$$  \hspace{1cm} (13)

Then we have

$$M_{A|B}(\Psi_{ABC}) = S(\Psi_{ABC}^\infty).$$  \hspace{1cm} (14)

Due to this theorem, the Markovianizing cost of pure states can be computed by the following algorithm, based on a matrix representation of CPTP maps.

1) Compute $d_A^2$-dimensional square matrices $\Lambda_1, \Lambda_2$ and $\Lambda$ given by

$$[\Lambda_1]_{kl, mn} = \langle k| (\Psi^A)^{1/2} |m\rangle \langle n| (\Psi^A)^{1/2} |l\rangle,$$

$$[\Lambda_2]_{kl, mn} = \text{Tr} \left[ (\Psi^{AC})^{1/2} |m\rangle \langle n| (\Psi^{AC})^{1/2} |l\rangle \right],$$

and $\Lambda = \Lambda_2 \Lambda_1^{-1}$, where the superscript $-1$ denotes the generalized inverse.

2) Check the hermiticity of $\Lambda$, which is equivalent to the self-adjointness of $E$. If it is Hermitian, continue to Step 3. If not, this algorithm is not applicable.

3) Compute a matrix $\tilde{\Lambda}_\infty \propto \Lambda_\infty$, which is given by the projection onto the eigensubspace of $\Lambda$ corresponding to the eigenvalue 1. Then compute $\Lambda_\infty = \Lambda_\infty \Lambda_1$.

4) Compute $\Omega_{AA'} \propto \Lambda_\infty$ given by

$$\Omega_{AA'} = \sum_{kl, mn} [\Lambda_\infty]_{kl, mn} [k]_{A} |m\rangle \langle n| A'.$$

5) Compute the Shannon entropy of the eigenvalues of $\Omega_{AA'}$, which is equal to $M_{A|B}(\Psi_{ABC})$.

IV. OUTLINE OF PROOFS OF THE MAIN THEOREMS

In this section, we describe the outline, main concepts and technical ingredients for the proofs of Theorem 9 and 10. Detailed proofs are given in Appendix B and C.

We first introduce an adaptation of the KI decomposition to tripartite pure states as follows.

Fig. 2. A graphical representation of the KI decomposition of tripartite pure states. Each vertex corresponds to a quantum system. The solid lines express pure states. The whole state is the superposition of the above states with the probability amplitude $\sqrt{p_j}$, namely, $\sum_j \sqrt{p_j} |j_{kn}^a| \omega_j^{|a_L b_k |(\varphi_j)^{aRB_nC}}.$
Lemma 11 Let $|\Psi\rangle_{ABC}$ be a tripartite pure state and suppose that the KI decomposition of $\Psi^A_{BC}$ on $A$ is given by

$$Gamma^A \Psi^A_{BC} Gamma^A = \sum_{j \in J} p_j |j\rangle^{\bar{b}_0} \otimes \omega_j^{a_L} \otimes \phi_j^{a_R}.$$  

There exists a linear isometry $Gamma' : H_B^{|\Psi\rangle} \rightarrow H_B^{a_L} \otimes H_B^{a_R} \otimes H_B^{a_C}$ that decomposes $|\Psi\rangle_{ABC}$ together with $Gamma$ as

$$(Gamma^A \otimes Gamma^B) |\Psi\rangle_{ABC} = \sum_{j \in J} \sqrt{p_j} |j\rangle^{a_L} |j\rangle^{a_R} |j\rangle^{a_C} |\phi_j\rangle^{a_R'C}.$$  

where $|\omega_j\rangle^{a_L}$ and $|\phi_j\rangle^{a_R'C}$ are purifications of $\omega^{a_L}$ and $\phi^{a_R'C}$, respectively, and $\langle j|j\rangle^{a_L} = \delta_{jj'}$. Moreover, $Gamma'$ is the sub-KI isometry on $B$ with respect to $\bar{\Psi}^{BC}$.

Proof: The existence of $Gamma'$ follows from Uhlmann’s theorem (24], see Appendix A-A). It is straightforward to verify that $Gamma'$ is the sub-KI isometry, since we have

$$Gamma^B \bar{\Psi}^{BC} Gamma^B = \sum_{j \in J} p_j |j\rangle^{b_0} \otimes \omega_j^{b_1} \otimes \phi_j^{b_C}$$

from (15).

We call (15) as the KI decomposition of $|\Psi\rangle_{ABC}$ on $A$ and $B$ (Figure 2), and denote it by $|\Psi_{K1}\rangle$.

A. For Theorem 9 Achievability

The direct part of Theorem 9 is formulated by the following inequality:

$$M_{A|B}(\Psi^{ABC}) \leq H(\{p_j\}_{j \in J}) + 2 \sum_{j \in J} p_j S(\phi_j^{a_R}).$$  

The outline of the proof is as follows. The state $|\Psi^{ABC}\rangle^{\otimes n}$ is local unitarily equivalent to $|\Psi_{K1}\rangle^{\otimes n}$, which is almost equal to the state defined by

$$|\Psi_{n,\delta}\rangle := \sum_{j \in J_{n,\delta}} \sqrt{p_j} |j\rangle^{a_L} |j\rangle^{a_R} |\omega_j\rangle^{a_L} |\phi_j\rangle^{a_R'C}$$

for sufficiently large $n$. Here, we have introduced notations $j = j_1 \cdots j_n$, $\varphi_j = \varphi_{j_1} \otimes \cdots \otimes \varphi_{j_n}$ and $\omega_j = \omega_{j_1} \otimes \cdots \otimes \omega_{j_n}$. $J_{n,\delta}$ is the $\delta$-strongly typical set with respect to the probability distribution $\{p_j\}_j$, and $\Pi_j^{\delta}$ is the projection onto the conditionally typical subspace of $\varphi_j^{a_R} \otimes \omega_j^{a_L}$ conditioned by $j$. Consider a unitary operation on $supp\{\Psi_{n,\delta}\}_{n,\delta}$ of the form

$$V := \sum_{j \in J_{n,\delta}} |j\rangle^{a_L} \otimes I^A_{j} \otimes v^B_{j}.$$  

where $I^A_{j}$ is the identity operator on $supp\omega^{a_L}$ and $v^B_{j}$ is a unitary on the support of $\Pi_j^{\delta}$. We apply $V^{n,\delta}_{n,\delta}$ on $\Psi_{n,\delta}$ by independently choosing $v^B_{j}$ from the Haar distributed random unitary ensemble for each $j$. By this random unitary operation, the state (17) is transformed to the following state

$$\bar{\Psi}_{n,\delta} := \sum_{j \in J_{n,\delta}} p_j |j\rangle^{a_L} \otimes |\omega_j\rangle^{a_L} \otimes |\phi_j^{a_R'}\rangle,$$

where $\phi_j^{a_R'} := Tr_B[p_{A|B}(j)|\varphi_j^{a_R}'\rangle]$ and $\bar{\Psi}_{n,\delta}$ is a Markov state conditioned by $B$ (Figure 3).

To Markovianize $|\Psi_{ABC}^{\otimes n}\rangle$, it is sufficient that we approximate the transformation from (17) to (19) by $\tilde{\Psi}$ with a vanishingly small error, where $\Psi_k$ in $\Psi_k^{\otimes n}$ are unitaries which are decomposed by $Gamma \otimes Gamma' \otimes Gamma'' \otimes \cdots \otimes Gamma'' \otimes \cdots$ as (18). By a random coding method and the operator Chernoff bound (19), it is shown that a sufficient number of unitaries in $\Psi_k$ for this approximation is almost equal to the inverse of the minimum nonzero eigenvalue of (19), and is given as $H(\{p_j\}_{j \in J}) + 2 \sum_{j \in J} p_j S(\phi_j^{a_R})$ per copy. We note that the error $\epsilon$ converges exponentially with $n$ to zero.

B. For Theorem 9 Optimality

The converse part of Theorem 9 is formulated by the following inequality:

$$M_{A|B}(\Psi^{ABC}) \geq H(\{p_j\}_{j \in J}) + 2 \sum_{j \in J} p_j S(\phi_j^{a_R}).$$  

Let us first assume tentatively that a Markov decomposition of $\bar{\Psi}_{ABC}$ in (9) is given by

$$(Gamma^B)^{\otimes n} \bar{\Psi}_{ABC} (Gamma^B)^{\otimes n} = \sum_{j \in J_n} p'_j |j\rangle^{a_L} \otimes \sigma_j^{A_L} \otimes \phi_j^{b_C},$$

with $Gamma'$ being the KI isometry on $B$ with respect to $\bar{\Psi}^{BC}$. In this case, it is not difficult to show that the amount of randomness per copy required for transforming $|\Psi_{ABC}^{\otimes n}\rangle$ to $\bar{\Psi}_{ABC}$ is bounded below by the R.H.S. of (20). Indeed, in order to transform $|\Psi_{K1}\rangle^{\otimes n}$ to a Markov state in the form of (20), it is necessary that (i) the off-diagonal terms with respect to $|j\rangle$ vanish, and (ii) the correlation between $\bar{a}_B$ and $b_C$ in the state $|\phi_j^{a_R'C}\rangle$ is destroyed for each $j$. An optimal way for satisfying these two conditions is transforming the state (17) close to a state of the form (19). Since the entropy of the state (19) is approximately equal to $n(H(\{p_j\}_{j \in J}) + 2 \sum_{j \in J} p_j S(\phi_j^{a_R}))$, the cost of randomness required for this transformation is at least about

$$H(\{p_j\}_{j \in J}) + 2 \sum_{j \in J} p_j S(\phi_j^{a_R}) \text{ bits per copy.}$$

However, it might be possible in general that the amount of randomness can be further reduced by appropriately choosing...
\[ \mathcal{E}_{\chi}^* = \left( \sum_i |i\rangle \langle i|^* \right) \Gamma_{\chi^*} \]

This is decomposed as \( \mathcal{E}_{\chi}^* = \mathcal{E}_1 \circ \mathcal{E}_2 \circ \mathcal{E}_3 \circ \mathcal{E}_4 \circ \mathcal{E}_5 \). The maps \( \mathcal{E}_1 \) and \( \mathcal{E}_2 \) are isometry channels corresponding to \( \Gamma_1 \) and \( \Gamma_2^* \), respectively; \( \mathcal{E}_3 \) is discarding of system \( \tilde{a}_0 \); \( \mathcal{E}_2 \) is the completely dephasing channel on \( \tilde{a}_0 \) with respect to the basis \( |i\rangle \); \( \mathcal{E}_3 \) is appending of the state \( \xi_{iL} \), conditioned by \( \tilde{a}_0 \) (Figure 5).
Let $\Gamma$ be the KI isometry on $A$ with respect to $\Psi^{ABC}$, and suppose the KI decomposition of $\Psi^{ABC}$ is given by (10). For each $E_{kl}$, we have
\[ \hat{E}_{kl} := \Gamma E_{kl} \Gamma^\dagger = \sum_{j \in J} |j\rangle\langle j| \otimes I_j^{\text{as}} \otimes e_{j,kl}^{an}, \tag{29} \]
where
\[ e_{j,kl} := \langle k|^{C} (\varphi^{a_{kl}})_{C}^{\dagger} |l\rangle^{C} (\varphi^{a_{kl}})^{-\frac{1}{2}}. \tag{30} \]
By an extension of Lemma 2, it follows that $\{e_{j,kl}\}_{kl}$ is irreducible in the sense that it satisfies Property 1 and 2. It is straightforward to verify from (30) that maps $E_{\{\}}$ on $S(\mathcal{H}_j^{as})$, defined by $E_{\{}}^{a_{kl}}(\cdot) := \sum_{k,l} e_{j,kl}(\cdot) e_{j,kl}^{\dagger} (j = 1, \ldots, |J|)$, are trace-preserving. Representations of $E$ and $E_\infty$ in the decomposed Hilbert space are given by
\[ \hat{E}(\cdot) = \sum_{kl} \hat{E}_{kl}(\cdot) \hat{E}^{\dagger}_{kl}, \tag{31} \]
and $\hat{E}_\infty := \lim_{n \to \infty} (1/N) \sum_{n=1}^{N} \hat{E}^{n}$, respectively, and that of $\Psi^{ABC}$ is given by $\hat{E}_\infty(\Psi^{ABC}(\Psi^{ABC})^\dagger |_{J})$.

