A Coding Theorem for Bipartite Unitaries in Distributed Quantum Computation

Eyuri Wakakuwa, Akihito Soeda and Mio Murao

Abstract—We analyze implementations of bipartite unitaries by means of local operations and classical communication (LOCC) assisted by shared entanglement. We employ concepts and techniques developed in quantum Shannon theory to study an asymptotic scenario, in which two distant parties perform the same bipartite unitary on infinitely many pairs of inputs. We analyze minimum cost of entanglement and classical communication per copy. For two-round LOCC protocols, we derive a single-letter formula for the minimum cost of entanglement and classical communication, under an additional requirement that the error converges to zero faster than $1/n^4$, where $n$ is the number of input pairs. The formula is given by the “Markovianizing cost” of a tripartite state associated with the unitary, which can be computed by a finite-step algorithm. We also derive a lower bound on the minimum cost of resources, which applies for protocols with arbitrary number of rounds.

Index Terms—Entanglement cost, classical communication cost, quantum Shannon theory, approximate recoverability.

I. INTRODUCTION

Distributed quantum computation is a task in which a group of distant parties collaborates to perform a large quantum computation, by using classical communication, quantum communication and shared entanglement as resources. One of the most extensively investigated tasks is implementation of bipartite unitaries by local operations and classical communication (LOCC) assisted by shared entanglement. Here, two distant parties, say Alice and Bob, have quantum systems $A$ and $B$ in an unknown state $|\psi\rangle^{AB}$, and aim to perform a known unitary $U^{AB}$ by LOCC, using some resource entanglement shared in advance. Although this task can be implemented simply by using quantum teleportation, it was shown that the cost of entanglement and classical communication can be reduced by constructing a more efficient protocol, depending on the unitary to be implemented [1].

The following two questions then naturally arise: (i) How can we find efficient protocols which consume less resources for a given bipartite unitary? and (ii) What are the minimum costs of resources required for implementing that unitary? Although these questions have been addressed, e.g. in [1]–[11], most of the studies so far assume particular forms of the resource entanglement or of the bipartite unitary to be implemented. A general method to address these problems is yet discovered.

In the present paper, we address the above questions in an information theoretical scenario for the first time, by applying the concept of “block coding”. Here, the two parties perform the same bipartite unitary on a sequence of input pairs at once. We consider an asymptotic limit of infinite pairs and vanishingly small error, and analyze the minimum cost of entanglement and classical communication per copy required for the task. We mainly focus on protocols consisting of two-round LOCC as the first nontrivial case, for simplicity. Our approach is different from previous ones which treated single-shot cases [1]–[11].

The main result of this paper is that we derive a single-letter formula for the minimum cost of entanglement, forward and backward classical communication in two-round protocols, under an additional assumption that the error converges to zero faster than $1/n^4$, where $n$ is the number of input pairs. The formula is represented in terms of the “Markovianizing cost” ([12], [13]) of a state associated with the unitary, which can be computed by a finite-step algorithm. The result is applicable for any bipartite unitary.

It is left open, however, whether the same converge bound holds when we drop the requirement on the convergence speed of the error. We relate this problem to another open problem regarding an “asymptotic symmetry” of approximate recoverability, that is, whether a tripartite quantum state $\rho^{ABC}$ is approximately recoverable from $\rho^{AB}$ if it is approximately recoverable from $\rho^{AB}$, up to a dimension-independent rescaling of error of recovery. We prove that an affirmative answer to the asymptotic symmetry implies an affirmative one to the converse bound.

We also derive a lower bound on the minimum cost of entanglement and classical communication, which is applicable for any protocol with arbitrary number of rounds, in terms of a parameter called the Schmidt strength of the unitary. It turns out that the lower bound is achievable for a class of bipartite unitaries called generalized Clifford operators.

The structure of this paper is as follows. In Section [1] we introduce the formal definition of the problem. The results are summarized in Section [11]. In Section [V] we review results on Markovianization and state merging. Section [V] analyzes single-shot two-round protocols for implementing.
a bipartite unitary. Outlines of the proofs of the main result are presented in Section [VI] In Section [VII] we discuss general properties of the Markovianizing cost of unitaries. The Markovianizing cost for two classes of bipartite unitaries is computed in Section [VIII] as examples. In Section [IX] we investigate an open problem regarding the convergence speed of the error from a viewpoint of approximate recoverability. Section [X] analyzes the power of a LOCC protocol for transmitting classical information. In Section [XI] we provide a lower bound on the cost of resources for an arbitrary LOCC protocol. Conclusions are given in Section [XII] The detailed proofs of lemmas and theorems in the main part are presented in Appendices.

Notations. $|\phi_d\rangle$, $|\Phi_{K_n}\rangle$, and $|\Phi_{L_n}\rangle$ represent the maximally entangled state with the Schmidt rank $d$, $K_n$ and $L_n$, respectively. $\pi_d$ is the maximally mixed state of rank $d$. The fidelity and the trace distance between two quantum states $\rho$ and $\sigma$ are denoted by $F(\rho, \sigma)$ and $\|\rho - \sigma\|_1$, respectively. We abbreviate $F(\rho, |\psi\rangle\langle\psi|)$ as $F(\rho, |\psi\rangle)$. For a quantum operation $E$, we abbreviate $E(|\psi\rangle\langle\psi|)$ as $E(|\psi\rangle)$. Otherwise we follow the notations introduced in [12].

II. Formulations

Suppose that Alice and Bob are given an unknown pure quantum state $\psi$ on a composite system $A^n B^n$, which is paired off as $A_1 B_1 \cdots A_n B_n$. We consider a task in which they perform the same bipartite unitary $U^{AB}$ on each pair of systems $A_1 B_1, \ldots, A_n B_n$, or equivalently, perform $U^{AB}$ on $A^n B^n$. For accomplishing this task, Alice and Bob apply a quantum operation consisting of local operations and classical communication (LOCC), in which case they need to consume an entangled state shared in advance as a resource. It is required that the error vanishes in the limit of $n$ to infinity. Our interest is to find the minimum cost of entanglement, forward and backward classical communication per copy for accomplishing this task, in an asymptotic limit of $n$ to infinity. Following the notations of [12], we denote composite systems $A^n B^n$ by $\bar{A}$ and $\bar{B}$, respectively.

We assume that the entangled state shared in advance is in the form of the maximally entangled state $\Phi_{K_n, A_0 B_0}$. In general, an entangled state $\Phi_{L_n}^{A_1 B_1}$ may be retrieved at the end of the operation. Thus we define a protocol for accomplishing this task by a triplet of $\Phi_{K_n, A_0 B_0}$ and a quantum operation $M_n$ that consists of local operations and classical communication. The entanglement cost of a protocol is quantified by the number of copies of Bell pairs in the state $\Phi_{K_n, A_0 B_0}$, subtracted by that in $\Phi_{L_n}^{A_1 B_1}$ (i.e., $\log K_n - \log L_n$).

For evaluating the error of a protocol for this task, we adopt the average fidelity over all pure input state $|\psi\rangle$ on $\bar{A}\bar{B}$. That is, defining

$$\rho(M_n, \psi) := M_n(|\psi\rangle\langle\psi|) \Phi_{K_n, A_0 B_0},$$

we will use the following function to evaluate the error of $M_n$:

$$F_{av}(M_n, K_n, L_n; U^{\otimes n}) := \int_{\text{Haar}} p(d\psi) F(\rho(M_n, \psi), (U^{\otimes n}|\psi\rangle) \Phi_{L_n, A_0 B_0}).$$  (1)

Here, the average is taken with respect to the Haar measure on $H^A \otimes H^B$.

Another task is one in which Alice and Bob apply $U^{AB}$ on $|\Phi_{K_n}^{AB}\rangle (U^{AB})^{\otimes n}$ by LOCC, using an entangled state $\Phi_{K_n, A_0 B_0}$ and retrieving $\Phi_{L_n}^{A_1 B_1}$ as well as the end. Here, $\bar{R}_A$ and $\bar{R}_B$ are imaginary reference systems that are inaccessible to Alice and Bob (see Figure [1]). We introduce notations

$$|\Psi_U\rangle := U^{AB} |\Phi_d^{AB}\rangle |\Phi_d^{AB}\rangle$$

and

$$\rho(M_n) := M_n(|\Phi_d^{AB}\rangle |\Phi_d^{AB}\rangle (U^{AB})^{\otimes n} |\Phi_{K_n}^{AB}\rangle A_0 B_0).$$  (2)

The error of a protocol $M_n$ is quantified by the following function:

$$F_{en}(M_n, K_n, L_n; U^{\otimes n}) := F(\rho(M_n), |\Psi_U^{\otimes n}|) \Phi_{L_n}^{A_1 B_1}).$$  (3)

These tasks are equivalent since

$$F_{av}(M_n, K_n, L_n; U^{\otimes n}) \approx 1$$  (4)

if and only if

$$F_{en}(M_n, K_n, L_n; U^{\otimes n}) \approx 1,$$  (5)

which is proven in Appendix [B] using the relation between the average fidelity and entanglement fidelity [14]. Note that, since $\bar{R}_A$ and $\bar{R}_B$ are inaccessible systems, we cannot apply the result of [13], which analyzed convertibility of bipartite pure entangled states by LOCC protocols.

Let us introduce a rigorous definition.

Definition 1 Let $U$ be a bipartite unitary acting on two $d$-dimensional quantum systems $A$ and $B$. Let Alice and Bob have quantum registers $\{A_0, A_1\}$ and $\{B_0, B_1\}$, respectively, and let $M_n$ be a quantum operation from $\bar{A}A_0 \otimes \bar{B}B_0$ to $\bar{A}A_1 \otimes \bar{B}B_1$. $M_n$ is called an $(r, n, \epsilon)$-protocol for implementing $U$ with the entanglement cost $E_n$, if $M_n$ is an $r$-round LOCC and there exist natural numbers $K_n, L_n$ such that

$$F_{en}(M_n, K_n, L_n; U^{\otimes n}) \geq 1 - \epsilon$$  (6).
and

\[ E_n := \log K_n - \log L_n. \]

Let \( r_{A \rightarrow B} \) and \( r_{B \rightarrow A} \) be numbers of times of classical communication rounds from Alice to Bob and from Bob to Alice, respectively, in an \( r \)-round protocol \( M_n \). By definition, we have \( r = r_{A \rightarrow B} + r_{B \rightarrow A} \). We denote by \( C_f(\gamma) \) the bit length of the classical message transmitted in the \( \gamma \)-th communication from Alice to Bob, and by \( C_b(\gamma) \) the one from Bob to Alice. The forward classical communication cost \( C_{f,n} \) of \( M_n \) is defined as the sum of numbers of classical bits transmitted from Alice to Bob in \( M_n \), that is,

\[ C_{f,n} := \sum_{\gamma=1}^{r_{A \rightarrow B}} C_f(\gamma). \]

In the same way, the backward classical communication cost \( C_{b,n} \) of \( M_n \) is defined as

\[ C_{b,n} := \sum_{\gamma=1}^{r_{B \rightarrow A}} C_b(\gamma). \]

As mentioned above, our interest is to find the minimum cost of entanglement, forward and backward classical communication per copy for accomplishing this task, in an asymptotic limit of \( n \to \infty \) and \( \epsilon \to 0 \). This leads us to define the achievability of the cost of resources.

**Definition 2** A rate triplet \((E, C_f, C_b)\) is said to be achievable by an \( r \)-round protocol if there exists a sequence of \((r, n, \epsilon_n)\)-protocols for implementing \( U \) (\( n = 1, 2, \cdots \)), with the entanglement cost \( E_n = n E \), forward classical communication cost \( C_{f,n} = n C_f \) and backward classical communication cost \( C_{b,n} = n C_b \) for each \( n \), such that \( \lim_{n \to \infty} \epsilon_n = 0 \).

### III. RESULTS

The results of this paper are summarized as follows.

**A. Result 1: Achievable Rate Region for Two-Round Protocols**

In this paper, we mainly consider two-round protocols (i.e. \( r = 2 \)) starting with Alice’s operation. In general, such a protocol proceeds as follows: Alice first performs a measurement and communicates the outcome to Bob; Bob then performs a measurement and communicates the outcome to Alice; and, finally, Alice performs an operation. The first main result of this paper is that the optimal rate of the cost of entanglement and of classical communication in a two-round protocol are given by a quantity called the Markovianizing cost of the unitary, under an additional requirement that the error vanishes faster than \( 1/n^4 \).

A tripartite quantum state \( \Upsilon^{ABC} \) is called a Markov state conditioned by \( B \) if it satisfies \( I(A : C|B)_\Upsilon = 0 \) \[16\]. Markovianization as formulated in \[12\] is a task in which \( n \) copies of a tripartite state \( \rho^{ABC} \) is transformed by a randomizing operation on \( \hat{A} \) to a Markov state conditioned by \( B \). The Markovianizing cost of \( \rho^{ABC} \) is defined as the minimum cost of randomness per copy required for the task, in an asymptotic limit of infinite copies and vanishingly small error. A rigorous definition is as follows.

**Definition 3** 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 \( V_n : \tau \mapsto 2^{-nR} \sum_{k=1}^{2^n} V_k \tau V_k^\dagger \) on \( \hat{A} \) and a Markov state \( \Upsilon^{ABC} \) conditioned by \( B \) that satisfy

\[ \left\| V_n^{\hat{A}}(\rho^{\otimes n}) - \Upsilon^{ABC} \right\|_1 \leq \epsilon. \]

The Markovianizing cost of \( \rho^{ABC} \) is defined as \( M_{A|B}(\rho^{ABC}) := \inf\{R \mid \rho^{ABC} \text{ is Markovianized with the randomness cost } R \text{ on } A, \text{ conditioned by } B\} \).

We extend the notion of Markovianizing cost to a bipartite unitary as follows.

**Definition 4** Let \( U \) be a bipartite unitary acting on two \( d \)-level systems \( A \) and \( B \), and consider a “triptite” state

\[ |\Psi_U\rangle^{AR_A(BRB)} := (U^{AB} \otimes I^{RA_B}) |\Phi_d\rangle^{AR_A} |\Phi_d\rangle^{BR_B} \]

by regarding \( B \) and \( RB \) as a single system. The Markovianizing cost of \( U \) is defined as \( M(U) := M_{A|R_A}(\Psi_{U}^{AR_A(BRB)}) \).

The main result of this paper is presented by the following theorem. The proofs are given in Section [VI] and the corresponding appendices, after preparatory arguments in Section [IV] and [V].

**Theorem 5**

- **Direct**: A rate triplet \((E, C_f, C_b)\) is achievable by a two-round protocol for implementing \( U \) if \( E, C_f, C_b \geq M(U^\dagger) \).

- **Converse**: A rate triplet \((E, C_f, C_b)\) is achievable by a two-round protocol for implementing \( U \) only if \( E, C_f, C_b \geq M(U^\dagger) \), if we additionally require in **Definition 2** that

\[ \lim_{n \to \infty} n^4 \cdot \epsilon_n = 0. \]

It is left open whether the same converse bound holds when we remove Condition [9]. As we will discuss in Section [IX] in detail, this question is directly related to another question of whether Equality \[17\] holds without Condition [9]. At the core of these questions lies an open problem regarding an “asymptotic symmetry” of approximate recoverability.

**B. Result 2: General Lower Bound on the Cost of Resources**

It is proven in \[17\], \[18\] that any bipartite unitary \( U \) on \( AB \) is decomposed as

\[ U^{AB} = \sum_{s=0}^{d^2-1} c_s E_s^A \otimes F_s^B, \]
where \( c_s (s = 0, \ldots, d^2 - 1) \) are nonnegative real numbers that satisfy
\[
\sum_{s=0}^{d^2-1} c_s^2 = 1, \quad c_s \geq 0 \quad (\forall s),
\]
and \( E_s \in \mathcal{L}(\mathcal{H}^A), \quad F_s \in \mathcal{L}(\mathcal{H}^B) \) are linear operators which are orthonormal with respect to the Hilbert-Schmidt inner product, i.e.,
\[
\frac{1}{d} \text{Tr}[E_s^d F_{s'}^d] = \frac{1}{d} \text{Tr}[F_s^d F_{s'}^d] = \delta_{s s'}.
\]
The Shannon entropy of \( \{c_s^2\}_s \) is called the Schmidt strength of \( U \). We denote it by \( K(U) \), that is,
\[
K(U) := H(\{c_s^2\}_s) = - \sum_s c_s^2 \log c_s^2.
\]

The following theorem states that a lower bound on the minimum cost of entanglement and classical communication, in a protocol with arbitrary number of rounds of communication, is given by the Schmidt strength of the unitary. Proofs are given in Section XI and the corresponding appendices.

**Theorem 6** A rate triplet \((E, C_f, C_b)\) is achievable only if \( E, C_f, C_b \geq K(U) \).

**Remark.** Instead of the average fidelity (1) and the entanglement fidelity (3), one could use the worst-case fidelity as a figure of merit of error. Let \( \psi' \) be a pure state on a system \( AB \), with \( B \) being a reference system, and define
\[
\rho(M_n, \psi') := M_n(\ketbra{\psi'}{\bar{A}B} | \Phi_{K_n}^{AB}) \rho_0^{AB}. \]

The worst-case fidelity is defined as
\[
F_{w.c.}(M_n, K_n, L_n; \rho^{AB}) := \inf \inf F(\rho(M_n, \psi'), (U^{\otimes n})^{\bar{A}B} | \Phi_{L_n}^{AB}) \rho_0^{AB},
\]
where the infimum is taken over all reference systems \( R \) and all pure states \( \psi' \) on \( \bar{A}B \). Since we have \( F_{w.c.} \leq F_{w.e.} \), the converse bounds in Theorem 5 and 6 remain to hold under this choice of the fidelity. However, we do not know whether the direct part in Theorem 5 holds as well. It is proven in [19] and [20] that the “superreplication” of an unknown unitary gate is possible with vanishingly small average error, while it is not possible with vanishingly small worst-case error. In analogy to these results, it could be natural to expect that there is a gap between the minimum cost of resources for implementing a bipartite unitary with vanishing average error and the one for implementing it with vanishing worst-case error.

**IV. Preliminaries**

We review an alternative definition of the Markovianizing cost [13], in addition to state merging [21], [22]. The results reviewed here are used in the following sections to prove Theorem 5.

Fig. 2. A graphical representation of Markovianization of a pure state by a measurement with an auxiliary entangled resource. After the measurement, the reduced state on \( A'BC \) should be an approximately recoverable state.

**A. Markovianization in terms of Recoverability**

It is proved in [16] that the following conditions are equivalent:

1) Vanishing QCMI: \( \rho_{ABC} \) is a Markov state conditioned by \( B \), i.e., it satisfies
\[
I(A : C | B)_\rho = 0.
\]

2) Recoverability: \( \rho_{ABC} \) is recoverable from its bipartite reduced state on \( AB \) and \( BC \), that is, there exist quantum operations \( \mathcal{R} : B \rightarrow AB \) and \( \mathcal{R'} : B \rightarrow BC \) such that
\[
\rho_{ABC} = \mathcal{R}(\rho_{BC}) = \mathcal{R}'(\rho_{AB}).
\]

Based on this fact, the Markovianizing cost in terms of recoverability is introduced in [13]. In the same way as Definition 5 we consider a task in which \( n \) copies of a tripartite quantum state \( \rho_{ABC} \) is transformed by a random unitary operation on \( A \). Instead of requiring that the state after the operation satisfies Condition 7, however, we now require that the state satisfies Condition (13) up to a small error \( \epsilon \). Rigorous definitions are as follows.

**Definition 7** A tripartite state \( \rho_{ABC} \) is said to be \( \epsilon \)-recoverable from \( BC \) if there exists a quantum operation \( \mathcal{R} : B \rightarrow AB \) such that
\[
\| \rho_{ABC} - \mathcal{R}(\rho_{BC}) \|_1 \leq \epsilon.
\]

\( \rho_{ABC} \) is \( \epsilon \)-recoverable from \( AB \) if there exists a quantum operation \( \mathcal{R}' : B \rightarrow BC \) such that
\[
\| \rho_{ABC} - \mathcal{R}'(\rho_{AB}) \|_1 \leq \epsilon.
\]

**Definition 8** A tripartite state \( \rho_{ABC} \) is Markovianized with the randomness cost \( R \) on \( A \), in terms of recoverability from \( BC \), if the following statement holds. That is, there exists a sequence of sets of unitaries \( \{V_{n,k}\}_{k=1}^{2^n R} \) acting on \( (\mathcal{H}^A)^{\otimes n} \), such that \( V_{n,k}((\rho_{ABC})^{\otimes n}) = \epsilon_n \)-recoverable from \( BC \) for
\[
V_n : \tau \mapsto 2^{-n R} \sum_{k=1}^{2^n R} V_{k} \tau V_{k}^\dagger
\]
and \( \lim_{n \to \infty} \epsilon_n = 0 \).

This allows us to define the Markovianizing cost of \( \rho^{ABC} \) in terms of recoverability from BC as
\[
M_{AB|BC}^R(\rho^{ABC}) := \inf \{ R \mid \rho^{ABC} \text{ is Markovianized with the randomness cost } R \text{ on } A, \text{ in terms of recoverability from } BC \}. 
\]

We also consider a Markovianization induced by a measurement, supplemented by auxiliary entanglement resource (Figure 2).

**Definition 9** Consider a tripartite pure state \( |\Psi\rangle^{ABC} \), and let \( A_0 \) and \( G \) be additional quantum systems. A pair of a pure state \( |\psi\rangle^{A_0G} \) and a measurement on \( A_0 \), which is described by a set of measurement operators \( \{ M_k^{A_0 \rightarrow A'} \}_{k \in \mathbb{K}} \), is called an \((n, R, \epsilon)\)-Markovianization pair for a state \( |\Psi\rangle^{ABC} \) if it satisfies the following conditions:

1) The measurement does not significantly change the reduced state on \( BC \) on average, i.e.,
\[
\sum_{k \in \mathbb{K}} p_k \left\| (\psi^{\otimes n})^{BC} - \psi_k^{BC} \right\|_1 \leq \epsilon, \tag{14}\]
where \( p_k \) is the probability of obtaining the outcome \( k \), and \( \Psi_k \) is the post-measurement state corresponding to the outcome \( k \).

2) The post-measurement state is approximately recoverable on average, that is, there exist linear CPTP maps \( R_k : \overline{B} \rightarrow \overline{BC} \) satisfying
\[
\sum_{k \in \mathbb{K}} p_k \left\| \psi'_k^{BC} - R_k(\psi_k^{B}) \right\|_1 \leq \epsilon. \tag{15}\]

3) The correlation between \( \overline{BC} \) and \( G \) produced by the measurement is at most \( nR \) bits in QMI, that is,
\[
I(\overline{BC} : G)_{av} := \sum_{k \in \mathbb{K}} p_k I(\overline{BC} : G)_{\psi_k} \leq nR. \tag{16}\]

A state \( |\Psi\rangle^{ABC} \) is said to be Markovianized with the correlation production \( R \) by a measurement on \( A \), in terms of recoverability from \( AB \), if there exists a sequence of \((n, R, \epsilon_n)\)-Markovianization pairs \((n = 1, 2, \cdots)\) such that \( \lim_{n \to \infty} \epsilon_n = 0 \).

Correspondingly, the measurement-induced Markovianizing cost of \( |\Psi\rangle^{ABC} \) in terms of recoverability from \( AB \) is defined as
\[
M_{A|BC}^R(\Psi^{ABC}) := \inf \{ R \mid |\Psi\rangle^{ABC} \text{ is Markovianized with the correlation production } R \text{ by a measurement on } A, \text{ in terms of recoverability from } AB \}. 
\]

The following lemma relates the Markovianizing cost to other entropic quantities characterizing the state transformation induced by a Markovianizing measurement.