Due to $E \circ E_\infty = E_\infty$ and the irreducibility of $\{e_{j,kl}\}_{kl}$, we have
\[ E_\infty(\Psi^{ABC}(\Psi^{ABC})^\dagger |_{J}) = \sum_{j \in J} p_j^{\dagger} |j\rangle_{j}^{a_{kl}} \otimes \varpi_j^{a_{kl}} \otimes \pi_j^{a_{kl}}, \tag{32} \]
where $\{p_j\}_{j \in J}$ is a probability distribution and $\varpi_j (j \in J)$ are states on $a_{l_j}b_{l_j}b_{l_j}C$. Explicit forms of $\varpi_j$ and $p_j$ are obtained as follows. First, from (29), (31) and the trace-preserving property of $\hat{E}_\infty$, we have
\[ \text{Tr}[|j\rangle\langle j|^{a_{kl}} \hat{E}(\hat{\tau}) |j\rangle^{a_{kl}}] = \text{Tr}[|\hat{\tau}\rangle\langle \hat{\tau}|^{a_{kl}} \otimes \hat{E}^{a}\otimes \hat{E}^{a}_{j} |j\rangle^{a_{kl}}] = \text{Tr}[|j\rangle\langle \hat{\tau}|^{a_{kl}} |j\rangle^{a_{kl}}], \]
for any $j$ and $\hat{\tau} = \Gamma \tau \Gamma^\dagger$ ($\tau \in S(\mathcal{H}_j^{as})$). This implies that the probability amplitude with respect to the basis $\{\{j\}\}_{j}$ is conserved by $\hat{E}_\infty$ as well as by $\hat{E}_\infty$. Thus we have $p_j^{\dagger} = p_j$. Observe from (29) that $\hat{E}$ and $\hat{E}_\infty$ do not affect the system $a_l$, which implies
\[ \varpi_j = |\omega_j\rangle |\omega_j\rangle^{a_{l_j}b_{l_j}b_{l_j}} \otimes |\phi_j\rangle^{a_{l_j}b_{l_j}}. \]
Hence the von Neumann entropy of $\Psi^{ABC}_\infty$, which is equal to that of (32), is given by
\[ S(\Psi^{ABC}_\infty) = H(\{p_j\}_{j}) + \sum_j p_j \left( S(\varpi^{a_{kl}}) + S(\pi^{a_{kl}}) \right). \]
In addition, that the self-adjointness of $E$ implies $\varphi^{a_{kl}} = \varphi^{a_{kl}}$. Since $|\varphi_j\rangle |\varphi_j\rangle^{a_{l_j}b_{l_j}b_{l_j}}$ is a purification of $\varphi^{a_{kl}}$, we finally obtain that
\[ S(\Psi^{ABC}_\infty) = H(\{p_j\}_{j}) + 2 \sum_j p_j S(\varphi^{a_{kl}}) = M_{A|B}(\Psi^{ABC}). \]

\section{Properties}

In this section, we describe properties of the Markovianizing cost of tripartite quantum states. We first consider arbitrary (possibly mixed) states, and then focus on the case of pure states.

\subsection{A. General Properties}

Let $\rho^{ABC}$ be an arbitrary tripartite state on finite dimensional quantum systems $A$, $B$, and $C$. The Markovianizing cost of $\rho^{ABC}$ satisfies
\[ I(\rho^{ABC}) \leq M_{A|B}(\rho^{ABC}) \leq I(\rho^{ABC}) \tag{33} \]
The second inequality directly follows from the fact that decoupling $A$ from $BC$ is sufficient for converting the state to a Markov state and that the cost of randomness for decoupling bipartite states is asymptotically given by QMI (11). The first inequality is proved in Appendix B. Consequently, the Markovianizing cost is equal to zero only for Markov states.

The Markovianizing cost satisfies a kind of the data processing inequality, namely, that
\[ M_{A|B}(\rho^{ABC}) \leq M_{A|B}(\rho^{ABC}) \]
under any quantum operation $E$ on $C$. This is because any random unitary operation $\rho_n$ on $A^n$ satisfying (3) also satisfies
\[ \left\| \left| \rho^{ABC} \otimes^n (E^{\otimes n})(\Psi^{ABC} \otimes^n (Y^{A^n B^n C^n}) \right) \right\| \leq \epsilon, \]
and $(E^{\otimes n})(\Psi^{ABC} \otimes^n (Y^{A^n B^n C^n})$ is a Markov state conditioned by $B^n$. As a consequence, an upper bound on the Markovianizing cost of a mixed state is obtained as
\[ M_{A|B}(\rho^{ABC}) \leq M_{A|B}(\rho^{ABC}), \]
where $|\psi^{ABC}\rangle$ is a purification of $\rho^{ABC}$. This is because there always exists a quantum operation $F_{\rho} : C' \to C$ such that $F_{\rho}(\psi^{ABC}) = \rho^{ABC}$.

\subsection{B. Pure States}

Let us now consider pure states, based on the result presented in Section III. First, we see that the Markovianizing cost $M_{A|B}$ of two pure states $\Psi_1$ and $\Psi_2$ are equal if there exist $\lambda (0 < \lambda \leq 1)$ and $\sigma^{C} \in S(\mathcal{H}^{C})$ such that
\[ \Psi^{ABC} = \lambda \Psi^{ABC} + (1 - \lambda) \Psi^{ABC} \tag{34} \]
This is because $\Psi^{ABC} = \Psi^{ABC}$ and the KI decompositions of $A$ with respect to $\Psi^{ABC}$ and $\Psi^{ABC}$ are equal, the latter of which follows from Lemma 3. Indeed, any quantum operation on $A$
which keeps \( \Psi_1^{AC} \) invariant also keeps \( \Psi_2^{AC} \) invariant and vice versa, as can be seen by observing from (34) that we have

\[
\Psi_1^{AC} = \frac{1}{\lambda} \Psi_2^{AC} - \frac{1 - \lambda}{\lambda} \Psi_2^A \otimes \sigma^C.
\]

Second, we obtain a necessary and sufficient condition for the Markovianizing cost of pure states to be equal to QCMI as follows (Figure 6).

**Theorem 13** Let \( |\Psi\rangle^{ABC} \) be a pure state, and let

\[
|\Psi_{KI}\rangle^{ABC} = \sum_{j \in \mathcal{J}} \sqrt{P_j} |j\rangle^{a} |j\rangle^{b} |\omega_j\rangle^{a_1 b_c} |\varphi_j\rangle^{a_2 b_c} |c\rangle,
\]

be its KI decomposition on \( A \) and \( B \) (see Lemma 11). Then we have \( M_{A|B}(\Psi^{ABC}) = I(A : C|B)_{\Psi} \) if and only if there exists an isometry \( \Gamma'' : \mathcal{H}_C \to \mathcal{H}_C \otimes \mathcal{H}_C \otimes \mathcal{H}_C \) such that \( \varphi_j \) is decomposed as

\[
\Gamma''\varphi_j^{a_2 b_c} = |j\rangle^{a} |\varphi_{j,1}\rangle^{a_1 c} |\varphi_{j,2}\rangle^{b_c c}.
\]

where \( \langle j|j'\rangle = \delta_{j,j'} \).

**Proof:** We have

\[
I(A : C|B)_{\Psi} = I(A : C)_{\Psi} = I(a_0 a_1 a_R : C)_{\Psi_{KI}}
\]

\[
= I(a_0 : C)_{\Psi_{KI}} + I(a_1 a_R : C|a_0)_{\Psi_{KI}}
\]

\[
= I(a_0 : C)_{\Psi_{KI}} + \sum_j p_j I(a_R : C)_{\varphi_j},
\]

as well as

\[
M_{A|B}(\Psi^{ABC}) = H(a_0) + \sum_j p_j S(a_R|\varphi_j).
\]

Since \( \Psi_{KI}^{C} \) is a classical-quantum state, we have \( I(a_0 : C)_{\Psi_{KI}} \leq H(a_0) \) with equality if and only if \( \{\text{supp}(\varphi_j^C)\}_j \) is mutually orthogonal. We also have \( 2S(a_R|\varphi_j) \geq I(a_R : C|\varphi_j) \), which is saturated if and only if \( 2S(a_R|\varphi_j) - I(a_R : b_R|\varphi_j) = I(a_R : b_R|\varphi_j) = 0 \) (see Inequality (35) in Appendix A-B). Hence we have \( M_{A|B}(\Psi^{ABC}) = I(A : C|B)_{\Psi} \) if and only if

\[
\Psi_{KI}^{AB} = \sum_{j \in \mathcal{J}} p_j |j\rangle\langle j|^{a b} \otimes |\omega_j\rangle^{a_1 b_c} \otimes \varphi_j^{a_2 b_c} \otimes \varphi_j^b,
\]

which concludes the proof due to Uhlmann’s theorem (24), see Appendix (A-A).

An example of states that satisfy the above conditions is given in Section VI-C.

### VI. Examples

In this section, we consider examples of pure states to illustrate discontinuity and asymmetry of the Markovianizing cost. We also give an example of states for which the Markovianizing cost is equal to QCMI.

**A. Discontinuity**

We consider tripartite pure states that are expressed as

\[
|\Psi_{\lambda}\rangle = \sqrt{\frac{d^2 \lambda - 1}{d^2 - 1}} |000\rangle^B |\Phi_d\rangle^{AC} + \sqrt{\frac{1 - \lambda}{d^2 - 1}} \sum_{k,l=1}^{d} |k\rangle^{B} |k\rangle^{A} |l\rangle^{C},
\]

where \( d = \dim \mathcal{H}_A = \dim \mathcal{H}_C, \dim \mathcal{H}_B = d^2 + 1, 1/d^2 \leq \lambda \leq 1, \) and \( \Phi_d \) is a maximally entangled state of Schmidt rank \( d \), defined as

\[
|\Phi_d\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |k\rangle |k\rangle.
\]

For this state, we have

\[
I(A : C|B)_{\Psi} = 2 \log d - h(\lambda) - (1 - \lambda) \log (d^2 - 1),
\]

where \( h \) denotes the binary entropy defined by \( h(\lambda) = -\lambda \log \lambda - (1 - \lambda) \log (1 - \lambda) \). The state is a Markov state if and only if \( \lambda = 1/d^2 \). The distance to the closest Markov state is bounded from above by

\[
\left\| |\Psi_{\lambda}^{ABC} - |\Phi_d\rangle^{ABC}\rangle_{1} \right\| = 2 \sqrt{\frac{d^2 \lambda - 1}{d^2 - 1}}
\]

The reduced state on \( AC \) is given by

\[
\Psi_{\lambda}^{AC} = \frac{d^2 \lambda - 1}{d^2 - 1} |\Phi_d\rangle |\Phi_d\rangle^{AC} + \frac{1 - \lambda}{d^2 - 1} I^A \otimes I^C
\]

where \( \pi \) is the \( d \)-dimensional maximally mixed state and \( \lambda' = (d^2 \lambda - 1)/(d^2 - 1) \). Hence the Markovianizing cost does not depend on \( \lambda' \) when \( \lambda' > 0 \), as we proved in Section VI-B. As directly verified by considering the case of \( \lambda' = 1 \), the Markovianizing cost is equal to \( 2 \log d \) for \( \lambda' > 0 \). Taking the symmetry of \( |\Psi_{\lambda}\rangle \) between \( A \) and \( C \) into account, we obtain

\[
M_{A|B}(\Psi_{\lambda}) = M_{C|B}(\Psi_{\lambda}) = \begin{cases} 2 \log d & (\lambda > 1/d^2) \\ 0 & (\lambda = 1/d^2) \end{cases}.
\]

Hence the Markovianizing cost is not a continuous function of states. In a particular case where \( \lambda = 2/d^2 \), the Markovianizing cost grows logarithmically with respect to the dimension of the system, whereas QCMI, as well as the distance to the closest Markov state, approaches zero as indicated by (36) and (37).

We note that \( |\Psi_{\lambda}^{ABC}\rangle \) is “approximately recoverable” if \( \lambda \) is close to \( 1/d^2 \), i.e., it satisfies Equalities (36) approximately if \( \lambda \approx 1/d^2 \). Indeed, since \( |\Psi_{1/d^2}\rangle \) is a Markov state, there exist quantum operations \( R : B \to BC \) and \( R' : B \to AB \) such that

\[
|\Psi_{1/d^2}^{ABC}\rangle = R(\Psi_{1/d^2}^{AB}) = R'(\Psi_{1/d^2}^{BC}).
\]
Due to the triangle inequality and the monotonicity of the trace distance (see Appendix A), we have
\[
\|\Psi_{A}^{ABC} - \mathcal{R}(\Psi_{A}^{AB})\|_1 \leq \|\Psi_{A}^{ABC} - \Psi_{1/d^2}^{ABC}\|_1 + \|\mathcal{R}(\Psi_{1/d^2}^{AB}) - \mathcal{R}(\Psi_{A}^{AB})\|_1
\]
\[
+ \|\mathcal{R}(\Psi_{1/d^2}^{AB}) - \mathcal{R}(\Psi_{A}^{AB})\|_1 \leq 2\|\Psi_{A}^{ABC} - \Psi_{1/d^2}^{ABC}\|_1,
\]
Thus Equality \[37\] implies
\[
\|\Psi_{A}^{ABC} - \mathcal{R}(\Psi_{A}^{AB})\|_1, \|\Psi_{A}^{ABC} - \mathcal{R}'(\Psi_{A}^{BC})\|_1 \leq 4\sqrt{d^2\lambda - 1/d^2 - 1}.
\]

B. Asymmetry

We consider tripartite pure states that are expressed as
\[
|\Psi_{A}\rangle = \sqrt{\lambda}|0\rangle_{F_{d}}^{A} + \sqrt{1 - \lambda}|0\rangle_{F_{d}}^{A} |\Phi_{d}\rangle_{BC},
\]
where \(d = \dim\mathcal{H}_{C}, \dim\mathcal{H}_{A} = \dim\mathcal{H}_{B} = d + 1, 0 < \lambda < 1\) and \(\Phi_{d}\) is a maximally entangled state defined by \[35\]. The reduced state on \(A\) is given by
\[
\Psi_{A}^{AC} = \lambda|\Phi_{d}\rangle_{BC}^{AC} + (1 - \lambda)|0\rangle_{0}^{A} \otimes \pi^{C}.
\]
(39)
Note that \(|\Phi_{d}\rangle_{0}^{AC}\) does not have any \(|0\rangle^{A} \otimes |0\rangle^{C}\) component. Hence the CPTP maps on \(A\) defined as \(11\) and \(12\) are given by
\[
\mathcal{E}(\tau) = \text{Tr}[P_{1}\tau \cdot \pi_{1} + \text{Tr}[P_{0}\tau] \cdot |0\rangle\langle 0|] + \frac{1}{d}P_{1}\tau P_{0} + \frac{1}{d}P_{0}\tau P_{1}
\]
and
\[
\mathcal{E}_{\infty}(\tau) = \text{Tr}[P_{1}\tau \cdot \pi_{1} + \text{Tr}[P_{0}\tau] \cdot |0\rangle\langle 0|],
\]
respectively, where \(P_{0} = |0\rangle\langle 0|, P_{1} = I - P_{0}\) and \(\pi_{1} = P_{1}/d\). It is straightforward to verify that \(\mathcal{E}\) is self-adjoint. By applying \(\mathcal{E}_{\infty}\) to \(|\Psi_{A}\rangle\) on \(A\), we obtain that
\[
\Psi_{A}^{ABC} = \lambda \pi_{1}^{A} \otimes |0\rangle\langle 0| \otimes \pi^{C} + (1 - \lambda) |0\rangle\langle 0| \otimes |\Phi_{d}\rangle_{BC} |\Phi_{d}\rangle_{BC}.
\]
Therefore, due to Theorem \[10\] the Markovianizing cost is given by
\[
M_{A|B}(\Psi_{A}) = \lambda(\lambda) + 2\lambda \log d.
\]
On the other hand, from \[39\], the Markovianizing costs \(M_{C|B}(\Psi_{A})\) does not depend on \(\lambda\) when \(\lambda > 0\) as proved in Section \[4\B\] Thus we have
\[
M_{C|B}(\Psi_{A}) = \begin{cases} 2\log d & (\lambda > 0) \\ 0 & (\lambda = 0) \end{cases}
\]
in the same way as \[38\]. Hence the Markovianizing cost is not symmetric in \(A\) and \(C\), as opposed to QCMI, which satisfies \(I(A : C|B) = I(C : A|B)\).

C. States for which the Markovianizing cost coincides QCMI

We consider states that are expressed as
\[
|\Psi_{\{\lambda\}}\rangle := \sum_{k=1}^{d} \sqrt{\lambda_{k}} |k\rangle^{A} |k\rangle^{B} |k\rangle^{C},
\]
where \(\lambda_{k} \geq 0\) and \(\sum_{k=1}^{d} \lambda_{k} = 1\). These states satisfy conditions in Theorem \[13\] thus the Markovianizing cost is given by
\[
M_{A|B}(\Psi_{\{\lambda\}}) = I(A : C|B)\psi_{\{\lambda\}} = H(\{\lambda_{k}\})k.
\]

VII. CONCLUSIONS AND DISCUSSIONS

We have introduced the task of Markovianization, and derived a single-letter formula for the minimum cost of randomness required for Markovianizing tripartite pure states. We have also proposed an algorithm to compute the Markovianizing cost of a class of pure states without obtaining an explicit form of the Koashi-Imoto decomposition. We then have computed the Markovianizing cost for certain pure states, and revealed its discontinuity and asymmetry. Our results have an application in analyzing optimal costs of resources for simulating a bipartite unitary gate by local operations and classical communication \[19\]. Some open questions are generalization to mixed states, formulation of a classical analog of Markovianization, in addition to finding an alternative formulation of Markovianization for which we obtain QCMI as the cost function.

In \[18\], we have introduced and analyzed an alternative formulation of Markovization and the Markovianizing cost. Instead of requiring Condition \[9\], we require that the state after a random unitary operation is “approximately recoverable”, i.e., it satisfies Equalities \[7\] approximately. For pure states, we have proved that the Markovianizing cost in that case is equal to the one obtained in this paper.

ACKNOWLEDGMENT

The authors thank Tomohiro Ogawa and Masato Koashi for useful discussions.

REFERENCES

[1] P. Hayden, R. Jozsa, D. Petz, and A. Winter, “Structure of states which satisfy strong subadditivity of quantum entropy with equality,” Comm. Math. Phys., vol. 246, pp. 359–374, 2004.
[2] I. Devetak and J. Yard, “Exact cost of redistributing multipartite quantum states,” Phys. Rev. Lett., vol. 100, p. 230501, 2008.
[3] J. T. Yard and I. Devetak, “Optimal quantum source coding with quantum side information at the encoder and decoder,” IEEE Trans. Inf. Theory, vol. 55, pp. 5339–5351, 2009.
[4] F. Buscemi, “Complete positivity, markovianity, and the quantum data-processing inequality, in the presence of initial system-environment correlations,” Phys. Rev. Lett., vol. 113, p. 140502, 2014.
[5] D. Poulin and M. B. Hastings, “Markov entropy decomposition: A variational dual for quantum belief propagation,” Phys. Rev. Lett., vol. 106, p. 080403, 2011.
[6] B. Ibinson, N. Linden, and A. Winter, “Robustness of quantum markov chains,” Comm. Math. Phys., vol. 277, pp. 289–304, 2008.
A. Trace Distance and Uhlmann’s Theorem

The trace distance between two quantum states $\rho, \sigma \in S(H)$ is defined by

$$d(\rho, \sigma) := \frac{1}{2} \|\rho - \sigma\|_1 = \frac{1}{2} \text{Tr} \left[ \sqrt{(\rho - \sigma)^2} \right].$$

In the following, we omit the coefficient 1/2 for simplicity. For pure states $|\psi\rangle, |\phi\rangle \in H$, the trace distance takes a simple form

$$\|\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 = 2\sqrt{1 - |\langle\psi|\phi\rangle|^2}.$$

For $\rho, \sigma, \tau \in S(H)$, we have

$$\|\rho - \tau\|_1 \leq \|\rho - \sigma\|_1 + \|\sigma - \tau\|_1,$$

which is called the triangle inequality. The trace distance is monotonically nonincreasing under quantum operations, i.e., it satisfies

$$\|\rho - \sigma\|_1 \geq \|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1$$

for any linear CPTP map $\mathcal{E} : S(H) \to S(H')$. As a particular case, the trace distance between two states on a composite system is nonincreasing under taking the partial trace, that is, for $\rho, \sigma \in S(H^A \otimes H^B)$ we have

$$\|\rho^{AB} - \sigma^{AB}\|_1 \geq \|\rho^A - \sigma^A\|_1.$$

Consider two states $\rho, \sigma \in S(H^A)$ satisfying $\|\rho - \sigma\|_1 \leq \epsilon$, and let $|\psi\rangle^{AB}$ and $|\phi\rangle^{AB}$ be purifications of the two states, respectively. If $d_{B} \leq d_{B'}$, there exists an embedding of $H^B$ into $H^{B'}$, represented by an isometry from $H^B$ to $H^{B'}$, such that

$$\left\| |\psi\rangle^{AB}\langle\psi| - |\phi\rangle^{AB}\langle\phi| \right\|_1 \leq 2\sqrt{\epsilon}.$$  

This relation is referred to as Uhlmann’s theorem [24], see also Lemma 2.2 in [29]. In the case of $\epsilon = 0$, the above statement implies that all purifications are equivalent up to a local isometry.

The gentle measurement lemma (Lemma 9.4.1 in [28]) states that for any $\rho \in S(H)$, $X \in \mathcal{L}(H)$ and $\epsilon \geq 0$ such that $0 \leq X \leq I$ and $\text{Tr}[\rho X] \geq 1 - \epsilon$, we have

$$\left\| \rho - \sqrt{X}\rho\sqrt{X} \right\|_1 \leq 2\sqrt{\epsilon}.$$  

As a corollary, when two bipartite states $\rho \in S(H^A \otimes H^B)$ and $\sigma \in S(H^C \otimes H^{B'})$ satisfies $\|\rho^A - \sigma^A\|_1 \leq \epsilon$, and $\Pi_\sigma$ is the projection onto $\text{supp}[\sigma^A] \subseteq H^A$, we have

$$\|\rho^{AB} - \Pi_\sigma^A\rho^{AB}\Pi_\sigma^A \|_1 \leq 2\sqrt{\epsilon}.$$  

This is because we have

$$\epsilon \geq \|\Pi_\sigma^A \rho^A\Pi_\sigma^A + \Pi_\sigma^A \rho^{AB} \Pi_\sigma^A - \sigma^A\|_1 = \|\sigma^A - \Pi_\sigma^A \rho^A \Pi_\sigma^A\|_1 + \text{Tr}[\rho^A \Pi_\sigma^A],$$

where $\Pi_\sigma^A$ denotes the projection onto the orthogonal complement of $\text{supp}[\sigma^A] \subseteq H^A$, and thus have

$$\text{Tr}[\rho^{AB} \Pi_\sigma^A] = \text{Tr}[\rho^A \Pi_\sigma^A] = 1 - \text{Tr}[\rho^A \Pi_\sigma^A] \geq 1 - \epsilon.$$
B. Quantum Entropies and Mutual Informatons