**Lemma 11** (See Appendix C-A for a proof.) There exists a nonnegative function \( \xi(x) \) such that \( \lim_{x \to 0} \xi(x) = 0 \), and the following inequalities hold for any \( n \in \mathbb{N}, \epsilon > 0 \) and \( \{ M_k^{A_0 \rightarrow A'} \}_{k} \) satisfying 16\( \epsilon < n \leq 1/4\epsilon \) and Inequalities (14) and (15):

\[
H(\{p_k\}_{k \in \mathbb{K}}) \geq \Delta S(A')_{av} \geq \Delta S(G)_{av} \geq nM_{A|B}(\Psi^{ABC}) - n\xi(ne). \tag{19}\]

Here, we defined
\[
\Delta S(A')_{av} := nS(A) + S(A_0)_{\phi} - \sum_{k \in \mathbb{K}} p_k S(A')_{\psi_k}, \quad \Delta S(G)_{av} := S(G)_{\phi} - \sum_{k \in \mathbb{K}} p_k S(G)_{\psi_k}. \]

**Remark.** It has been left open whether Equality (17) holds when we remove Condition (18). See Proposition 13 (unproven) of [13].

**B. State Merging**

Suppose Alice and Bob share a tripartite pure state \( |\Psi\rangle^{ABR} \) with an inaccessible reference system \( R \). State merging ([21]) is a task in which Bob sends his share of \( |\Psi\rangle^{ABR} \) to Alice. \( R \) is an inaccessible reference system. For the sake of presentation, we consider Bob as the sender and Alice as the receiver.

The following lemma relates the Markovianizing cost to other entropic quantities characterizing the state transformation induced by a Markovianizing measurement.

**Definition 12** Consider a tripartite pure state \( |\Psi\rangle^{ABR} \). Let Alice and Bob have quantum systems \( \{ A_0, A_1 \} \) and \( \{ B_0, B_1 \} \), respectively, where \( A_0 \) is assumed to be identical to \( B \). The following protocol \( \mathcal{N} \) consisting of a sequence of quantum operations is called state merging of \( \Psi \) with error \( \epsilon \), entanglement cost \( \log K - \log L \) and classical communication cost \( C \). Here, \( \mathcal{N} \) : \( AA_0BB_0 \rightarrow AA_0A_1B_1 \) is a one-way LOCC from Bob to Alice, such that
\[
F(\rho(\mathcal{N}), |\Psi\rangle^{A_0R} |\Phi_L\rangle^{A_1B_1}) \geq 1 - \epsilon \tag{20}\]
for
\[ \rho(\mathcal{N}) := \mathcal{N}(|\Psi\rangle^{ABR}|\Phi^K_{L}\rangle^{A_0B_0}). \]

\( K \) and \( L \) are natural numbers, and \( \Phi^K \) and \( \Phi^L \) are the maximally entangled states with the Schmidt rank \( K \) and \( L \), respectively. \( C \) is the total amount of classical communication transmitted from Bob to Alice in \( \mathcal{M} \), measured in bits.

There always exists a state merging with an error determined by the initial state. The following theorem is obtained as a corollary of Proposition 3 and 4 in [22] by letting \( L = 1 \).

**Theorem 3** Let \( D_A := (\text{Tr}(|\Psi\rangle^A)^2)^{-1} \) and \( r_B := \text{rank}|\Psi|B\rangle \).

There exists a state merging of \( \Psi \) with entanglement cost 0, classical communication cost \( C = \log r_B \) and error
\[ \epsilon \leq 2\sqrt{2(\sqrt{d_R/D_A} + 1/r_B)^{1/2}}. \]

(21)

It is also proved in [22] that the cost of entanglement and classical communication in a state merging are bounded from below as
\[ \log K - \log L \geq nS(B|A)_{\Psi}, \quad C \geq nI(B : R)_{\Psi}, \]

(22)

when \( \epsilon \) is sufficiently small. See Appendix [CB] for details.

**V. SINGLE-SHOT TWO-ROUND PROTOCOLS**

In this section, we consider \( n = 1 \) (single-shot) case, and analyze a single-shot protocol \( \mathcal{M} \) for implementing \( U \) by two-round LOCC assisted by shared entanglement. The results obtained here are then applied to the asymptotic situation in Section [VI].

Let \( \mathcal{M} : AA_0 \otimes BB_0 \rightarrow AA_1 \otimes BB_1 \) be a two-round LOCC protocol for implementing \( U \). \( \mathcal{M} \) succeeds in implementing \( U \) with high fidelity, if
\[ F(\rho(\mathcal{M})^{AB}|\Psi_U\rangle) \geq 1 - \epsilon \]

(23)

for some small \( \epsilon \), where
\[ \rho(\mathcal{M}) := \mathcal{M}(|\Phi^d\rangle^A_R|\Phi^d\rangle^B_R|\phi^{\text{res}}_{\text{res}}\rangle^{A_0B_0}). \]

and \( \phi^{\text{res}} \) is a pure resource state shared in advance. Since we have \( \Phi^d_A \otimes \Phi^d_B = \Psi^{AB}_U \), and all purifications are equivalent up to local unitary transformations, there exists a unitary \( \hat{U} \) on \( AA_RBR \) such that
\[ |\Phi^d\rangle^A_R|\Phi^d\rangle^B_R = \hat{U}^{ARBR}|\Psi_U\rangle^{ARBR}. \]
assume that the two-round protocol $\mathcal{M}$ proceeds as follows (Figure 5):

1. Alice performs a measurement on $AA_0$, which is described by a set of measurement operators $\mathbb{M} = \{M_{kA_0}^{AA_0} \to A'\}_k$, and obtains an outcome $k$.
2. Alice communicates $k$ to Bob.
3. Bob performs a measurement on $BB_0$, described by $N_k = \{N_{kBB_0}^{BB_0} \to BB_1\}_l$, and obtains an outcome $l$.
4. Bob communicates $l$ to Alice.
5. Alice performs an operation which is described by a linear CPTP map $\mathcal{O}_{kl} : A' \to AA_1$.

Here, $A'$ is the output system of Alice’s measurement such that

$$\dim \mathcal{H}_A' \leq \dim \mathcal{H}_A \times \dim \mathcal{H}_{A_0}. \quad (26)$$

We denote the set of outcomes of Alice’s measurement by $\mathbb{K}$. (See Remark at the end of Appendix E-A for a treatment of protocols in which not all information about the measurement outcome $k$ is communicated to Bob.)

A. Conditions on Alice’s Measurement

Let us first discuss general conditions regarding state transformations by Alice’s measurement. For a fixed $\phi_{\text{res}}$, and for any linear operator $\mathcal{M} : \mathcal{H}^A \otimes \mathcal{H}_{A_0} \to \mathcal{H}_A$ such that $\mathcal{M}M \leq I$, we define a map $\mathcal{E}_M$ by

$$\mathcal{E}_M(\tau) := p_M^{-1} M(\tau^A \otimes \phi_{\text{res}}^{A_0}) M^\dagger,$$

$$p_M := \text{Tr}[M(\tau^A \otimes \phi_{\text{res}}^{A_0}) M^\dagger].$$

We call $\mathcal{E}_M$ as an $M$-induced map. $M$ is supposed to be an element of $\mathbb{M}$, in which case $p_M$ describes the probability of obtaining a certain measurement outcome corresponding to $M$. Note that $p_M$ depends on the input state $\tau$ in general. Consequently, the $M$-induced map is not necessarily a linear map. The linearity of $\mathcal{E}_M$ is equivalent to the independence of $p_M$ from $\tau$, which indicates that the measurement is oblivious to the input, in the sense that it does not extract any information about the input state. This obliviousness condition plays an important role in proofs of most of the lemmas in this section (as well as in [10]). Thus we introduce an equivalent definition of approximate obliviousness as follows.

Definition 14 An $M$-induced map is $\varsigma$-oblivious if it satisfies

$$\left\| \Phi_{M}^{RA} - \pi_d^{RA} \right\|_1 \leq \varsigma,$$

where $\Phi_{M}^{RA} := \mathcal{E}_M(\Phi_{\text{res}}^{RA})$. A measurement $\mathbb{M}$ is $\varsigma$-oblivious if an $M_k$-induced map is $\varsigma_k$-oblivious for each $k$ and $\sum_{k \in \mathbb{K}} p_k \varsigma_k \leq \varsigma$.

We introduce some other conditions on Alice’s measurement. In the following, we denote $\mathcal{E}_M(\Psi)$ as $\Psi_M$.

Definition 15 An $M$-induced map is $\mu$-decoupling between $A'R_A$ and $R_B$ if it satisfies

$$\left\| \Psi_{M}^{A'R_{A}R_{B}} - \Psi_{M}^{A'R_{A}} \otimes \pi_d^{R_B} \right\|_1 \leq \mu. \quad (27)$$

A measurement $\mathbb{M}$ is $\mu$-decoupling between $A'R_A$ and $R_B$ if an $M_k$-induced map is $\mu_k$-decoupling between $A'R_A$ and $R_B$ for each $k$ and $\sum_{k \in \mathbb{K}} p_k \mu_k \leq \mu$.

As in Definition 4 we now consider $B$ and $R_B$ as a single system and regard $\Psi_{M}^{A'R_{A}B'R_{B}}$ as a “tripartite” state on $A', R_A$ and $B'R_B$.

**Definition 16** An $M$-induced map is $\nu$-Markovianizing from $R_A B'R_B$ if $\Psi_{M}^{A'R_{A}B'R_{B}}$ is $\nu$-recoverable from $R_A B'R_B$, that is, if there exists a linear CPTP map $\mathcal{R} : R_A \to A'R_A$ such that

$$\left\| \Psi_{M}^{A'R_{A}B'R_{B}} - \mathcal{R}(\Psi_{M}^{A'R_{A}B'R_{B}}) \right\|_1 \leq \nu.$$

A measurement $\mathbb{M}$ is $\nu$-Markovianizing from $R_A B'R_B$ if an $M_k$-induced map is $\nu_k$-Markovianizing from $R_A B'R_B$ for each $k$ and $\sum_{k \in \mathbb{K}} p_k \nu_k \leq \nu$.

**Definition 17** An $M$-induced map is $\iota$-Markovianizing from $A'R_A$ if $\Psi_{M}^{A'R_{A}B'R_{B}}$ is $\iota$-recoverable from $A'R_A$, that is, if there exists a linear CPTP map $\mathcal{R}' : R_A \to R_A B'R_B$ such that

$$\left\| \Psi_{M}^{A'R_{A}B'R_{B}} - \mathcal{R}'(\Psi_{M}^{A'R_{A}}) \right\|_1 \leq \iota.$$

A measurement $\mathbb{M}$ is $\iota$-Markovianizing from $A'R_A$ if an $M_k$-induced map is $\iota_k$-Markovianizing from $A'R_A$ for each $k$ and $\sum_{k \in \mathbb{K}} p_k \iota_k \leq \iota$.

The following two lemmas are at the core of the proofs of the main result, which translates the problem of finding the optimal costs for implementing a bipartite unitary to that of computing the Markovianizing cost of a particular state.

**Lemma 18** A measurement $\mathbb{M}$ is $(3\varsigma + 2\nu)$-decoupling between $A'R_A$ and $R_B$ if it is $\varsigma$-oblivious and $\nu$-Markovianizing from $R_A B'R_B$.

**Lemma 19** A measurement $\mathbb{M}$ is $(\varsigma + \mu)$-Markovianizing from $A'R_A$ if it is $\varsigma$-oblivious and $\mu$-decoupling between $A'R_A$ and $R_B$.