The Shannon entropy of a probability distribution \( \{p_x\}_{x \in X} \) is defined as
\[
H(\{p_x\}_{x \in X}) := - \sum_{x \in X} p_x \log p_x.
\]
The von Neumann entropy of a quantum state \( \rho^A \in S(\mathcal{H}^A) \) is defined as
\[
S(\rho^A) = S(A)_{\rho} := -\text{Tr}[\rho^A \log \rho^A].
\]
If \( \rho^A \) is a probabilistic mixture of pure states as \( \rho^A = \sum_j p_j |\phi_j\rangle \langle \phi_j| \), we have \( S(\rho^A) \leq H(\{p_j\}) \) with equality if and only if \( \{ |\phi_j\rangle\}_j \) is mutually orthogonal. The von Neumann entropy is monotonically nondecreasing under random unitary operations, that is, we have \( S(A)_{\rho} \leq S(A)_{\nu(\rho)} \) for any random unitary operation \( \nu \) on \( A \). For a bipartite pure state \( |\psi\rangle^{AB^C} \), we have
\[
S(\rho^A) = S(\rho^C) = S(AB^C). (42)
\]
For a bipartite state \( \rho \in S(\mathcal{H}^A \otimes \mathcal{H}^B) \), the quantum conditional entropy and the quantum mutual information (QMI) are defined as
\[
S(A|B)_\rho = S(AB)_\rho - S(B)_\rho, \quad I(A : B)_\rho = S(A)_\rho + S(B)_\rho - S(AB)_\rho,
\]
respectively. The von Neumann entropy satisfies the subadditivity, expressed as
\[
S(A)_\rho + S(B)_\rho \geq S(AB)_\rho, \quad (43)
\]
which guarantees the nonnegativity of QMI. The equality holds if and only if \( \rho^{AB} = \rho^A \otimes \rho^B \). Applying (43) to \( |\psi\rangle^{ABC} \), which is a purification of \( \sigma^{ABC} \), and by using (42), we obtain
\[
S(C)_\sigma - S(A)_\sigma \leq S(AC)_\sigma. \quad (44)
\]
Hence QMI is bounded above as
\[
I(A : C)_\rho \leq \min\{2S(A)_\rho, 2S(C)_\rho\}. \quad (45)
\]
For any \( \rho \in S(\mathcal{H}^A \otimes \mathcal{H}^B) \) and quantum operation \( \mathcal{E} \) on \( B \), we have
\[
S(A|B)_\rho \leq S(A|B)_{\mathcal{E}(\rho)}, \quad I(A : B)_\rho \geq I(A : B)_{\mathcal{E}(\rho)}. \quad (46)
\]
Inequalities (46) are called the data processing inequality. For a tripartite state \( \rho \in S(\mathcal{H}^A \otimes \mathcal{H}^B \otimes \mathcal{H}^C) \), the quantum conditional mutual information (QCFMI) is defined as
\[
I(A : C|B)_\rho = S(AB)_\rho + S(BC)_\rho - S(B)_\rho - S(ABC)_\rho.
\]
QCFMI is nonnegative because of the strong subadditivity of the von Neumann entropy \([30]\), which is also equivalent to the data processing inequality. QMI and QCFMI are related by a simple relation as
\[
I(A : BC)_\rho = I(A : B)_\rho + I(A : C|B)_\rho,
\]
which is called the chain rule.

For a class of states called the classical-quantum states, the quantum conditional entropy and QCFMI take simple forms.

That is, for states \( \rho \in S(\mathcal{H}^X \otimes \mathcal{H}^A) \) and \( \sigma \in S(\mathcal{H}^X \otimes \mathcal{H}^A \otimes \mathcal{H}^B) \), given as
\[
\rho^{XA} = \sum_i p_i |i\rangle \langle i| \otimes \rho^A_i, \quad \sigma^{XAB} = \sum_i q_i |i\rangle \langle i| \otimes \sigma^{AB}_i,
\]
where \( \{ |i\rangle \} \) is an orthonormal basis of \( \mathcal{H}^X \), we have
\[
S(A|X)_\rho = \sum_i p_i S(\rho^A_i), \quad I(A : B|X)_\sigma = \sum_i q_i I(A : B|\sigma)_i.
\]
QMI of a classical-quantum state takes the form of
\[
I(X : A)_\rho = S(\tilde{\rho}^A) - \sum_i p_i S(\tilde{\rho}^A_i),
\]
where \( \tilde{\rho}^A = \sum_i p_i \rho^A_i \). This quantity is equal to the Holevo information \([31]\), and satisfies
\[
I(X : A)_\rho \leq S(X)_\rho = H(\{p_i\}_i)
\]
with equality if and only if \( \{\text{supp}[\rho^A_i]\}_i \) is mutually orthogonal.

C. Continuity of Quantum Entropies

Define
\[
\eta_0(x) := \begin{cases} -x \log x & (x \leq 1/e) \\ \frac{1}{e} & (x \geq 1/e) \end{cases},
\]
\[
\eta(x) = x + \eta_0(x) \quad \text{and} \quad h(x) := \eta_0(x) + \eta_0(1 - x), \quad \text{where } e \text{ is the base of the natural logarithm. For two states } \rho \text{ and } \sigma \text{ in a } d \text{-dimensional quantum system (} d < \infty \text{) such that } ||\rho - \sigma||_1 \leq \epsilon, \text{ we have}
\]
\[
|S(\rho) - S(\sigma)| \leq \epsilon \log d + \eta_0(\epsilon) \leq \eta(\epsilon) \log d,
\]
which is called the Fannes inequality \([32]\). It follows that for two bipartite states \( \rho, \sigma \in S(\mathcal{H}^A \otimes \mathcal{H}^B) \) such that \( ||\rho - \sigma||_1 \leq \epsilon \), we have
\[
|S(A|B)_\rho - S(A|B)_\sigma| \leq \eta(\epsilon) \log (d_A d_B^2) \quad (48)
\]
and
\[
|I(A : B)_\rho - I(A : B)_\sigma| \leq 2\eta(\epsilon) \log (d_A d_B). \quad (49)
\]

D. Typical Sequences and Subspaces \([33], [34]\), see also Appendices in \([12]\) for further details.

Let \( X \) be a discrete random variable with finite alphabet \( X \) and probability distribution \( p_x = \Pr\{X = x\} \) where \( x \in X \). A sequence \( x = (x_1, \ldots, x_n) \in X^n \) is said to be \( \delta \)-weakly typical with respect to \( \{p_x\}_{x \in X} \) if it satisfies
\[
2^{-n(H(X) + \delta)} \leq \prod_{i=1}^{N} p_{x_i} \leq 2^{-n(H(X) - \delta)}.
\]
The set of all δ-weakly typical sequences is called the δ-weakly typical set, and is denoted by \( T_{n,\delta} \) in the following. Denoting \( \prod_{i=1}^{N} p_{x_i} \) by \( p_{x} \), we have
\[
1 = \sum_{x \in \mathcal{X}^{n}} p_{x} \geq \sum_{x \in T_{n,\delta}} p_{x} \geq |T_{n,\delta}| \cdot 2^{-n(H(X)+\delta)},
\]
which implies that
\[
|T_{n,\delta}| \leq 2^{n(H(X)+\delta)}, \quad (50)
\]
A sequence \( x = (x_1, \cdots, x_n) \in \mathcal{X}^{n} \) is called δ-strongly typical with respect to \( \{p_{x}\}_{x \in \mathcal{X}} \) if it satisfies
\[
\frac{1}{n} N_{x|\bar{x}}|p_{x}| < \delta \quad \text{for all} \ x \in \mathcal{X} \text{ and } N_{x|\bar{x}} = 0 \text{ if } p_{x} = 0.
\]
Here, \( N_{x|\bar{x}} \) is the number of occurrences of the symbol \( x \) in the sequence \( \bar{x} \). The set of all δ-strongly typical sequences is called the δ-strongly typical set, and denoted by \( T_{n,\delta}^{s} \) in the following. From the weak law of large numbers, we have that for any \( \epsilon, \delta > 0 \) and sufficiently large \( n \),
\[
\Pr\{(X_1, \cdots, X_n) \in T_{n,\delta}^{s}\} \geq 1 - \epsilon, \quad (51)
\]
\[
\Pr\{(X_1, \cdots, X_n) \in T_{n,\delta}^{s}\} \geq 1 - \epsilon. \quad (52)
\]
Suppose the spectral decomposition of \( \rho \in S(\mathcal{H}) \) is given by \( \rho = \sum_{x} p_{x}|x\rangle\langle x| \). The δ-weakly typical subspace \( \mathcal{H}_{n,\delta} \subset \mathcal{H}_{\otimes n}^{s} \) with respect to \( \rho \) is defined as
\[
\mathcal{H}_{n,\delta} := \text{span}\{|x_1, \cdots, x_n\rangle \in \mathcal{H}_{\otimes n}^{s}|x_1, \cdots, x_n\rangle \in T_{n,\delta}^{s}\},
\]
where \( T_{n,\delta}^{s} \) is the δ-strongly typical set with respect to \( \rho_{x} \). Similarly, the δ-strongly typical subspace \( \mathcal{H}_{n,\delta}^{s} \subset \mathcal{H}_{\otimes n}^{s} \) with respect to \( \rho \) is defined as
\[
\mathcal{H}_{n,\delta}^{s} := \text{span}\{|x_1, \cdots, x_n\rangle \in \mathcal{H}_{\otimes n}^{s}|x_1, \cdots, x_n\rangle \in T_{n,\delta}^{s}\}.
\]
Suppose the Schmidt decomposition of \( |\psi\rangle^{AB} \in \mathcal{H}_{\otimes n}^{s} \) is given by \( |\psi\rangle^{AB} = \sum_{x} \sqrt{p_{x}}|x\rangle^{A}|x\rangle^{B} \). For any \( \delta > 0 \) and \( n \), let \( \mathcal{H}_{n,\delta} \) and \( \mathcal{H}_{n,\delta}^{s} \) be the δ-weakly and strongly typical subspace of \( \mathcal{H}_{\otimes n}^{s} \) with respect to \( |\psi\rangle^{A} = \text{Tr}_{B}(|\psi\rangle^{AB}) \), and let \( \Pi_{n,\delta} \) and \( \Pi_{n,\delta}^{s} \) be the projection onto those subspaces, respectively. From \( (50) \), we have
\[
\text{rank} \Pi_{n,\delta} = \dim \mathcal{H}_{n,\delta} \leq 2^{n(H(A)_{\psi}+\delta)}.
\]
From \( (51) \) and \( (52) \), we have
\[
\text{Tr}[\Pi_{n,\delta}^{s}|\psi\rangle^{AB}]_{\otimes n} = \sum_{x \in T_{n,\delta}^{s}} p_{x} \geq 1 - \epsilon, \quad (53)
\]
\[
\text{Tr}[\Pi_{n,\delta}^{s}|\psi\rangle^{AB}]_{\otimes n} = \sum_{x \in T_{n,\delta}^{s}} p_{x} \geq 1 - \epsilon. \quad (54)
\]
for any \( \epsilon, \delta > 0 \) and sufficiently large \( n \).

**Appendix B**

**Proof of Theorem 9**

In this Appendix, we show a detailed proof of Theorem 9. In the following, we informally denote the composite systems \( a_{i}, a_{i}, a_{i} \) by \( A \) and \( b_{i}, b_{i}, b_{i} \) by \( B \), when there is no fear of confusion.

### A. Proof of Achievability (Inequality 16)

Fix arbitrary \( n \) and \( \delta \in (0,1] \). Let \( J_{n,\delta} \subset J^{n} \) be the δ-strongly typical set with respect to \( \{p_{j}\}_{j \in J} \). For each \( j \in J \) and \( j = j_{1} \cdots j_{n} \in J_{n,\delta} \), define \( \mathcal{E}_{j,j} := \{j_{1} = j, 1 \leq l \leq n\} \). The number of elements in the set is bounded as
\[
n \left( p_{j} - \frac{\delta}{|J|} \right) \leq |\mathcal{E}_{j,j}| \leq n \left( p_{j} + \frac{\delta}{|J|} \right). \quad (50)
\]
For each \( j \in J_{n,\delta} \), we sort \( \left( H^{a_{j}} \right)_{\otimes n} = \mathcal{H}_{\otimes n}^{a_{j}} \otimes \cdots \otimes \mathcal{H}_{\otimes n}^{a_{n}} \) as
\[
\left( H^{a_{j}} \right)_{\otimes n} = \bigotimes_{j \in J_{n,\delta}} \left( \bigotimes_{\mathcal{E}_{j,j}} \mathcal{H}_{\otimes n}^{a_{j}} \right). \quad (55)
\]
For each \( j \) and \( j \), let \( \mathcal{H}_{j,j,\delta} \) be the δ-weakly typical subspace of \( \otimes_{j \in J_{n,\delta}} \mathcal{H}_{\otimes n}^{a_{j}} \) with respect to \( \varphi_{j}^{a_{j}}, \Pi_{j,j,\delta} \) be the projection onto \( \mathcal{H}_{j,j,\delta} \), and let \( \Pi_{j,j,\delta}^{s} := \otimes_{j \in J_{n,\delta}} \Pi_{j,j,\delta} \). Define
\[
\Pi_{\delta}^{s} := \sum_{j \in J_{n,\delta}} |j\rangle\langle j|^{\otimes n} \otimes \Pi_{j,j,\delta}^{s} \quad \text{as (18).}
\]
and
\[
|\Psi_{n,\delta}^{AB}C :\rangle := \otimes_{j \in J_{n,\delta}} |\Psi_{j,j,\delta}^{AB}C :\rangle = \sum_{j \in J_{n,\delta}} \sqrt{p_{j}}|j\rangle^{\otimes n} \otimes \Pi_{j,j,\delta}^{s} \otimes \varphi_{j}^{a_{j}}C_{j}^{\otimes n}, \quad (56)
\]
where we introduced notations \( p_{j} = p_{j_{1}} \times \cdots \times p_{j_{n}}, \varphi_{j} = \varphi_{j_{1}} \otimes \cdots \otimes \varphi_{j_{n}}, \) and \( \omega_{j} = \omega_{j_{1}} \otimes \cdots \otimes \omega_{j_{n}} \).

Let \( v_{j} \) be any unitary acting on \( \otimes_{j \in J_{n,\delta}} \mathcal{H}_{j,j,\delta} \), and define a unitary on \( \otimes_{j \in J_{n,\delta}} \otimes_{\mathcal{E}_{j,j}} \mathcal{H}_{\otimes n}^{a_{j}} \) by
\[
V_{\delta}^{A} := \sum_{j \in J_{n,\delta}} |j\rangle\langle j|^{\otimes n} \otimes \Pi_{j,j,\delta}^{s} \otimes v_{j}^{\otimes n}, \quad (57)
\]
for any \( \epsilon, \delta > 0 \) and sufficiently large \( n \).

### B. Proof of Theorem 9

In the following, we informally denote the composite systems \( a_{i}, a_{i}, a_{i} \) by \( A \) and \( b_{i}, b_{i}, b_{i} \) by \( B \), when there is no fear of confusion.
which is a subnormalized Markov state conditioned by $\tilde{B}$ corresponding to [19] (see Figure 3).

The minimum nonzero eigenvalue of $\tilde{\Psi}_{n,\delta}$ is calculated as follows. First, due to the definition of $J_{n,\delta}$, we have

$$p_j \geq \prod_{j \in J} p_j^{n(p_j+\delta/|J|)} = 2^{-n(H((p_j)_j)+\delta H'((p_j)_j))},$$

where

$$H'((p_j)_j) := \frac{1}{|J|} \sum_j \log p_j > -\infty.$$ 

Second, since the spectrums of $\varphi_{j}^{\bar{n}_{a}}$ and $\varphi_{j}^{\bar{n}_{a}\bar{C}}$ are the same, the minimum nonzero eigenvalue $\mu_j$ of $\varphi_{j}^{\bar{n}_{a}\bar{C}}$ is bounded from below as

$$\mu_j \geq \prod_{j \in J} 2^{-\nu_j |S(\varphi_{j}^{\bar{n}_{a}})+\delta)}$$

$$\geq \prod_{j \in J} 2^{-n(p_j+\delta/|J|)(S(\varphi_{j}^{\bar{n}_{a}})+\delta)}$$

$$\geq 2^{-n \sum_j p_j S(\varphi_{j}^{\bar{n}_{a}})+\delta \log (4d_A)},$$

where the last line follows from

$$\sum_j \left( p_j + \frac{\delta}{|J|} \right) \left( S(\varphi_{j}^{\bar{n}_{a}})+\delta \right)$$

$$= \sum_j p_j S(\varphi_{j}^{\bar{n}_{a}})+\delta \left( 1 + \frac{\delta}{|J|} \sum_j S(\varphi_{j}^{\bar{n}_{a}})+\delta \right)$$

$$\leq \sum_j p_j S(\varphi_{j}^{\bar{n}_{a}})+\delta \left( 2 + \log d_A \right)$$

$$= \sum_j p_j S(\varphi_{j}^{\bar{n}_{a}})+\delta \log (4d_A).$$

Third, we have

$$\text{rank } \Pi_{j,j,\delta} \leq 2^{|J_j|} \leq 2^n \left( n \log (2d_A) \right),$$

and

$$\text{rank } \Pi_{j,\delta} \leq \prod_{j \in J} \text{rank } \Pi_{j,j,\delta} \leq \prod_{j \in J} 2^n \left( n \log (2d_A) \right).$$

Thus the nonzero eigenvalue $\nu_j$ of $\bar{n}_{a}$ is, in the same way as $\mu_j$, bounded from below as

$$\nu_j \geq \prod_{j \in J} 2^{-n(p_j+\delta/|J|)(S(\varphi_{j}^{\bar{n}_{a}})+\delta)}.$$ 

All in all, the minimum nonzero eigenvalue $\lambda$ of $\tilde{\Psi}_{n,\delta}$ is bounded as

$$\lambda = p_j H_{j,j'j} \geq 2^{-n\left( H((p_j)_j)+2 \sum_j p_j S(\varphi_{j}^{\bar{n}_{a}})+\delta \left( H'((p_j)_j)+2 \log (4d_A) \right) \right)}.$$ 

We also have

$$\text{rank } \tilde{\Psi}_{n,\delta} \leq |J_{n,\delta}| \times \text{rank } \bar{n}_{a} \times \text{rank } \varphi_{j}^{\bar{n}_{a}} \leq d_A^3.$$ 

Suppose $V_1, \ldots, V_N$ are unitaries that are randomly and independently chosen from the ensemble $\{p(dV), V\}$. Due to the operator Chernoff bound (Lemma 3 in [11]), we have

$$\Pr \left\{ \frac{1}{N} \sum_{i=1}^{N} \Psi'_{n,\delta}(V_i) \notin \left( (1-\epsilon_1)\tilde{\Psi}_{n,\delta}, (1+\epsilon_1)\tilde{\Psi}_{n,\delta} \right) \right\}$$

$$\leq 2d_A^n \exp \left( -\frac{N \lambda^2}{2} \right)$$

for any $\epsilon_1 \in (0, 1]$, which implies that

$$\Pr \left\{ \left\| \frac{1}{2nR} \sum_{i=1}^{2nR} \Psi'_{n,\delta}(V_i) - \tilde{\Psi}_{n,\delta} \right\|_1 \leq 2\epsilon_1 \right\}$$

$$\geq 1 - 2d_A^n \exp \left( -\frac{2nR \lambda^2}{2} \right)$$

(60)

for an arbitrary $R > 0$. Therefore, if $R$ satisfies

$$R > H((p_j)_j) + 2 \sum_j p_j S(\varphi_{j}^{\bar{n}_{a}}) + \delta \left( H'((p_j)_j) + 2 \log (4d_A) \right),$$

and if $n$ is sufficiently large so that the R.H.S. in (60) is greater than 0, there exists a set of unitaries $\{V_i\}_{i=1}^{2nR}$ such that

$$\left\| \frac{1}{2nR} \sum_{i=1}^{2nR} \Psi'_{n,\delta}(V_i) - \tilde{\Psi}_{n,\delta} \right\|_1 \leq 2\epsilon_1.$$ 

(62)

Using unitaries in the set, construct a random unitary operation $\mathcal{V}_n$ on $\hat{a}_0\hat{a}_2\hat{a}_R$ as $\mathcal{V}_n(\cdot) = 2^{-nR} \sum_{k=1}^{2nR} V_k(\cdot)V_k^\dagger$. Let us evaluate the total error. First, from (53), (54), (55) and (62), we have

$$D_{n,\delta} := \text{Tr}[\Pi_{\delta}^{\dagger}\Psi_{KL}^{\otimes n}(\Psi_{KL}^{\otimes n})] = \langle \Psi_{n,\delta}^\dagger | \langle \Psi_{n,\delta} | \geq 1 - \epsilon_2$$

(63)

for any $\epsilon_2 > 0$ and sufficiently large $n$. Thus, by the gentle measurement lemma (40), we have

$$\left\| \frac{\Psi'_{n,\delta}}{D_{n,\delta}} - \frac{\Psi_{n,\delta}}{D_{n,\delta}} \right\|_1 \leq 2\sqrt{\epsilon_2},$$

which leads to

$$\left\| \mathcal{V}_n \left( \frac{\Psi'_{n,\delta}}{D_{n,\delta}} \right) - \mathcal{V}_n \left( \frac{\Psi_{n,\delta}}{D_{n,\delta}} \right) \right\|_1 \leq 2\sqrt{\epsilon_2}.$$ 

Second, from (62) and (63), we have

$$\left\| \mathcal{V}_n \left( \frac{\Psi'_{n,\delta}}{D_{n,\delta}} \right) - \frac{\Psi_{n,\delta}}{D_{n,\delta}} \right\|_1 \leq \frac{2\epsilon_1}{D_{n,\delta}} \leq \frac{2\epsilon_1}{1 - \epsilon_2}.$$ 

Therefore, by the triangle inequality, we obtain

$$\left\| \mathcal{V}_n \left( \frac{\Psi_{n,\delta}}{D_{n,\delta}} \right) - \frac{\Psi_{n,\delta}}{D_{n,\delta}} \right\|_1 \leq 2\sqrt{\epsilon_2} + \frac{2\epsilon_1}{1 - \epsilon_2}.$$ 

(64)

From (58) and (59), we have $\text{Tr}[\Psi_{n,\delta}] = D_{n,\delta}$, which implies that $\Psi_{n,\delta}/D_{n,\delta}$ is a normalized Markov state conditioned by $B^n$. Since the relation (64) holds for any $\epsilon_1, \epsilon_2 > 0$, $R > H((p_j)_j) + 2 \sum_j p_j S(\varphi_{j}^{\bar{n}_{a}})$, any $\delta \in (0, 1]$ that satisfies (61) and sufficiently large $n$, we obtain (16).
B. Convergence Speed of the Error

We prove that, in the direct part of Theorem 9, the error $\epsilon$ vanishes exponentially in the asymptotic limit of $n \to \infty$. More precisely, we prove the following theorem.