Let us describe a simplified version of the proof of the above two lemmas in the case of $\mu = \nu = \varsigma = 0$. The conditions of $\varsigma$-Markovianizing in Definition 16 and 17 are then equivalent to the condition that $\Psi_{M}^{A'R_{A}B'R_{B}}$ is a Markov state conditioned by $R_A$. Suppose an $M$-induced map is $0$-oblivious, which implies $\Phi_{M}^{RA} = \pi_d^{RA}$. Using (24), we see that

$$\Psi_{M}^{A'R_{A}B'R_{B}} = \hat{U} R_A R_B (\Phi_{M}^{A'R_{A}} \otimes \Phi_{M}^{B'R_{B}}) \hat{U} R_A R_B, \quad (28)$$

and consequently,

$$\Psi_{M}^{R_A R_B} = \hat{U} R_A R_B (\pi_d^{RA} \otimes \pi_d^{RB}) \hat{U} R_A R_B$$

$$= \pi_d^{RA} \otimes \pi_d^{RB}. \quad (29)$$

Therefore, for the state $\Psi_{M}^{A'R_{A}B'R_{B}}$, we have $I(A' : B | R_A R_B) = 0$ due to the local unitary invariance of QCMI, as well as $I(R_A : R_B) = 0$. It follows that

$$I(A' : B|R_B|R_A) = I(A' : R_B|R_A) + I(A' : B|R_A R_B)$$

$$= I(A'R_A : R_B) - I(R_A : R_B)$$

$$= I(A'R_A : R_B).$$
which implies the equivalence between the conditions of decoupling and Markovianizing under the condition of obliviousness.

For rigorous proofs, we need to relax the “exact” condition ($\mu = \nu = \zeta = 0$) to the “approximate” condition ($\mu, \nu, \zeta > 0$). See Appendices D-A and D-B for details.

B. Conditions for Achievability

For the proof of the direct part of Theorem 5 let us consider how to construct a successful protocol. Let $\hat{B}$ be a register on Bob’s side which has a sufficiently large dimension. The following lemma states that Markovianization by Alice’s measurement is a sufficient condition for the success of the first half of $\mathcal{M}$, in which $\Phi^{BBB}_d$ is obtained from $|\Psi\rangle$.

**Lemma 20** (See Appendix D-C for a proof.) Suppose that a measurement $\mathcal{M}$ is $0$-oblivious and $\mu$-decoupling between $A^1R_A$ and $R_B$, $\mu \in (0, 1]$, and that $\Psi^A_{M_k}R_A^RBB_R$ does not depend on $k$. Then, there exist pure states $|\psi^p\rangle A^1R_A\hat{B}, |\psi^p\rangle A^1R_A\hat{B}BB_R$ and isometries $W_k^{BBB}_B \rightarrow \hat{B}$ ($k \in \mathbb{K}$) such that

$$|||\psi^p\rangle A^1R_A\hat{B} - \pi_d^{BB}_A||_1 \leq 5\sqrt{\mu}$$

and $|\psi^p\rangle = W_k|\Psi_{M_k}\rangle$ for any $k \in \mathbb{K}$. In addition, $|\psi^p\rangle$ satisfies

$$|||\psi^p\rangle A^1R_A - \pi_d^{BB}_A||_1 \leq 3\sqrt{\mu}, \text{ rank}[|\psi^p\rangle A^1R_A\hat{B}] \leq \dim \mathcal{H}_{B_0}.$$  

The following lemma immediately follows from Lemma 18 and 20.

**Lemma 21** Suppose that a measurement $\mathcal{M}$ is $0$-oblivious and $\nu$-Markovianizing from $R_A^RBB_R$, $\nu \in (0, 1/2]$, and that $\Psi^A_{M_k}R_A^RBB_R$ does not depend on $k$. Then, there exist pure states $|\psi^p\rangle A^1R_A\hat{B}, |\psi^p\rangle A^1R_A\hat{B}BB_R$ and isometries $W_k^{BBB}_B \rightarrow \hat{B}$ ($k \in \mathbb{K}$) that satisfy

$$|||\psi^p\rangle A^1R_A\hat{B} - \pi_d^{BB}_A||_1 \leq 5\sqrt{2\nu},$$

$$|\psi^p\rangle = W_k|\Psi_{M_k}\rangle$$

and

$$|||\psi^p\rangle A^1R_A - \pi_d^{BB}_A||_1 \leq 3\sqrt{2\mu}, \text{ rank}[|\psi^p\rangle A^1R_A\hat{B}] \leq \dim \mathcal{H}_{B_0}.$$  

The task remaining after obtaining $\Phi_d^{BBB}$ is to obtain $\Phi_d^{A_R}$ from tripartite pure states $|\psi^p\rangle A^1R_A\hat{B}$, which is equivalent to performing state merging from Bob to Alice. Consequently, we can construct a successful protocol by combining Markovianization of $|\Psi\rangle$ and the subsequent state merging of $|\psi^p\rangle$ from Bob to Alice.

C. Conditions for Optimality

For the proof of the converse part of Theorem 5 let us analyze conditions on Alice’s measurement imposed by (25). Let $\mathcal{M} = \{\mathcal{M}_k\}_{k \in \mathbb{K}}$ be Alice’s measurement in protocol $\mathcal{M}$ that satisfies (25). First, conservation of the maximal entanglement between systems $A$ and $R_A^RBB_B$ immediately implies that Alice’s measurement must be oblivious. Second, since the final state is close to $|\Phi_d\rangle^{A_RA_B} |\Phi_d\rangle^{BBB_B}$, correlation between $A^1R_A$ and $R_B$ is destroyed by $\mathcal{M}$. This part of decoupling must be accomplished by Alice’s measurement alone, which implies that Alice’s measurement must be Markovianizing due to Lemma 19. Hence we obtain the following lemma.

**Lemma 22** (Appendices E-A and E-B) The measurement $\mathcal{M}$ is $4\sqrt{\epsilon}$-oblivious and $12\sqrt{\epsilon}$-Markovianizing from $A^1R_A$.

Let us continue to analyze conditions on Bob’s measurement imposed by (25). Let $\mathcal{B}$ be an ancillary system, and let $W_k : BB_0 \rightarrow BB_1B_2$ be an isometry such that the Naimark extension of Bob’s measurement $\mathcal{N}_k$ is given by $N_{ik} = \{i\}^{B_2}W_k$. Define $|\Psi_k\rangle := W_k|\Psi_k\rangle$. The following lemma states that Bob’s measurement is decomposed into two parts: (i) performing an isometry to obtain $\Phi_d^{BBB_B}$, and (ii) performing a measurement on his share of a “tripartite” pure state on $A^1, R_A$ and $B_1B_2$.

**Lemma 23** (Appendix E-C) There exist pure states $|\psi^p_k\rangle A^1R_A B_1B_2$ ($k \in \mathbb{K}$) such that

$$\sum_{k \in \mathbb{K}} p_k |||\psi_k^p\rangle A^1R_A B_1B_2 - |\Phi_d\rangle^{AAA}_A B_1B_2 \otimes \Phi_d^{BBB_B}\rangle|| \leq 4\sqrt{\epsilon}.$$  

By the measurement on $B_2$ described by $\{\{i\}\}^B$ and an operation on $A^1$ depending on $k$ and $l$, the maximal
entanglement $\Phi_d^{AR_A}$ must be obtained from $|\Psi_k^n\rangle^{A'RA_RB_R}$. This transformation is equivalent to state merging from Bob to Alice. Consequently, any successful protocol is described as a combination of Markovianization of $|\Psi\rangle$ and the subsequent state merging of $|\Psi_k^n\rangle$ from Bob to Alice.

VI. PROOF OF THEOREM 5

Let us return to the asymptotic scenario and prove Theorem 5. We consider protocols that transforms a state $|\Psi\rangle^{A'RA_RB_R}|\Phi_{K_n}\rangle_{A_0B_0}$ into $|\Phi_d^{AR_A}\rangle^{A_0B_0}$ as depicted in the right side of Figure 6. Conditions obtained in Section V directly apply by the following correspondence:

$$A, B, R_A, R_B \rightarrow \tilde{A}, \tilde{B}, \tilde{R}_A, \tilde{R}_B, \ U \rightarrow U^{\otimes n}, \ \Phi_d \rightarrow \Phi_d^{\otimes n}, \ \phi_{res} \rightarrow \Phi_{K_n}, \ \psi \rightarrow \Phi_{\otimes n}, \ \pi_d \rightarrow \pi_d^{\otimes n}.$$

As presented in Section V, two-round protocols for this task is decomposed into two steps (see the left side of Figure 6). The first step is composed of Alice’s measurement, forward classical communication and Bob’s isometry. Markovianization by Alice’s measurement satisfying the obliviousness condition is necessary and sufficient in order that Bob is able to obtain $(\Phi_d^{AR_R})^{\otimes n}$. The second step is composed of Bob’s measurement, backward classical communication and Alice’s local operation. To obtain $(\Phi_d^{AR_A})^{\otimes n}$, it is necessary and sufficient that the second step implements state merging of a particular tripartite state.

A. Direct Part

We prove the direct part of Theorem 5. We assume $L_n = 1$ in Definition 4 i.e., we consider a case where no entanglement is left after the protocol. The proof is by construction. Take arbitrary $R > M(U^\dagger)$, small $\epsilon, r > 0$, choose sufficiently large $n$ and let $K_n = 2^{n(R+r)}$. Divide the resource state $\Phi_{K_n}^{A_0B_0}$ as

$$\Phi_{K_n}^{A_0B_0} \rightarrow \Phi_{2nR}^{A_0B_0} \otimes \Phi_{2nR}^{R_R}.$$

Consider a protocol consisting of the following steps.

1) Alice’s measurement: By Definition 4 and Theorem 10, there exists a random unitary operation $V_n : \tau \rightarrow 2^{-nR} \sum_{j=1}^{2^{2nR}} V_j \tau V_j^\dagger$ on $\tilde{A}$ such that $V_n(\Phi_{\otimes n})$ is $c$-recoverable from $\tilde{R}_A\tilde{B}_R$. Using $V_j$ in $\tilde{V}_n$, construct Alice’s measurement $M = \{ M_k \}_{k=1}^{2^{2nR}}$ as

$$M_k^{A_0\rightarrow \tilde{A}} = \frac{1}{\sqrt{2^{2nR}}} \sum_{j=1}^{2^{2nR}} \exp(\frac{2\pi i j k}{2^{2nR}}) (j|A_0 \otimes \tilde{V}_j)^\dagger.$$

$M$ is 0-oblivious and $c$-Markovianizing from $\tilde{R}_A\tilde{B}_R$. In addition, the reduced state of the post-measurement state on $\tilde{A}\tilde{R}_A\tilde{B}_R$ does not depend on $k$. Indeed, we have $p_k = 2^{-nR}$ and

$$p_k^{-1} M_k (|\tau^\tilde{A} \otimes \Phi_{2nR}^{A_0B_0}) M_k^\dagger = V_n(\tau)$$

for $k = 1, \ldots, 2^{nR}$. Alice performs the measurement defined above.

2) Forward classical communication: Alice sends the measurement outcome $k$ to Bob.

3) Bob’s isometry: Due to Lemma 20, there exist pure states $|\tilde{\Phi}^{AR_A}\rangle$, $|\tilde{\Psi}^{AR_A}\rangle$ and isometries $\{ W_k^{B_0\rightarrow B_R} |_{k=1}^{2^{2nR}} \}$ that satisfy

$$|\tilde{\Psi}^{\otimes n}\rangle = p_k^{-1/2} (M_k \otimes W_k)|\Phi_{K_n}\rangle.$$  

for any $k$, and satisfy

$$|\tilde{\Phi}^{\otimes n}\rangle \approx \tilde{\Phi}^{AR_A} \tilde{B}_R,$$

$$|\tilde{\Psi}^{\otimes n}\rangle \approx \pi_d^{\otimes n} \tilde{B}_R,$$

with a small error. Bob performs $W_k$.

4) State merging: Alice and Bob perform state merging of

$$|\tilde{\Phi}^{A'\tilde{R}_A\tilde{B}_R}'\rangle := |\tilde{\Phi}^{AR_A\tilde{B}_R}\rangle |\Phi_{K_n}\rangle^{A_0},$$

where $A' = A\tilde{A}_0$ and $B' = \tilde{B}_B\tilde{B}_0$. Alice obtains a purification of $(\tilde{\Phi}^{\otimes n})^{AR_A}$ with a small error.