**Theorem 14** There exists a constant $c_\psi > 0$ such that for any $R > M_{A|B}(\Psi^{ABC})$, sufficiently small $\delta > 0$ and any sufficiently large $n$, we find a random unitary operation $V_n: \tau \mapsto 2^{-nR} \sum_{k=1}^{2^n} V_k \tau V_k^\dagger$ on $A^n$ and a Markov state $\Upsilon nA^nB^nC^n$ conditioned by $B^n$ that satisfy
\[
\|V_n(\rho^{\otimes n}) - \Upsilon nA^nB^nC^n\|_1 \leq 6 \exp\left(-\frac{c_\psi \delta^2 n}{2}\right).
\]

**Proof:** Let $X_1, \cdots, X_n$ be a sequence of i.i.d. random variables obeying a probability distribution $\{p_x\}_x$. It is proved in [35] that there exists a constant $c > 0$, which depends on $\{p_x\}_x$, such that for any $\delta > 0$ and $n$, we have
\[
\Pr\{\{X_1, \cdots, X_n\} \in T_{n,\delta}\} \geq 1 - \exp(-c\delta^2 n),
\]
\[
\Pr\{\{X_1, \cdots, X_n\} \in T_{n,\delta}\} \geq 1 - \exp(-c\delta^2 n).
\]
As a consequence, there exists a constant $c_\psi > 0$ such that we have
\[
\text{Tr}_n[2\delta n] \|\rho^{\otimes n}\| \leq 1 - \exp(-c_\psi \delta^2 n),
\]
\[
\text{Tr}_n[2\delta n] \|\rho^{\otimes n}\| \leq 1 - \exp(-c_\psi \delta^2 n).
\]
for any $\delta > 0$ and $n$, corresponding to (63) and (64). Thus, for any $\delta > 0$, $n$, and $D_{n,\delta}$ defined by (63), we obtain
\[
D_{n,\delta} \geq 1 - \exp(-c_\psi \delta^2 n),
\]
where $c_\psi > 0$ is a constant. Hence we have
\[
\|V_n(\rho^{\otimes n}) - \Upsilon nA^nB^nC^n\|_1 \leq 6 \exp\left(-\frac{c_\psi \delta^2 n}{2}\right)
\]
for any $\delta, c_\psi > 0$ and $n$, corresponding to (64). Substituting $\exp(-c_\psi \delta^2 n/2)$ into $c_\psi$, we obtain
\[
\|V_n(\rho^{\otimes n}) - \Upsilon nA^nB^nC^n\|_1 \leq 6 \exp\left(-\frac{c_\psi \delta^2 n}{2}\right)
\]
for any $\delta > 0$ and $n \geq (\ln 2)/c_\psi \delta^2$.

For an arbitrary $R$, choose sufficiently small $\delta > 0$ such that
\[
R > H(\{p_j\}_j) + 2 \sum_j p_j S(\sigma_j^{\psi n}) + \delta (H'(\{p_j\}_j) + 2 \log (4d_A) + c_\psi \delta).
\]
Inequality (65) then holds for sufficiently large $n$, while keeping the R.H.S. in (64) strictly greater than 0. This completes the proof. ■

C. Proof of Optimality (Inequality 20)

We assume, without loss of generality, that $d_A \geq d_Bd_C$. This condition is always satisfied by associating a sufficiently large Hilbert space $H_A$ to system $A$.

Take an arbitrary $R > M_{A|B}(\Psi^{ABC})$. By definition, for any $\epsilon \in (0, 1]$ and sufficiently large $n$, there exist a random unitary operation $V_n: \tau \mapsto 2^{-nR} \sum_{k=1}^{2^n} V_k \tau V_k^\dagger$ on $\tilde{A}$ and a Markov state $\Upsilon nA^nB^nC^n$ conditioned by $B^n$ such that
\[
\|V_n(\rho^{\otimes n}) - \Upsilon nA^nB^nC^n\|_1 \leq \epsilon.
\]
By tracing out $A^n$, we have
\[
\|\rho^{\otimes n} - \Upsilon nB^nC^n\|_1 \leq \epsilon.
\]
Due to Uhlmann’s theorem [24], see Appendix A-A, there exists a purification $\chi$ of $\Upsilon nB^nC^n$ such that we have
\[
\|\rho^{\otimes n} - \Upsilon nB^nC^n\|_1 \leq 2\sqrt{\epsilon}.
\]
Let $\tilde{\Gamma}': \tilde{H}_A \to \tilde{H}_B \otimes \tilde{H}_a \otimes \tilde{H}_b$ be the KI isometry on $\tilde{B}$ with respect to $\chi B^n$. From Lemma 11, there exists such a sub-KI isometry $\Gamma': \tilde{H}_A \to \tilde{H}_a \otimes \tilde{H}_a \otimes \tilde{H}_b$ such that the KI decomposition of $\chi$ on $\tilde{B}$ and $\tilde{A}$ is given by
\[
\|\chi_{\tilde{A}B}\| = \|\tilde{\Gamma}' \otimes \tilde{\Gamma}B\| = \sum_i \sqrt{q_i} |i\rangle_{\tilde{A}} |i\rangle_{\tilde{B}} |i\rangle_{\tilde{B}} \otimes |i\rangle_{\tilde{B}} \chi_{\tilde{A}B} C^n.
\]
From Theorem 7 and $\chi B^n = \Upsilon nB^nC^n$, a Markov decomposition of $\Upsilon nB^nC^n$ is obtained by $\tilde{\Gamma}'$ as
\[
\|\chi_{\tilde{A}B}\| = \|\tilde{\Gamma}' \otimes \tilde{\Gamma}B\| = \sum_i \sqrt{q_i} |i\rangle_{\tilde{A}} |i\rangle_{\tilde{B}} \otimes |i\rangle_{\tilde{B}} \chi_{\tilde{A}B} C^n.
\]
Thus from (66) and the triangle inequality, we obtain
\[
\|V_n(\rho^{\otimes n}) - \Upsilon nA^nB^nC^n\|_1 \leq 2\sqrt{\epsilon} + \epsilon < 3\sqrt{\epsilon}.
\]
Applying $\tilde{\Gamma}'$ yields
\[
\|V_n(\rho^{\otimes n}) - \Upsilon nA^nB^nC^n\|_1 \leq 3\sqrt{\epsilon},
\]
due to (70). Hence we obtain from (69) that
\[
\|V_n(\rho^{\otimes n}) - \Upsilon nA^nB^nC^n\|_1 \leq 3\sqrt{\epsilon}.
\]
Let $D_B$ be the completely dephasing operation on $\tilde{b}_n$ with respect to the basis $\{|b_i\}_n$. From (70), we have $D_B(\Upsilon nA^nB^nC^n) = \Upsilon nA^nB^nC^n$. Thus we obtain from (71) that
\[
\|\{T \otimes D_B\}|\chi_{\tilde{A}}\rangle\rangle - \Upsilon nA^nB^nC^n\|_1 \leq 3\sqrt{\epsilon}.
\]
Here, we defined a random isometry operation $T_n = V_n \circ \mathcal{E}$, where $\mathcal{E}$ is an isometry operation corresponding to $\tilde{\Gamma}'$. ■
Due to (69), we have
\[ D_{\hat{b}}(\{X_{\kappa},X_{\lambda}\}) = \sum_i q_i |i\rangle |i\rangle^{\hat{b}} \otimes |\xi\rangle_{\kappa} |\xi\rangle_{\lambda} |\phi_i\rangle_{\hat{a}} |\phi_i\rangle^{\hat{b}} |\phi_i\rangle^{\hat{a}} |\phi_i\rangle^{\hat{b}} |\phi_i\rangle^{\hat{a}} \hat{b} |\phi_i\rangle^{\hat{b}} |\phi_i\rangle^{\hat{a}} \hat{C}, \]
which leads to
\[ \text{Tr}_{\hat{b}\hat{C}} \left[ D_{\hat{b}}(\{X_{\kappa},X_{\lambda}\}) \right] \]
\[ = \sum_i q_i |i\rangle |i\rangle^{\hat{b}} \otimes |\xi\rangle_{\kappa} |\xi\rangle_{\lambda} |\phi_i\rangle_{\hat{a}} |\phi_i\rangle^{\hat{b}} |\phi_i\rangle^{\hat{a}}. \]
Hence we have
\[ \text{Tr}_{\hat{b}\hat{C}} \left[ (T_n^T \otimes D_{\hat{b}})(\{X_{\kappa},X_{\lambda}\}) \right] = \sum_i q_i |i\rangle |i\rangle^{\hat{b}} \otimes \phi_i^{\hat{a}} \phi_i^{\hat{b}}. \]
where we define
\[ \phi_i^{\hat{a}} = T_n'(|i\rangle |\xi\rangle_{\kappa} |\xi\rangle_{\lambda} |\phi_i\rangle^{\hat{a}} \phi_i^{\hat{b}}). \]
(73)
From (70) we have
\[ \text{Tr}_{\hat{b}\hat{C}} \left[ \sum_i q_i |i\rangle |i\rangle^{\hat{b}} \otimes |\phi_i\rangle^{\hat{a}} \phi_i^{\hat{b}} \right] \leq 3. \]
Hence we obtain
\[ \frac{1}{n} S(A\hat{B}\hat{C}) = \frac{1}{n} S(A\hat{b} |\phi_i\rangle^{\hat{a}} |\phi_i\rangle^{\hat{b}}), \]
where (75) follows from (70). Hence we have
\[ \frac{1}{n} S(A\hat{B}\hat{C}) \geq \frac{1}{n} S(A\hat{b} |\phi_i\rangle^{\hat{a}} |\phi_i\rangle^{\hat{b}}), \]
which implies (20) by taking the limit of \( \epsilon \to 0 \).

D. Proof of Lemma 12

The key idea for the proof of Lemma 12 is similar to the one used in [25]. Let \( \psi^{AC'} \) be a state such that the KI isometry on \( A \) with respect to \( \psi^{AC'} \) is the same as that with respect to \( \psi^{AC} \), and that it is decomposed as
\[ \psi_{KI}^{AC'} := (\Gamma^\Lambda \otimes \psi_{C'}^{AC'}) \psi^{AC'} (\Gamma^\Lambda \otimes \psi_{C'}^{AC'})^\dagger = \sum_{j} p_j |j\rangle |\phi_j\rangle |\bar{\phi}_j\rangle |\bar{\phi}_j\rangle^{ac'} \otimes |j\rangle |j\rangle^{ac'}, \]
(77)
where \( \psi_{KI} : \mathcal{H}^C \rightarrow \mathcal{H}^{C'} \otimes \mathcal{H}^{C'} \) is an isometry and \( |\bar{\phi}_j\rangle^{ac'} \) is a purification of \( \phi_j^{ac'} \). The state satisfies \( \psi^A = \Psi^A \). Note that we have
\[ d_{C'} = \sum_{j} \text{rank} |\bar{\phi}_j\rangle \leq d_A. \]
Let \( \mathcal{E} \) be the set of all linear CPTP maps on \( \mathcal{H}^A \), and define two functions \( f,g : \mathcal{E} \rightarrow \mathbb{R} \) by
\[ f(\mathcal{E}) = \| \mathcal{E}(\psi^{AC}) - \Psi^{AC} \|_1, \]
\[ g(\mathcal{E}) = \| \mathcal{E}(\psi^{AC'}) - \psi^{AC'} \|_1. \]
Since the KI decomposition of \( A \) with respect to \( \psi^{AC} \) and that with respect to \( \psi^{AC'} \) are the same, \( f(\mathcal{E}) = 0 \) if and only if \( g(\mathcal{E}) = 0 \) (see Lemma 5). Define
\[ \psi_{KI}^{\epsilon} := \sup_{\mathcal{E} \in \mathcal{E}} \{ g(\mathcal{E}) | f(\mathcal{E}) \leq \epsilon \}. \]
(78)
This is a monotonically nondecreasing function of \( \epsilon \) by definition, and satisfies \( \lim_{\epsilon \to 0} \psi_{KI}^{\epsilon} = 0 \) as we prove in Appendix B-E.