5) Alice’s isometry: Alice performs an isometry and obtains $(\tilde{\Phi}^{\otimes n})^{AR_A}$ within a small error.

The forward classical communication cost $nC_f$ is simply equal to $nR$ bits. As for Step 4), we consider a state merging in which no entanglement is obtained afterward. Thus the total entanglement cost is equal to the amount of entanglement that Alice and Bob have initially shared, i.e., $nE = n(R+r)$ bits of (32). Applying Theorem 13 and the rank inequality in (31) for $\tilde{\Phi}$, the backward classical communication cost is bounded above by

$$nC_b \leq \log \text{rank}[(\tilde{\Phi}^{\otimes n})^{B'}] = \log \text{rank}[(\tilde{\Phi}^{\otimes n})] + nr \leq \log \dim[H] + nr = n(R+r).$$

In total, we have $(E, C_f, C_b) = (R, R + r, R + r)$. The total error is evaluated by counting errors of (33), (34) and one induced by state merging (see Appendix I for the detail). Due to Lemma 21 the first two errors are bounded above by $5\sqrt{2}\epsilon$ and $3\sqrt{2}\epsilon$, respectively. Theorem 13 implies that the merging error $\epsilon_{\text{merg}}$ is bounded as

$$\epsilon_{\text{merg}} \leq 4 \cdot 2^{-nr/4}.$$ 

A simple calculation then yields an upper bound on the total error $\epsilon_{\text{tot}}$:

$$\epsilon_{\text{tot}} \leq 2\sqrt{5} \sqrt{2}\epsilon + 5\sqrt{2}\epsilon + 4 \cdot 2^{-nr/8}. \quad (36)$$

Since $\epsilon, r > 0$ can be arbitrarily small, we conclude that a rate triplet $(E, C_f, C_b) = (R, R, R)$ is achievable if $R > M(U^\dagger)$.

B. Reverse Part (Outline)

We prove the reverse part of Theorem 5 by combining (19) and (22). Let $M_n$ be a two-round LOCC protocol that satisfies Condition (6). We then have

$$F(\rho(M_n, U^\dagger), |\tilde{\Phi}^{\otimes n})^{B_0\rightarrow B_R}|\Phi_{K_n}\rangle^{A_0B_0}) \geq 1 - \epsilon$$

for

$$\rho(M_n, U^\dagger) := M_n(|\tilde{\Phi}^{\otimes n})^{A_0B_0}.$$
corresponding to (25). From Lemma 22, the map induced by Alice’s measurement in $M_i$ is $4\sqrt{\epsilon}$-oblivious and $12\sqrt{\epsilon}$-Markovianizing from $A'R_A$. Then the first two conditions in Definition 9 are satisfied by the following correspondence:

$$\tilde{A}, \tilde{B}, \tilde{C}, A_0, G \to \tilde{A}, \tilde{R}_A, (\tilde{B} \tilde{R}_B), A_0, B_0,$$

$$\langle \Psi \rangle_{ABC}^\epsilon \to \langle \Psi \rangle_{A'R_A(BR_B)}^\epsilon, \quad \phi_{\text{res}} \to \Phi_{K_\epsilon},$$

$$\epsilon \to 12\sqrt{\epsilon}.$$ (38)

Therefore, from (19), we have

$$H((p_k)_k) \geq \Delta S(A')_{av}
\geq \Delta S(A')_{av} - \Delta S(B_0)_{av}
\geq n M(U^\dagger) - n \zeta(12\sqrt{\epsilon} \cdot n).$$ (39)

Suppose a rate triplet $(E,C_f,C_b)$ is achievable by a two-round protocol. Due to Definition 2 and Assumption (9), there exists a sequence of $(2,n,\epsilon_n)$-protocols for implementing $U$, with the entanglement cost $nE$ and classical communication costs $nC_f$ and $nC_b$ for each, such that

$$\lim_{n \to \infty} 12\sqrt{\epsilon_n} \cdot n = 0.$$ (40)

The optimality of the forward classical communication cost immediately follows from (39) and $nC_f \geq H((p_k)_{k \in E})$.

As for the backward classical communication cost, recall that Bob’s measurement is decomposed into an isometry operation for obtaining $(\Phi_{d}^{\otimes n})^{BR_B}$ and a projective measurement on an ancillary system $B_E$. The latter forms state merging of $\Psi^p_{k'}$, together with the backward classical communication and the subsequent local operation by Alice. Thus the backward classical communication cost is equal to the one required for performing state merging of $\Psi^p_{k'}$. Due to (22) with the correspondence $A \to A'$, $B \to B_1 B_E$ and $R \to R_A$, the cost is given by $I(B_1 B_E : R_A)_{BP}$. Because of $\Delta S(A')_{av} \geq n M(U^\dagger)$ in (39), this cost turns out not to be smaller than $n M(U^\dagger)$.

The amount of entanglement obtained after state merging is bounded by (22) as $\log L_n \leq \tau(B_1 B_E | A')_{\text{BP}}$, which implies the optimality of the total cost of entanglement when combined with $\Delta S(A')_{av} - \Delta S(B_0)_{av} \geq n M(U^\dagger)$. See Appendix C for a detailed proof.

VII. PROPERTIES OF THE COST

In this section, we investigate properties of the Markovianizing cost of units. The results obtained here will be used in the next section for analyzing examples.

Consider a tripartite pure state $|\Psi_U\rangle^{AR_A(BR_B)}$ defined by (8). The Petz recovery map $R_U : A \to A(BR_B)$ corresponding to $|\Psi_U\rangle^{AR_A(BR_B)}$ is defined by

$$R_U(\tau) = (\Psi_U^{AR_A(BR_B)})^\epsilon \cdot \tau(\Psi_U^{AR_A(BR_B)})^\epsilon = U^{AB} \text{Tr}_B[U^{AB}(\tau A^\otimes B U^{AB} \otimes \Phi_{d}^{BR_B}) U^{AB}]$$

for $\tau \in S(\mathcal{H}^A)$. Define CPTP maps $\mathcal{E}_U$ and $\mathcal{E}_{U,\infty}$ on $A$ by

$$\mathcal{E}_U := \text{Tr}_{BR_B} \circ R_U, \quad \mathcal{E}_{U,\infty} := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^N \mathcal{E}_U^n,$$ (41)

and consider the states

$$\Psi_{U,\infty}^{AR_A(BR_B)} := \mathcal{E}_{U,\infty}(|\Psi_U\rangle\langle\Psi_U|^{AR_A(BR_B)}),$$

$$\Psi_U^{AR_A} := \mathcal{E}_U(|\Psi_U\rangle\langle\Psi_U|^A R_B),$$ (42)

As we prove below, the map $\mathcal{E}_U$ is self-adjoint in the sense that $\mathcal{E}_U = \mathcal{E}_U^*$. Therefore, due to Theorem 9 in [12], the Markovianizing cost of $U$ is given by

$$M(U) = M|_{AR_A}(\Psi_U^{AR_A(BR_B)}) = S(\Psi_U^{AR_A(BR_B)}),$$

which can be computed by a finite-step algorithm proposed in [12] (see Section III therein). Due to the unitary invariance of the von Neumann entropy, it immediately follows that

$$M(U) = S(\Phi_{d}^{AR_A}).$$ (43)

As a consequence, the Markovianizing costs of unitaries $U_1$ and $U_2$ are equal if they are local unitarily equivalent, that is, if there exist unitaries $v, v'$ on $\mathcal{H}^A$ and $w, w'$ on $\mathcal{H}^B$ such that $U_1 = (v \otimes w) U_2 (v' \otimes w')$.

Let us also analyze the Schmidt strength of unitaries. Consider Decomposition (10) of a bipartite unitary $U$. A CPTP map $\mathcal{E}_U$ defined by (41) takes the form

$$\mathcal{E}_U(\tau) = \sum_{ss'} c^2_{s} E^s_s \tau E^s_s,\quad E^s_s := \sum_{kk} \tilde{E}^s_{kk},$$ (44)

where we introduced notations $\tilde{E}^s_{kk} := c_{s} c^*_s$ and $\tilde{E}^s_{ss'} := E^s_s \tau E^s_{s'}$. It is straightforward to verify that $\mathcal{E}_U$ is self-adjoint, that is, it satisfies

$$\mathcal{E}_U(\tau) = \mathcal{E}_U^*(\tau) := \sum_{ss'} c^2_{s} \tilde{E}^s_{ss'} \tau E^s_{ss'}.$$ (45)

Due to the orthonormality of $\{E_s|\Phi_{d}\rangle^{AR_A}\}_s$, and $\{F_s|\Phi_{d}\rangle^{BR_B}\}_s$, which follows from (11), the eigen decomposition of $\Psi_U^{AR_A}$ is given by

$$\Psi_U^{AR_A} = \sum_{s=0}^{d^2-1} c^2_{s} E^s_s |\Phi_{d}\rangle \langle\Phi_{d}|^{AR_A} E^s_s A.$$ (46)

Thus we have

$$K(U) = S(AR_A|\Phi_{d}_U).$$

The following lemma provides a lower bound on the Markovianizing cost of unitaries.

Lemma 24 $M(U) \geq K(U)$ holds for any bipartite unitary $U$.

Proof: Define quantum operations $e$ and $e^*$ on $S(\mathcal{H}^A)$ by

$$e(\tau) = \sum_{s} c^2_{s} E^s_s \tau E^s_s,\quad e^*(\tau) = \sum_{s} c^2_{s} E^s_s \tau E^s_s.$$ (47)

From (45) and (46), the Schmidt strength of the unitary is given by

$$K(U) = S(e^A(|\Phi_{d}\rangle \langle\Phi_{d}|^{AR_A})).$$

It immediately follows from (44) that $\mathcal{E}_U = e^* \circ e$. We have

$$e(\tau) = \text{Tr}_B[U^{AB}(\tau A^\otimes B U^{AB})],$$

$$e^*(\tau) = \text{Tr}_B[U^{AB}(\tau A^\otimes B U^{AB})].$$
due to (11) and (11), which implies that \( e \) and \( e^* \) are unitary, i.e., \( e(I) = e^*(I) = I \). Therefore, owing to the monotonicity of the von Neumann entropy under unital maps, we have
\[
S((\mathcal{E}_U^A)^n(|\Phi_d\rangle\langle\Phi_d|^{AR})) \geq S(e^A(|\Phi_d\rangle\langle\Phi_d|^{AR})) \tag{48}
\]
for any \( n \geq 1 \). Due to Definitions (41), (42) and the concavity of the von Neumann entropy, we obtain
\[
S(\Phi_d^{AR}) \geq \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} S((\mathcal{E}_U^A)^n(|\Phi_d\rangle\langle\Phi_d|^{AR})). \tag{49}
\]
Expressions (43), (47), (48) and (49) yields \( M(U) \geq K(U) \).

VIII. Examples

In this section, we consider two classes of bipartite unitaries and compute their Markovianizing costs.

A. Two-Qubit Unitaries

It is proven in [17] that all two-qubit unitaries are classified into the following categories:

1) Unitaries that can be written as a tensor product of local unitaries as \( U = u^A \otimes u^B \). We do not consider this type of unitaries because of its triviality.

2) Unitaries that can be written in the form of
\[
U = \cos \left( \frac{\theta}{2} \right) I^A \otimes I^B + i \sin \left( \frac{\theta}{2} \right) \sigma_{s}^A \otimes \sigma_{s}^B \tag{50}
\]
up to local unitaries, where \( \theta \in (0, \pi/2] \). Controlled-unitary gates are examples of such unitaries.

3) Unitaries that can be written in the form of
\[
U = c_0 e^{i \theta_0} I^A \otimes I^B + c_1 e^{i \theta_1} \sigma_{s}^A \otimes \sigma_{s}^B + c_2 e^{i \theta_2} \sigma_{s}^A \otimes \sigma_{s}^B + c_3 e^{i \theta_3} \sigma_{s}^A \otimes \sigma_{s}^B
\]
up to local unitaries, where \( c_s, \theta_s \in \mathbb{R} \) (\( s = 0, 1, 2, 3 \)) are nonnegative real parameters satisfying \( \sum_{s=0}^{3} c_s^2 = 1 \). All two-qubit unitaries that are not classified to the first two categories are of this category.

Let us consider unitaries of the form (50), which is local unitarily equivalent to the following controlled phase gate:
\[
U_{\theta} = |0\rangle\langle 0|^A \otimes I^B + |1\rangle\langle 1|^A \otimes (e^{i \theta_{s}})^B.
\]

We have
\[
c_0 = \cos \left( \frac{\theta}{2} \right), \quad c_1 = \sin \left( \frac{\theta}{2} \right),
\]
\[
c_{00}^2 + c_{11}^2 = \frac{1}{2}(1 + \cos^2 \theta), \quad c_{01}^2 + c_{10}^2 = \frac{1}{2} \sin^2 \theta,
\]
\[
E_0 = F_0 = I, \quad E_1 = \sigma_z, \quad F_1 = i \sigma_z,
\]
\[
\bar{E}_0 = \bar{E}_{11} = I, \quad \bar{E}_1 = \bar{E}_{10} = \sigma_z.
\]

Thus a map corresponding to (44) is given by
\[
\mathcal{E}(\tau) = \frac{1 + \cos^2 \theta}{2} \cdot \tau + \frac{1}{2} \sin^2 \theta \cdot \sigma_z \tau \sigma_z,
\]
which leads to
\[
\mathcal{E}_{U,\infty}(\tau) = \frac{1}{2}(\tau + \sigma_z \tau \sigma_z) = |0\rangle\langle 0| \tau |0\rangle\langle 0| + |1\rangle\langle 1| \tau |1\rangle\langle 1|
\]
from (41). Hence we have
\[
\Phi_{U,\infty}^{AR} = \frac{1}{2}(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|)
\]
corresponding to (42), which implies \( M(U) = 1 \) due to (43).

Consequently, we obtain the following theorem:

Theorem 25 A rate triplet \((E, C_f, C_b)\) is achievable by two-round protocols for implementing a qubit controlled-unitary gate only if \( E, C_f, C_b \geq 1 \), if we additionally require in Definition 2 that Condition (9) holds.

The above theorem implies that, counterintuitively, at least one ebit of entanglement consumption per copy is necessary for implementing two-qubit controlled-unitary gate by two round protocols, regardless of how close the unitary is to the identity operation (i.e., regardless of how small \( \theta \) is). In [23], we prove that a certain class of two-qubit controlled-unitary gates can be implemented by a four-round protocol with the entanglement cost strictly smaller than one ebit per copy. Thereby we reveal a trade-off relation between the entanglement cost and the number of rounds for a LOCC task.

B. Generalized Clifford Operators

The generalized Pauli operators \( \sigma_{pq}^A(p, q \in \{1, \ldots, d\}) \) on a \( d \)-dimensional Hilbert space is defined as
\[
\sigma_{pq} := \sum_{t=1}^{d} |t-p\rangle\langle t|, \quad \sigma_{0q} := \sum_{t=1}^{d} e^{2\pi iqt/d} |t\rangle\langle t|,
\]
\[
\sigma_{pq} := \sigma_{0q}, \tag{51}
\]
with a fixed basis \( \{|t\rangle\}_{t=1}^{d} \). Here, subtraction is taken with mod \( d \). A bipartite unitary \( U \) is called a generalized Clifford operator if, for any \( p, q, r \) and \( s \), there exist \( p', q', r', s' \) and a phase \( \theta_{pqrs} \in \mathbb{R} \) such that
\[
U(\sigma_{pq} \otimes \sigma_{rs}) U^\dagger = e^{i \theta_{pqrs}} \sigma_{p'q'} \otimes \sigma_{r's'}.
\]
The Markovianizing cost of generalized Clifford operators can be simply computed by the following theorem, a proof of which is given in Appendix [H]

Theorem 26 \( M(U) = K(U) \) holds for any generalized Clifford operator \( U \).

As a corollary of Theorem 5 and 26, the Schmidt strength \( K(U) \) is equal to the minimum cost of entanglement and classical communication for implementing generalized Clifford operator by two-round protocols under additional assumption [9]. A stronger statement, represented by the following theorem, immediately follows from Theorem 6 and 26.

Theorem 27 The following statements hold for any generalized Clifford operator \( U \) and \( r \geq 2 \):

- **Direct:** A rate triplet \((E, C_f, C_b)\) is achievable by \( r \)-round protocols for implementing \( U \) if \( E, C_f, C_b > K(U) \).
- **Converse:** A rate triplet \((E, C_f, C_b)\) is achievable by \( r \)-round protocols for implementing \( U \) only if \( E, C_f, C_b \geq K(U) \).
IX. Open Problems

We have derived a converse bound on the cost of entanglement and classical communication for implementing a bipartite unitary by two-round protocols. However, we do not know whether the converse bound remains to hold when we remove the additional requirement on the convergence speed of error, represented by Inequality \( (40) \). In this section, we investigate a relation between this open problem and another regarding a property of approximate recoverability.

In the proof of the converse part, Condition \( (9) \) is exploited in the form of Inequality \( (40) \). This inequality is required to prove the convergence of an error term in \( (39) \), which depends not only on \( \epsilon \) but also on \( n \). The \( n \)-dependence of the error term originates from that in Inequality \( (19) \), and the latter arises due to the fact that Condition \( (18) \) is required to prove \( (17) \). In summary, we require Condition \( (9) \) to prove the converse part of Theorem 5 because we require Condition \( (18) \) to prove \( (17) \).

In \([13]\), we proved that Condition \( (18) \) in Theorem 10 can be eliminated if a conjecture about approximate recoverability is true. The conjecture states that a tripartite quantum state \( \rho^{ABC} \) is approximately recoverable from \( \rho^{AB} \) by an operation \( \mathcal{R} : B \rightarrow BC \) if it is approximately recoverable from \( \rho^{BC} \) by an operation \( \mathcal{R}' : B \rightarrow AB \), up to a dimension-independent rescaling of error of recovery. A rigorous statement is as follows:

Conjecture 28 (Proposition 13 (unproven) in [13]) There exists a nonnegative function \( g(\epsilon) \), independent of the dimension of quantum systems and satisfies \( \lim_{\epsilon \rightarrow 0} g(\epsilon) = 0 \), such that the following statement holds for an arbitrary tripartite state \( \rho^{ABC} \) and \( \epsilon > 0 \): The state \( \rho^{ABC} \) is \( g(\epsilon) \)-recoverable from \( BC \) if it is \( \epsilon \)-recoverable from \( AB \).

Condition \( (9) \) in Theorem 5 can be eliminated if the above conjecture is true. See also Appendix G-C

X. Communication Power of a LOCC Protocol

In this section, we analyze classical communication power of a LOCC protocol with an arbitrary preshared resource state. The results obtained here will be used in the next section to prove Theorem 6.

Consider the following scenario in which Alice aims to transmit \( nR \) bits of classical message to Bob by a bidirectional LOCC protocol that transforms a preshared quantum state \( \rho^{AB} \).

- Alice and Bob initially share a bipartite quantum state \( \rho^{AB} \).
- Alice is given an array of uniformly random classical bits \( \vec{X} = X_1 \cdots X_{nR} \).
- Alice and Bob transforms \( \rho^{AB} \) by an LOCC protocol.
- Alice’s operations during the protocol, as well as the message from Alice to Bob, may depend on \( \vec{X} \).
- After the completion of the protocol, Bob performs a measurement on \( B \) to decode \( \vec{X} \).

Let \( \vec{X}' \) be the result of Bob’s decoding measurement. The decoding error is defined by

\[
P_e := \Pr\{\vec{X} \neq \vec{X}'\}.
\]

In the following, we prove that the length \( nR \) of classical bits transmitted from Alice to Bob during the protocol, if the decoding error is vanishingly small.

Without loss of generality, we assume that the protocol proceeds as follows. Here, \( \Gamma \) is a natural number, and \( K_\gamma, L_\gamma, \hat{K}_\gamma, \hat{L}_\gamma \) are random variables which take values in finite sets \( K_\gamma, L_\gamma, \hat{K}_\gamma, \hat{L}_\gamma \), respectively.

1) Alice and Bob recursively apply the following operation from \( \gamma = 1 \) to \( \gamma = \Gamma \):
   a) Alice performs a measurement \( \mathcal{M}_\gamma \) on her system and obtains an outcome \( K_\gamma \).
   b) Alice transmits a classical message \( \hat{K}_\gamma \) to Bob.
   c) Bob performs a measurement \( \mathcal{N}_\gamma \) on his system and obtains an outcome \( L_\gamma \).
   d) Bob transmits a classical message \( \hat{L}_\gamma \) to Alice.
2) Alice performs a quantum operation on her system.

Fig. 7. A graphical representation of the \( \gamma \)-th step in an LOCC protocol is depicted. We denote system \( A \) and \( B \) after the \( \gamma \)-th step by \( A_\gamma \) and \( B_\gamma \) for \( \gamma = 1, \cdots, \Gamma \), respectively.
The total number of classical bits, transmitted from Alice to Bob during the protocol, is given by

\[ C_{\text{tot}} := \sum_{\gamma=1}^{\Gamma} \log |\mathcal{K}_\gamma|. \]

Let us introduce the following notations:

\[ K^\gamma := (K_1, \cdots, K_\gamma), \quad L^\gamma := (L_1, \cdots, L_\gamma) \]

\[ \hat{K}^\gamma := (\hat{K}_1, \cdots, \hat{K}_\gamma), \quad \hat{L}^\gamma := (\hat{L}_1, \cdots, \hat{L}_\gamma) \]

In general, Alice and Bob’s measurement in the protocol, as well as classical messages, may depend on the previous measurement outcomes and messages in the following way (Figure 7).

- \( M_\gamma \) depends on \( (\hat{X}, K^{\gamma-1}, \hat{L}^{\gamma-1}) \).
- \( N_\gamma \) depends on \( (L^{\gamma-1}, \hat{K}^\gamma) \).
- \( \hat{K}_\gamma \) depends on \( (\hat{X}, K^\gamma, \hat{L}^\gamma) \).
- \( \hat{L}_\gamma \) depends on \( (L^\gamma, \hat{K}^{\gamma-1}) \).

The following lemma states that the mutual information between \( \hat{X} \) and all that Bob has after the protocol is bounded above by the total amount of classical communication transmitted from Alice to Bob during the protocol. See Appendix [A] for a proof.

**Lemma 29** The following inequalities hold:

\[ I(\hat{X} : B_\gamma^\Gamma, K^{\Gamma^\Gamma}) \leq C_{\text{tot}}, \quad (52) \]

\[ nR \leq C_{\text{tot}} + h(P_\epsilon) + nR P_\epsilon. \quad (53) \]

Here, \( h(x) \) is the binary entropy defined by

\[ h(x) := -x \log x - (1 - x) \log (1 - x), \]

and \( B_\gamma^\Gamma \) denotes system \( B \) after the \( \Gamma \)-th step of the protocol.

**Remark.** An upper bound on the classical communication power of a two-way LOQC (local operations and quantum communication) protocol, which is similar to (53), has been proven in [24].

### XI. PROOF OF THEOREM 6

We prove Theorem 6 in this section, based on the idea that the cost of entanglement and classical communication for implementing a unitary is not smaller than powers of the unitary for generating entanglement and transmitting classical information ([10], [11], [25]).

Let us analyze power of a bipartite unitary for transmitting classical information. The following lemma states that the Schmidt strength is a lower bound on the classical communication power of a bipartite unitary. See Appendix [1-B] for a proof.