We consider a general situation in which the relation (24) does not necessarily hold. Let \( \Pi_{\chi} \) be the projection onto \( \mathcal{H}^A = \mathcal{H}^A_{\chi} \) and \( \Pi^\perp_{\chi} \) be that onto its orthogonal complement. Using a quantum channel \( \mathcal{E}_{\chi} \) on \( \mathcal{H}^A_{\chi} \) defined by (26), construct another quantum channel \( \mathcal{E}_{\chi}^* \) on \( \mathcal{H}(\mathcal{H}^A) \) by
\[ \mathcal{E}_{\chi}^*(\tau) = \mathcal{E}_{\chi}(\Pi_{\chi} \tau \Pi_{\chi}) + \Pi^\perp_{\chi} \tau \Pi^\perp_{\chi} (\forall \tau \in \mathcal{H}(\mathcal{H}^A)). \]
(79)
Define quantum channels $\mathcal{E}_l$ on $A_l$ ($1 \leq l \leq n$) by

$$\mathcal{E}_l(\tau^{A_l}) = \text{Tr}_{\bar{A} \setminus A_l}[\mathcal{E}_x^{\ast}(\Psi^{A_1} \otimes \cdots \otimes \Psi^{A_{l-1}} \otimes \tau^{A_l} \otimes \Psi^{A_{l+1}} \otimes \cdots \otimes \Psi^{A_n})],$$

where $\text{Tr}_{\bar{A} \setminus A_l}$ denotes the partial trace over $A_1 \cdots A_{l-1}A_{l+1} \cdots A_n$. From (25), we have $\mathcal{E}_x^{\ast}(\chi \otimes \tilde{\chi}) = \mathcal{E}_x^{\ast}(\tilde{\chi} \otimes \chi)$, and thus from (22) and the triangle inequality, we have

$$\left\| \mathcal{E}_x^{\ast}(\Psi^{\otimes n})^{\tilde{\chi}} - (\Psi^{\otimes n})^{\chi} \right\|_1 \leq 2\epsilon,$$

which implies

$$\left\| \mathcal{E}_l(\Psi^{A_l}) - \Psi^{A_l} \right\|_1 \leq 2\epsilon$$

by taking the partial trace. Thus we have

$$\left\| \mathcal{E}_l(\Psi^{A_l}) - \Psi^{A_l} \right\|_1 \leq \zeta_\psi(2\epsilon)$$

for any $1 \leq l \leq n$. By Inequality (49) and $d_{\mathcal{C}_l} \leq d_A$, it follows that

$$I(A : C')_\psi - I(A_l : C_l')\mathcal{E}_l(\psi) \leq 4n\eta(\zeta_\psi(2\epsilon)) \log d_A,$$

and consequently, that

$$nI(A : C')_\psi - \sum_{l=1}^n I(A_l : C_l')\mathcal{E}_l(\psi) \leq 4n\eta(\zeta_\psi(2\epsilon)) \log d_A.$$ (80)

We also have

$$I(\tilde{A} : \tilde{C'})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})} = S(\tilde{C'})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})} - S(\tilde{C'}|\tilde{A})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})}$$

$$= S(\tilde{C'})_{\mathcal{H}^{\otimes n}} - \sum_{l=1}^n S(C_l'|A_1 \cdots A_n C_l' \cdots C_{l-1}'|\tilde{C})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})}$$

$$\geq \sum_{l=1}^n S(C_l'|\tilde{A})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})} - \sum_{l=1}^n S(A_l C_l'|\tilde{A})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})}$$

$$= \sum_{l=1}^n S(C_l'|\tilde{A})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})} - \sum_{l=1}^n S(C_l'|A_l)_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})}$$

$$= \sum_{l=1}^n I(A_l : C_l')_{\mathcal{E}_l(\psi)}.$$ (81)

Here, we used the fact that $\mathcal{E}_x^{\ast}$ on $\tilde{A}$ does not change the reduced state on $C'$, and that

$$\text{Tr}_{\bar{A}}[\mathcal{E}_x^{\ast}(\psi^{\otimes n})] = \mathcal{E}_l(\psi),$$

because of $\Psi^{\otimes n}_{\tilde{A}} = \Psi^A_{\tilde{A}}$. Combining (80) and (81), we obtain

$$nI(A : C')_\psi \leq I(\tilde{A} : \tilde{C'})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})} + 4n\eta(\zeta_\psi(2\epsilon)) \log d_A.$$ (82)

Define

$$\psi^{\tilde{A}C'}_{n,\chi,\psi} := \frac{\Pi^A(\psi^{AC'})^{\otimes n} \Pi^A}{\text{Tr}(\Psi^{\otimes n})^{\otimes n}}$$

and

$$\psi^{\tilde{A}C'}_{n,\psi} := (\mathcal{E}_2 \circ \mathcal{E}_1 \circ \mathcal{E}_{\Gamma_{\chi}})(\psi^{\tilde{A}C'})$$

as depicted in Figure 5. From Condition (22) and $\psi^A = \Psi^A$, we have $\|\psi^{\otimes n}_{\tilde{A}} - \chi^A\|_1 \leq \epsilon$. Thus, due to (79), we have

$$\left\| \mathcal{E}_x(\psi^{AC'}_{n,\chi}) - \mathcal{E}_x(\psi^{\otimes n})^{\tilde{A}C'} \right\|_1 \leq \left\| \psi^{AC'}_{n,\chi} - (\psi^{AC'})^{\otimes n} \right\|_1 \leq 2\sqrt{\epsilon},$$

which leads to

$$I(\tilde{A} : \tilde{C'})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})} \leq I(\tilde{A} : \tilde{C'})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})} + 4n\eta(2\sqrt{\epsilon}) \log d_A.$$ (83)

by Inequality (49) and $d_{\mathcal{C}_l} \leq d_A$. By the data processing inequality, we also have

$$I(\tilde{A} : \tilde{C'})_{\mathcal{E}_x^{\ast}(\mathcal{H}^{\otimes n})} \leq I(\tilde{A}_i \tilde{a}_R : \tilde{C'})_{\psi}. $$ (84)

Consequently, we obtain from (82), (83) and (84) that

$$nI(A : C')_\psi \leq I(\tilde{A}_i \tilde{a}_R : C')_{\psi} + 4n(\eta(2\sqrt{\epsilon}) + \eta(\zeta_\psi(2\epsilon))) \log d_A.$$ (85)

The QMIs in (85) are calculated as follows. First, from (25), we have

$$I(\tilde{E}_2 \circ \tilde{E}_1 \circ \tilde{E}_{\Gamma_{\chi}})(\tilde{\psi}) = \sum_{i=1}^d q_i |\tilde{\psi},i^\tilde{A} \otimes \tilde{\psi}_i^{\tilde{a}}.$$ (86)

Therefore, from the monotonicity of the trace distance under $\tilde{E}_2 \circ \tilde{E}_1 \circ \tilde{E}_{\Gamma_{\chi}}$, Equality (22) and $\psi^\tilde{A} = \psi^A$, we obtain

$$\left\| \psi^{\tilde{A}C'}_{n,\psi} - \chi^{\tilde{A}C'} \right\|_1 \leq \left\| \psi^{\otimes n}_{\tilde{A}} - \chi^A \right\|_1 \leq \epsilon.$$ (87)

Due to (47), (48) and

$$\dim \mathcal{H}^{\tilde{A}}, \dim \mathcal{H}^{\tilde{a}} \leq \dim \mathcal{H}^A = d_A^n,$$

implies that

$$|S(\tilde{a}_0)_{\chi,\mathcal{K}^A} - S(\tilde{a}_0)_{\psi}|_1 \leq n\eta(\epsilon) \log d_A,$$

$$|S(\tilde{a}_0)_{\chi,\mathcal{K}^A} - S(\tilde{a}_R)_{\psi}|_1 \leq 3n\eta(\epsilon) \log d_A,$$

and consequently, that

$$S(\tilde{a}_0)_{\psi} + 2S(\tilde{a}_R)_{\psi} \leq S(\tilde{a}_0)_{\chi,\mathcal{K}^A} + 2S(\tilde{a}_R)_{\chi,\mathcal{K}^A} + 7n\eta(\epsilon) \log d_A.$$ (88)

Since $\psi^{\tilde{a},\tilde{a}C'}$ is a classical-quantum state between $\tilde{a}_0$ and $\tilde{a}_R C'$, we obtain

$$I(\tilde{a}_0 \tilde{a}_R : \tilde{C'})_{\psi}$$

$$= I(\tilde{a}_0 : C')_{\psi} + I(\tilde{a}_R : C'|\tilde{a}_0)_{\psi}$$

$$\leq S(\tilde{a}_0)_{\psi} + 2S(\tilde{a}_R)_{\psi}$$

$$\leq S(\tilde{a}_0)_{\chi,\mathcal{K}^A} + 2S(\tilde{a}_R)_{\chi,\mathcal{K}^A} + 7n\eta(\epsilon) \log d_A.$$ (89)

where the last equality follows from (86).

It is straightforward to obtain from (77) that

$$I(A : C')_{\psi} = I(a_0 a_R : C'|c_{\psi})_{\psi}$$

$$= H(|\{p_j\}|) + 2 \sum_j p_j S(\phi^{a_j}_{\psi})$$

(89)
Combining (83), (88) (89), we obtain
\[
n \left( H\{p_j\}_j + 2 \sum_j p_j S(\varphi_j^{\alpha_j}) \right) \leq H\{q_i\}_i + 2 \sum_i q_i S(\phi_i^{\alpha_i}) + n \zeta'_\epsilon(\epsilon) \log d, \]
where
\[
\zeta'_\epsilon(\epsilon) = 11\eta(2\sqrt{\epsilon}) + 4\eta(\zeta(2\epsilon)) \tag{90}
\]
and \(\zeta_\epsilon\) is a function defined by (78). Thus we finally arrive at (23).

E. Convergence of \(\zeta_\epsilon\)

We prove that \(\zeta_\epsilon(\epsilon)\) defined by (78) satisfies \(\lim_{\epsilon \to 0} \zeta_\epsilon(\epsilon) = 0\), based on an idea used in [25]. Due to the Choi-Jamiolkowski isomorphism, \(\mathcal{E}\) can be identified with \(S(\mathcal{H}^A \otimes \mathcal{H}^A)\). Hence \(\mathcal{E}\) is compact, which implies that the supremum in (78) can actually be the maximum:
\[
\zeta_\epsilon(\epsilon) = \max_{\mathcal{E} \in \mathcal{E}} \{ g(\mathcal{E}) | f(\mathcal{E}) \leq \epsilon \}.
\]
Hence we have that
\[
\forall \epsilon > 0, \exists \mathcal{E} \in \mathcal{E} : g(\mathcal{E}) = \zeta_\epsilon(\epsilon), f(\mathcal{E}) \leq \epsilon.
\]
Define \(\alpha := \lim_{\epsilon \to 0} \zeta_\epsilon(\epsilon)\). Due to the monotonicity, we have \(\zeta_\epsilon(\epsilon) \geq \alpha\) for all \(\epsilon > 0\). Consequently, we have that
\[
\forall \epsilon > 0, \exists \mathcal{E} \in \mathcal{E} : g(\mathcal{E}) \geq \alpha, f(\mathcal{E}) \leq \epsilon. \tag{91}
\]
Define \(\mathcal{E}_\alpha := \{ \mathcal{E} | g(\mathcal{E}) \geq \alpha \}\). Due to the continuity of \(g\), \(\mathcal{E}_\alpha\) is a closed subset of \(\mathcal{E}\). Hence
\[
\beta := \min_{\mathcal{E} \in \mathcal{E}_\alpha} f(\mathcal{E})
\]
exists due to the continuity of \(f\). By definition, we have that
\[
\forall \mathcal{E} \in \mathcal{E} : g(\mathcal{E}) \geq \alpha \Rightarrow f(\mathcal{E}) \geq \beta. \tag{92}
\]
Suppose now that \(\alpha > 0\). We have \(f(\mathcal{E}) > 0\) for all \(\mathcal{E} \in \mathcal{E}_\alpha\) due to Lemma [5]. Thus we have \(\beta > 0\), in which case \(\beta = \epsilon\) contradicts with (91) because \(\epsilon\) can be arbitrarily small.

APPENDIX C
PROOF OF THEOREM 10

In this Appendix, we prove Theorem [10] based on irreducibility of the KI decomposition.

A. Irreducibility of the KI decomposition

Similarly to the irreducibility of the KI decomposition of a set of states presented in Lemma 2, the KI decomposition of a bipartite state defined by Definition 3 also has a property of irreducibility as follows.

Lemma 15 Suppose the KI decomposition of \(\Psi^{AC}\) on \(A\) is given by
\[
\Psi^{AC}_{KI} = \sum_{j \in J} p_j |j\rangle^a |\omega_j^{a}\rangle \otimes \varphi_j^{aC},
\]
and define \(\varphi_{j,k,l} := (|k\rangle^C \varphi_j^{aC} |l\rangle^C\), where \(|k\rangle\) is an orthonormal basis of \(\mathcal{H}^C\). Then the following two properties hold.

1) If a linear operator \(N\) on \(\mathcal{H}_p^{a}\) satisfies \(p_j N \varphi_{j,k,l} = p_j \varphi_{j,k,l} N\) for all \(k\) and \(l\), then \(N = cI_{F}^{a}\) for a complex number \(c\), where \(I_{F}^{a}\) is the identity operator on \(\mathcal{H}_p^{a}\).

2) If a linear operator \(N : \mathcal{H}_p^{a} \rightarrow \mathcal{H}_p^{a}\) \((j \neq j')\) satisfies \(p_j N \varphi_{j,k,l} = p_j \varphi_{j',k,l} N\) for all \(k\) and \(l\), then \(N = 0\).

Proof: Define
\[
p_{M} := \text{Tr}[M^{C} \Psi^{AC} M^{IC}],
\]
\[
\Psi_{M}^{a} := p_{M}^{-1} \text{Tr}_{C}[M^{C} \Psi^{AC} M^{IC}],
\]
\[
\varphi_{j,M}^{a} := p_{M}^{-1} \text{Tr}_{C}[M^{C} \varphi_{j,K}^{aC} M^{IC}]
\]
for \(M \in \mathcal{L}(\mathcal{H}^C)\). The set of steerable states corresponding to \(\Psi\) is given by \(\mathcal{E}_{\Psi^{aC} \rightarrow a} := \{ \Psi_{M} | M \in \mathcal{L}(\mathcal{H}^C)\} \). Hence the KI isometry \(\Gamma\) on \(A\) with respect to \(\Psi^{AC}\) is equal to that with respect to \(\mathcal{E}_{\Psi^{aC} \rightarrow a}\) by Definition 3. Thus \(\Psi_{M}^{a}\) is decomposed as
\[
\Gamma \Psi_{M}^{a} \Gamma^{\dagger} = \sum_{j \in J} p_{j} |j\rangle^{a} \otimes \omega_{j}^{a} \otimes \varphi_{j,M}^{aC},
\]
where \(\{ \varphi_{j,M}^{aC} \}_{M \in \mathcal{L}(\mathcal{H}^C)}\) is a set of states which is irreducible in the sense of Lemma 2.

To prove Property 1), suppose that \(N \in \mathcal{L}(\mathcal{H}_p^{a})\) satisfies \(p_j N \varphi_{j,k,l} = p_j \varphi_{j,k,l} N\) for all \(k\) and \(l\). Since \(\varphi_{j,M}^{aC}\) is decomposed as \(\varphi_{j,M}^{aC} = \sum_{k,l} (|k\rangle^{C} M^{C} |k\rangle_{C} \varphi_{j,k,l})\), it follows that \(p_j N \varphi_{j,k,l} = p_j \varphi_{j,k,l} N\) for all \(M \in \mathcal{L}(\mathcal{H}^C)\). Hence we obtain Property 1) due to the irreducibility of \(\{ \varphi_{j,M}^{aC} \}_{M \in \mathcal{L}(\mathcal{H}^C)}\). Property 2) is proved in a similar vein.

B. Proof of Theorem 10

Let us first adduce a useful lemma regarding fixed points of the adjoint map of a linear CPTP map.

Lemma 16 (See Lemma 11 in [11].) Let \(\mathcal{E}\) be a linear CPTP map on \(S(\mathcal{H})\), the Kraus representation of which is given by \(\mathcal{E}(\cdot) = \sum_k E_k(\cdot) E_k^{\dagger}\). Let \(\mathcal{E}^{*}\) be the adjoint map of \(\mathcal{E}\) defined by \(\mathcal{E}^{*}(\cdot) = \sum_k E_k^{\dagger}(\cdot) E_k\). Then \(X \in \mathcal{L}(\mathcal{H})\) satisfies \(\mathcal{E}^{*}(X) = X\) if and only if \([E_k, X] = [E_k^{\dagger}, X] = 0\) for all \(k\).

The proof of Theorem 10 proceeds as follows. Let \(\Gamma\) be the KI isometry on \(A\) with respect to \(\Psi^{AC}\), and let
\[
\hat{\Psi}^{AC} := \Gamma^{A} \Psi^{AC} \Gamma^{\dagger A} = \sum_{j \in J} p_j |j\rangle^{a} \otimes \omega_{j}^{a} \otimes \varphi_{j,K}^{aC}.
\]
be the KI decomposition of \(\Psi^{AC}\) on \(A\). We have
\[
(\hat{\Psi}^{AC})^{\frac{1}{2}} = \sum_{j \in J} \sqrt{p_j} |j\rangle^{a} \otimes (\omega_{j}^{a})^{\frac{1}{2}} \otimes (\varphi_{j,K}^{aC})^{\frac{1}{2}},
\]
\[
(\hat{\Psi}^{A})^{-\frac{1}{2}} = \sum_{j \in J} \frac{1}{\sqrt{p_j}} |j\rangle^{a} \otimes (\omega_{j}^{a})^{-\frac{1}{2}} \otimes (\varphi_{j,K}^{aC})^{-\frac{1}{2}},
\]
and
\[
(\hat{\Psi}^{A})^{\frac{1}{2}} (\hat{\Psi}^{A})^{-\frac{1}{2}} = \sum_{j \in J} |j\rangle^{a} \otimes I_{j}^{a} \otimes (\varphi_{j,K}^{aC})^{\frac{1}{2}} (\varphi_{j,K}^{aC})^{-\frac{1}{2}}.
\]
Therefore, from (93), we obtain

\[
\hat{E}_{\infty,j'j}(|\varphi_j\rangle^a|\varphi_{j'}\rangle^{a_{bc}}) = \mathcal{E}_{\infty,j'j}(|\varphi_j\rangle^a|\varphi_{j'}\rangle^{a_{bc}}) e^\dagger_{j',kl} (\forall j,j',k,l).
\]

Hence we have

\[
\mathcal{E}_{\infty,j'j}(|\varphi_j\rangle^a|\varphi_{j'}\rangle^{a_{bc}}) = \begin{cases} 
\pi^a_n \otimes \varphi_j^{b_{C}} & (j = j') \\
0 & (j \neq j') 
\end{cases}
\]

due to the irreducibility of \{e_{j,kl}\}_{k,l}. From (93), we obtain

\[
\hat{E}_{\infty}(|\Psi_{KL}\rangle) = \sum_{j \in J} p_j j, j, \omega_j (j, j, \omega_j |a_{bc}a_{bc}) \otimes \pi^a_j \otimes \varphi_j^{b_{C}}.
\]

Since \(\hat{E}_{\infty}(|\Psi_{KL}\rangle)\) and \(\Psi_{ABC}\) are equivalent up to local isometries on A and B, we finally obtain

\[
S(\Psi_{ABC}) = H(|\Psi_{KL}\rangle) + \sum_j p_j \left( S(\pi^a_n) + S(\varphi_j^{b_{C}}) \right)
\]

where we used the fact that \(\varphi_j^{a} = \pi^a_j\) and that \(\varphi_j^{a_{bc}}\) is a pure state. ☐

Remark: From (13) and (14), it is straightforward to verify that the statement of Theorem 10 does not depend on a particular choice of a purification of \(\Psi_A\). That is, if for any \(\Omega^{AA'}\) of \(\Psi_A\), we have \(M_{AB}(\Psi^{ABC}) = S(\Omega^{AA'})\), where \(\Omega^{AA'} := \mathcal{E}_{\infty}(|\Omega\rangle|\Omega\rangle|AA'\rangle\). A purification of \(\Psi_A\) is simply obtained by

\[
|\Omega\rangle^{AA'} = (\Psi_A)^{\frac{1}{2}} \sum_{k=1}^{d_A} |k\rangle^A |k\rangle^{A'}
\]

and its matrix representation is given by

\[
|\Omega\rangle^{AA'} = \sum_{klmn} [\Lambda_{kl, mn}] |k\rangle^A \otimes |m\rangle|n\rangle^{A'}.
\]

The matrix elements of the generalized inverse matrix \(\Lambda^{-1}\) are given by

\[
[\Lambda_{kl, mn}] = (\langle k | (\Psi_A)^{-\frac{1}{2}} | m\rangle) (|n\rangle (\Psi_A)^{-\frac{1}{2}} | l\rangle).
\]

In addition, the dimension of the eigensubspace of \(\Lambda_{\infty}\) corresponding to the eigenvalue 1 is at least 1, since we have \(\mathcal{E}_{\infty}(I) = I\) due to the self-adjointness of \(\mathcal{E}\). These facts justify the algorithm described in Section III.

APPENDIX D

PROOF OF INEQUALITY (33)

The first inequality in (33) is proved as follows. For an arbitrary \(n\) and \(\epsilon > 0\), let \(V_n : \tau \mapsto 2^{-nR} \sum_{k=0}^{a^nR} V_k \tau V_k^\dagger\) be a random unitary operation on \(A^n\), and let \(\Gamma^{AB^nC^n}\) be a Markov state conditioned by \(B^n\) such that

\[
\left\| V_n(\rho^{\otimes n}) - Y^{AB^nC^n} \right\|_1 \leq \epsilon.
\]

(94)
Let $|\psi\rangle^{ABCD}$ be a purification of $\rho^{ABC}$, and $E$ be a quantum system with dimension $2^{nR}$. Defining an isometry $W : A^n \rightarrow EA^n$ by $W = \sum_{k=1}^{2^{nR}} |k\rangle^E \otimes V_k^A^n$, a Stinespring dilation of $V_n$ is given by $V_n(\tau) = Tr_E[WLW^\dagger]$. Then a purification of $\rho_n^{ABC} := V_n(\rho^{\otimes n})$ is given by $|\psi'_n\rangle^{EA^nB^nC^nR^n} := W(|\psi\rangle^{ABCR^R})^\otimes n$. For this state, we have

$$nR \geq S(E)_{\psi_n'} = S(A^nB^nC^nR^n)_{\psi'_n} \geq S(A^nB^nC^n)_{\psi'_n} - S(R^n)_{\psi'_n} = S(A^nB^nC^n)_{\rho'_n} - nS(ABC)_{\rho_n}, \quad (95)$$

where the second line follows from (44). From (94), we also have

$$S(A^nB^nC^n)_{\rho'_n} \geq S(A^nB^nC^n)_T - n\eta(\epsilon) \log (d_A d_B d_C)$$

$$= S(A^nB^n)_T + S(B^nC^n)_T - S(B^n)_T - n\eta(\epsilon) \log (d_A d_B d_C)$$

$$\geq S(A^nB^n)_{\rho'_n} + S(B^nC^n)_{\rho'_n} - S(B^n)_{\rho'_n} - 4n\eta(\epsilon) \log (d_A d_B d_C)$$

$$\geq S(A^nB^n)_{\rho_{\otimes n}} + S(B^nC^n)_{\rho_{\otimes n}} - S(B^n)_{\rho_{\otimes n}} - 4n\eta(\epsilon) \log (d_A d_B d_C)$$

$$= n(S(AB)_{\rho} + S(BC)_{\rho} - S(B)_{\rho}) - 4n\eta(\epsilon) \log (d_A d_B d_C). \quad (96)$$

Here, the second line follows by Inequality (47); the third line because of $\Upsilon$ being a Markov state conditioned by $B^n$; the fourth line by Inequality (47); and the fifth line by the von Neumann entropy being nondecreasing under random unitary operations, in addition to $\rho_n^{B^nC^n} = (\rho^{BC})_{\otimes n}$. From (95) and (96), we obtain

$$R \geq I(A:C|B)_{\rho} - 4\eta(\epsilon) \log (d_A d_B d_C),$$

which concludes the proof. \[\square\]