**Lemma 30** For any \( \epsilon \in (0, 1) \) and sufficiently large \( n \), let \( U_n \) be a quantum operation on \( AB \) that satisfy

\[ F(\rho(U_n), |\Psi_U\rangle^\otimes n) \geq 1 - \epsilon \quad (54) \]

for

\[ \rho(U_n) := U_n(|\Phi^{AR}_d\rangle^\otimes n|\Phi^{BR}_d\rangle^\otimes n). \]

Then \( U_n \) has a capacity to transmit \( n(K(U) - \epsilon) \) bits of classical information from Alice to Bob up to an error \( 5\sqrt{\epsilon} \), when assisted by shared entanglement.

Theorem 6 is then proved as follows.

**Proof of Theorem 6** Suppose a rate triplet \( (C_f, C_b, C_h) \) is achievable. By definition, for any \( \epsilon > 0 \) and sufficiently large \( n \), there exist \( K_n \) and \( L_n \) that satisfy \( \log K_n - \log L_n = n\hat{E} \), and a LOCC protocol \( \mathcal{M}_n \) that satisfies (6), with the forward and backward classical communication cost \( nC_f \) and \( nC_b \), respectively.

Define a quantum operation \( \hat{\mathcal{M}}_n \) on \( \hat{A}\hat{B} \) by

\[ \hat{\mathcal{M}}_n : \tau \rightarrow \text{Tr}_{A_1B_1}[\mathcal{M}_n(\tau\hat{A}\hat{B} \otimes \Phi_{K_0B_0})]. \]

Due to Lemma 30, \( \hat{\mathcal{M}}_n \) has a capacity to transmit \( n(K(U) - \epsilon) \) bits of classical information from Alice to Bob up to an error \( 5\sqrt{\epsilon} \), when assisted by shared entanglement. By definition, \( \mathcal{M}_n \) has the same capacity. Applying Lemma 29 yields

\[ n(K(U) - \epsilon) \leq nC_f + h(5\sqrt{\epsilon}) + n\epsilon(K(U) - \epsilon), \]

which leads to

\[ (1 - \epsilon)(K(U) - \epsilon) \leq C_f + \frac{1}{n}h(5\sqrt{\epsilon}). \]

Since \( \epsilon > 0 \) can be arbitrarily small, we obtain \( C_f \geq K(U) \).

Exchanging roles of Alice and Bob, we also have \( C_b \geq K(U) \).

To prove \( E \geq K(U) \), we assume for simplicity that \( K_n \) and \( L_n \) is bounded above as

\[ \log K_n, \log L_n \leq n \log \kappa \quad (55) \]

with a constant \( \kappa > 0 \). We quantify entanglement of states between systems \( \hat{A}\hat{R}_A \hat{A}_0 \) and \( \hat{B}\hat{R}_B \hat{B}_0 \) (or between \( \hat{A}\hat{R}_A \hat{A}_1 \) and \( \hat{B}\hat{R}_B \hat{B}_1 \)) by an entanglement measure that satisfy asymptotic continuity ([26], see Appendix [A-D]). We denote it by \( \hat{E} \).

Since \( \hat{E} \) is equal to the entanglement entropy for pure states, we have

\[ \hat{E}(|\Psi_U\rangle^\otimes n|\Phi_{L_n}\rangle^\otimes n|\Phi_{A_0B_0}\rangle) = nK(U) + \log L_n, \]

\[ \hat{E}(|\Phi^{AR}_d\rangle^\otimes n|\Phi^{BR}_d\rangle^\otimes n|\Phi_{K_0A_0B_0}\rangle) = \log K_n \]

from (46). Due to asymptotic continuity, Condition (6) and (55) implies

\[ \hat{E}(\rho(\mathcal{M}_n)) \geq nK(U) + \log L_n - n\delta(2\sqrt{\epsilon}) \log (d^4\kappa^2), \quad (56) \]

where \( \delta(\epsilon) \) is an \( n \)-independent nonnegative function that satisfies \( \lim_{n \to 0} \delta(\epsilon) = 0 \). Equality (2) and the monotonicity of \( \hat{E} \) under LOCC operations yield

\[ \hat{E}(\rho(\mathcal{M}_n)) \leq \log K_n. \quad (57) \]

Combining (56) and (57), we obtain

\[ E = \frac{1}{n} (\log K_n - \log L_n) \geq K(U) - \delta(2\sqrt{\epsilon}) \log (d^4\kappa^2), \]

which implies \( E \geq K(U) \) by taking the limit of \( \epsilon \to 0 \).
A. Fidelity, Trace Distance and Uhlmann’s Theorem

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

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

It satisfies

$$0 \leq \|\rho - \sigma\|_1 \leq 2$$

and

$$\|\rho - \sigma\|_1 = 2 \max_{\Lambda} \text{Tr}[\Lambda(\rho - \sigma)],$$

where the maximization is taken over all linear operators $\Lambda$ on $\mathcal{H}$ that satisfy $0 \leq \Lambda \leq I$.

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

$$\|\rho - \tau\|_1 \leq \|\rho - \sigma\|_1 + \|\sigma - \tau\|_1,$$  \hspace{1cm} (59)

which is called the triangle inequality. For two ensembles $\{p_i, \rho_i\} \text{ and } \{\rho_i, \sigma_i\}$, we have

$$\sum_i p_i \|\rho_i - \sigma_i\|_1 \geq \left\| \sum_i p_i \rho_i - \sum_i p_i \sigma_i \right\|_1.$$  \hspace{1cm} (60)

The trace distance takes a simple form under tensor product, i.e., for any $\rho, \sigma \in S(\mathcal{H}^A)$ and $\tau \in S(\mathcal{H}^B)$, we have

$$\|\rho^A \otimes \tau^B - \sigma^A \otimes \tau^B\|_1 = \|\rho^A - \sigma^A\|_1.$$  \hspace{1cm} (61)

The fidelity between two quantum states $\rho, \sigma \in S(\mathcal{H})$ is defined by

$$F(\rho, \sigma) := \left( \text{Tr} \left[ \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right] \right)^2,$$

and satisfies

$$0 \leq F(\rho, \sigma) \leq 1.$$  \hspace{1cm} (62)

The fidelity takes a simple form for pure states as

$$F(|\psi\rangle, |\phi\rangle) = |\langle \psi | \phi \rangle|^2$$  \hspace{1cm} (63)

and

$$F(\rho, |\phi\rangle) = \langle \phi | \rho |\phi\rangle,$$

the latter of which yields

$$\sum_k p_k F(\rho_k, |\phi\rangle) = F \left( \sum_k p_k \rho_k, |\phi\rangle \right)$$  \hspace{1cm} (64)

for any ensemble $\{p_k, \rho_k\}_k$.

Let $|\psi_\rho\rangle, |\psi_\sigma\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ be arbitrary purifications of $\rho, \sigma \in S(\mathcal{H}^A)$, respectively. Due to Uhlmann’s theorem [33], we have

$$F(\rho, \sigma) = \max_{W} \left| \langle \psi_\rho | (I^A \otimes W^B) | \psi_\sigma \rangle \right|^2$$  \hspace{1cm} (65)

and

$$F(\rho, |\phi\rangle) = \max_{|\psi_\rho\rangle} \left| \langle \psi_\rho | \psi_\sigma \rangle \right|^2.$$  \hspace{1cm} (66)

Here, the maximization in the first line is taken over all unitaries $W$ acting on $\mathcal{H}^B$, and in the second line over all purifications $|\psi_\rho\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ of $\sigma$. It immediately follows that, for an arbitrary pure states $|\Psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ and $|\phi\rangle \in \mathcal{H}^A$, we have

$$F \left( |\Psi\rangle, |\phi\rangle \right) = \max_{|\psi_\rho\rangle \in \mathcal{H}^B} F \left( |\Psi\rangle^A B, |\phi_\rho\rangle^A |\varphi\rangle^B \right),$$  \hspace{1cm} (67)

where the maximization is taken over all pure states on system $B$.

The trace distance and the fidelity are monotonic under quantum operations, i.e., it satisfies

$$\|\rho - \sigma\|_1 \geq \|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1,$$  \hspace{1cm} (68)

$$F(\rho, \sigma) \leq F(\mathcal{E}(\rho), \mathcal{E}(\sigma))$$

for any $\rho, \sigma \in S(\mathcal{H})$ and any linear CPTP map $\mathcal{E} : S(\mathcal{H}) \to S(\mathcal{H}')$. In particular, the two functions are monotonic under taking the partial trace, that is, for any $\rho, \sigma \in S(\mathcal{H}^A \otimes \mathcal{H}^B)$ we have

$$\|\rho^{AB} - \sigma^{AB}\|_1 \geq \|\rho^A - \sigma^A\|_1,$$  \hspace{1cm} (69)

$$F(\rho^{AB}, \sigma^{AB}) \leq F(\rho^A, \sigma^A).$$  \hspace{1cm} (70)
The two functions are invariant under unitary operations, namely, for any unitary \( U \) acting on \( \mathcal{H} \) we have
\[
\|\rho - \sigma\|_1 = \|U\rho U^\dagger - U\sigma U^\dagger\|_1, \tag{71}
\]
\[
F(\rho, \sigma) = F(U\rho U^\dagger, U\sigma U^\dagger). \tag{72}
\]

The trace distance and the fidelity satisfy the following relation in general:
\[
1 - \sqrt{F(\rho, \sigma)} \leq \frac{1}{2} \|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)}. \tag{73}
\]

Therefore, if \( F(\rho, \sigma) \geq 1 - \epsilon \) then \( \|\rho - \sigma\|_1 \leq 2\sqrt{\epsilon} \).

Conversely, if \( \|\rho - \sigma\|_1 \leq \epsilon \) then \( F(\rho, \sigma) \geq 1 - \epsilon \).

Let \( \mathcal{E} \) be a quantum operation on a system described by a \( d \)-dimensional Hilbert space \( \mathcal{H} \), and \( U \) be a unitary acting on \( \mathcal{H} \). How precisely \( \mathcal{E} \) approximates a unitary operation \( U(\cdot) = U(\cdot)U^\dagger \) is evaluated by the average fidelity and the entanglement fidelity, which are defined as
\[
F_{av}(\mathcal{E}; U) := \int_{\text{Haar}} p(d\psi) F(\mathcal{E}(|\psi\rangle), |U\psi\rangle), \tag{74}
\]
and
\[
F_{en}(\mathcal{E}; U) := F((\mathcal{E} \circ \text{id})(|\Phi\rangle), (U \otimes I)|\Phi\rangle), \tag{75}
\]
respectively. Here, the integral in \( (74) \) is taken with respect to the Haar measure on \( \mathcal{H} \), and \( |\Phi\rangle \in \mathcal{H} \) is a maximally entangled state with the Schmidt rank \( d \). As proved in \([14]\), it holds that
\[
F_{av}(\mathcal{E}; U) = \frac{dF_{en}(\mathcal{E}; U) + 1}{d + 1}. \tag{76}
\]

Let us introduce two lemmas that will be used in the following Appendices.

**Lemma 31.** For any two bipartite pure states \( |\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B \) and \( |\phi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B' \) that satisfy
\[
\|\psi^A - \phi^A\|_1 \leq \epsilon, \tag{77}
\]
the following statements hold:

1) There exists a linear CPTP map \( \mathcal{T} : B \to B' \) that satisfies
\[
\|\mathcal{T}(|\psi\rangle|\psi\rangle) - |\phi\rangle|\phi\rangle\|_1 \leq 2\sqrt{\epsilon}. \tag{78}
\]

2) If \( \dim \mathcal{H}^A \leq \dim \mathcal{H}^B \), then there exists an isometry \( \tilde{W} : \mathcal{H}^B \to \mathcal{H}^{B'} \) that satisfies
\[
\|\tilde{W}|\psi\rangle|\psi\rangle|\tilde{W}^\dagger - |\phi\rangle|\phi\rangle\|_1 \leq 2\sqrt{\epsilon}. \tag{79}
\]

**Proof:** To prove 2), let \( |\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B \) be a purification of \( \psi^A \). Since all purifications are equivalent up to a local isometry, there exists an isometry \( W_1 : \mathcal{H}^B \to \mathcal{H}^{B'} \) that satisfies \( W_1|\psi\rangle = |\tilde{\psi}\rangle \). From (77) and (73), the states satisfy
\[
F(\tilde{\psi}^A, \tilde{\phi}^A) \geq 1 - \epsilon.
\]

Due to (62) and (65), there exists a unitary \( W_2 \) acting on \( \mathcal{H}^{B'} \) such that
\[
F(W_2|\tilde{\psi}\rangle, |\phi\rangle) \geq 1 - \epsilon.
\]

Using (75) once again, we obtain
\[
\|W_2W_1|\psi\rangle|\psi\rangle|W_1^\dagger W_2^\dagger - |\phi\rangle|\phi\rangle\|_1 \leq 2\sqrt{\epsilon},
\]
which implies (79) by \( \tilde{W} := W_2W_1 \).

To prove 1), let \( B'' \) be an ancillary system such that
\[
\dim \mathcal{H}^{B''} \leq \dim \mathcal{H}^B \times \dim \mathcal{H}^{B''}.
\]

Due to 2), for any \( |\varphi\rangle \in \mathcal{H}^{B''} \) there exists an isometry \( \tilde{W} : \mathcal{H}^B \to \mathcal{H}^{B''} \otimes \mathcal{H}^{B''} \) that satisfies
\[
\|\tilde{W}|\psi\rangle|\psi\rangle|\tilde{W}^\dagger - |\varphi\rangle|\varphi\rangle\|_1 \leq 2\sqrt{\epsilon}. \tag{80}
\]

Define a linear CPTP map \( \mathcal{T} : B \to B' \) by
\[
\mathcal{T} : \tau \to \text{Tr}_{B''}[\tilde{W}\tau\tilde{W}^\dagger].
\]

From (80) and (69), we obtain (78).

**Lemma 32.** For any pure states \( |\psi\rangle \in \mathcal{H}^A \), \( |\phi\rangle \in \mathcal{H}^B \) and any state \( \rho \in S(\mathcal{H}^A \otimes \mathcal{H}^B) \) satisfying
\[
1 - F(\rho^A, |\psi\rangle) \leq \epsilon, \tag{81}
\]
\[
1 - F(\rho^B, |\phi\rangle) \leq \epsilon, \tag{82}
\]
we have
\[
1 - F(\rho^{AB}, |\psi\rangle^A|\phi\rangle^B) \leq 6\sqrt{\epsilon}. \tag{83}
\]

**Proof:** Let \( |\chi_\rho\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B \otimes \mathcal{H}^E \) be a purification of \( \rho^{AB} \), and \( |\tau\rangle \in \mathcal{H}^E \) be an arbitrary pure state. Due to Equality (65), Condition (81) and the fact that \( |\chi_\rho\rangle \) is a purification of \( \rho^{AB} \), there exists a unitary \( W \) on \( \mathcal{H}^B \otimes \mathcal{H}^E \) such that
\[
F(|\chi_\rho\rangle, |\psi\rangle^A(W|\phi\rangle|\tau\rangle)^{BE}) \geq 1 - \epsilon,
\]
which implies
\[
\|\chi_\rho |\psi\rangle^A |\phi\rangle^B W^\dagger (|\phi\rangle^B \otimes |\tau\rangle^E W^\dagger)|^{BE} \|_1 \leq 2\sqrt{\epsilon}
\]
due to (73). From the monotonicity of the trace distance under the partial trace (68), we have
\[
\|\rho^{AB} - |\psi\rangle^A \otimes |\tilde{\phi}^B\rangle\|_1 \leq 2\sqrt{\epsilon}, \quad \|\rho^B - |\tilde{\phi}\rangle_1\|_1 \leq 2\sqrt{\epsilon},
\]
where we defined \( \tilde{\phi} := \tri_{BE}[W(|\phi\rangle^B \otimes |\tau\rangle^E W^\dagger)] \). The triangle inequality (59) and (61) yield
\[
\|\rho^{AB} - |\psi\rangle^A \otimes |\phi\rangle^B\|_1 \leq \|\rho^{AB} - |\psi\rangle^A \otimes |\tilde{\phi}^B\rangle\|_1
\]
\[
+ \||\psi\rangle^A \otimes |\phi\rangle^B - |\psi\rangle^A \otimes |\phi\rangle^B\|_1 \leq 2\sqrt{\epsilon} + \|\tilde{\phi}^B - |\phi\rangle\|_1
\]
\[
\leq 2\sqrt{\epsilon} + \|\tilde{\phi}^B - |\phi\rangle\|_1 + \|\rho^B - |\phi\rangle\|_1 \leq 2\sqrt{\epsilon} + 2\sqrt{\epsilon} + 2\sqrt{\epsilon} = 6\sqrt{\epsilon},
\]
where we used Condition (82) and (73) in the last inequality. Using (73) once again, we obtain (83).
**B. Gentle measurement lemma**

The gentle measurement lemma (Lemma 9.4.1 in [32]) states that for any \( \rho \in S(\mathcal{H}) \), \( X, Y \in \mathcal{L}(\mathcal{H}) \) and \( \epsilon \geq 0 \) such that \( 0 \leq X \leq I \) and \( \text{Tr}[\rho X] \geq 1 - \epsilon \), we have

\[
\left\| \rho - \frac{\sqrt{X} \rho \sqrt{X}}{\text{Tr}[\rho X]} \right\|_1 \leq 2\sqrt{\epsilon}.
\]

Let us introduce extensions of the gentle measurement lemma. Although similar lemmas have been used in the literature, we provide rigorous proofs for completeness.

**Lemma 33** For any \( \rho \in S(\mathcal{H}) \), \( X, Y \in \mathcal{L}(\mathcal{H}) \) and \( \epsilon \in [0,1] \) such that

\[
0 \leq X \leq I, \ 0 \leq Y \leq I
\]

and

\[
\text{Tr}[\rho X] \geq 1 - \epsilon, \ \text{Tr}[\rho Y] \geq 1 - \epsilon,
\]

define

\[
D_{XY} := \text{Tr}[\sqrt{Y} \sqrt{X} \rho \sqrt{X} \sqrt{Y}], \quad \rho_{XY} := \frac{\sqrt{Y} \sqrt{X} \rho \sqrt{X} \sqrt{Y}}{D_{XY}}.
\]

Then we have

\[
D_{XY} \geq 1 - 2\sqrt{\epsilon}, \ \|\rho - \rho_{XY}\|_1 \leq 5\sqrt{\epsilon}.
\]

**Proof:** Define

\[
\rho_X := \frac{\sqrt{X} \rho \sqrt{X}}{\text{Tr}[\sqrt{X} \rho \sqrt{X}]}
\]

Due to the gentle measurement lemma, Condition (84) implies

\[
\|\rho - \rho_X\|_1 \leq 2\sqrt{\epsilon}.
\]

Consequently, we have

\[
D_{XY} = \text{Tr}[\rho_X Y] \\
\geq \text{Tr}[\rho Y] - \frac{1}{2} \|\rho - \rho_X\|_1 \\
\geq 1 - \epsilon - \sqrt{\epsilon} \geq 1 - 2\sqrt{\epsilon},
\]

where the second line follows from (85), which leads to

\[
\|\rho_X - \rho_{XY}\|_1 \leq 2\sqrt{2\sqrt{\epsilon}}
\]

by (84). Thus we obtain

\[
\|\rho - \rho_{XY}\|_1 \leq \|\rho - \rho_X\|_1 + \|\rho_X - \rho_{XY}\|_1 \\
\leq 2\sqrt{\epsilon} + 2\sqrt{2\sqrt{\epsilon}} \leq 5\sqrt{\epsilon}.
\]

**Lemma 34** Suppose \( \rho, \sigma \in S(\mathcal{H}) \) satisfy \( \|\rho - \sigma\|_1 \leq \epsilon \). Let \( \{\lambda_i^\dagger\}_{i=1}^d \) be the eigenvalues of \( \sigma \) sorted in decreasing order, where \( d := \text{dim}\mathcal{H} \), and let

\[
\sigma = \sum_{i=1}^d \lambda_i^\dagger |i\rangle\langle i|
\]

be the eigen decomposition of \( \sigma \). Define a projection operator

\[
\tilde{\Pi} := \sum_{i=1}^{\text{rank}[\rho]} |i\rangle\langle i|.
\]

Then we have

\[
\text{Tr}[\tilde{\Pi}\sigma] \geq 1 - \epsilon.
\]

**Proof:** Let \( \Pi_\rho \) be the projection onto \( \text{supp}[\rho] \), and \( \Pi_\rho^\perp \) be that onto its orthogonal complement (i.e., \( \Pi_\rho^\perp = I - \Pi_\rho \)). Then we have

\[
\epsilon \geq \|\rho - \sigma\|_1 \geq \|\rho - \Pi_\rho \sigma \Pi_\rho - \Pi_\rho^\perp \sigma \Pi_\rho^\perp\|_1
\]

\[
= \|\rho - \Pi_\rho \sigma \Pi_\rho\|_1 + \text{Tr}[\Pi_\rho^\perp \sigma] \geq 1 - \text{Tr}[\Pi_\rho\sigma], \tag{85}
\]

where the second inequality follows from the monotonicity of the trace distance under a linear CPTP map defined by

\[
\tau \rightarrow \Pi_\rho \tau \Pi_\rho + \Pi_\rho^\perp \tau \Pi_\rho^\perp.
\]

We also have

\[
\text{Tr}[\tilde{\Pi}\sigma] = \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\dagger
\]

\[
= \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\dagger \langle i| \Pi_\rho |i\rangle + \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\dagger \langle i| \Pi_\rho^\perp |i\rangle
\]

\[
\geq \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\dagger \langle i| \Pi_\rho |i\rangle + \lambda_i^\dagger \sum_{i=1}^{\text{rank}[\rho]} \langle i| \Pi_\rho |i\rangle
\]

\[
= \sum_{i=1}^{\text{rank}[\rho]} \lambda_i^\dagger \langle i| \Pi_\rho |i\rangle + \lambda_i^\dagger \sum_{i=\text{rank}[\rho]+1}^{d} \langle i| \Pi_\rho |i\rangle
\]

\[
\geq \sum_{i=1}^{d} \lambda_i^\dagger \langle i| \Pi_\rho |i\rangle = \text{Tr}[\Pi_\rho \sigma], \tag{86}
\]

where the fourth line follows due to

\[
\sum_{i=1}^{d} \langle i| \Pi_\rho |i\rangle = \text{Tr}[\Pi_\rho] = \text{rank}[\rho].
\]

From (85) and (86), we obtain

\[
\text{Tr}[\tilde{\Pi}\sigma] \geq \text{Tr}[\Pi_\rho \sigma] \geq 1 - \epsilon.
\]

**C. Continuity of Quantum Entropies**

Define

\[
\eta_0(x) := \begin{cases} 
-x \log x & (x \leq 1/e) \\
1/e & (x \geq 1/e)
\end{cases}
\]

and \( \eta(x) = x + \eta_0(x) \), where \( e \) is the base of the natural logarithm. Define also

\[
h(x) := -x \log x - (1 - x) \log (1 - x).
\]

For two states \( \rho \) and \( \sigma \) in a \( d \)-dimensional quantum system (\( d < \infty \)) such that \( \|\rho - \sigma\|_1 \leq \epsilon \), we have

\[
|S(\rho) - S(\sigma)| \leq \epsilon \log d + \eta_0(\epsilon), \tag{87}
\]

where \( S(\rho) \) is the von Neumann entropy of \( \rho \).
which is called the Fannes inequality [34]. A simple calculation then yields
\[ |S(\rho) - S(\sigma)| \leq \eta(\epsilon) \log d. \] (88)
For two bipartite states \( \rho, \sigma \in S(\mathcal{H}^A \otimes \mathcal{H}^B) \) such that \( \|\rho - \sigma\|_1 \leq \epsilon < 1 \), we have
\[ |S(A|B)_\rho - S(A|B)_\sigma| \leq 4\epsilon \log d_A + 2h(\epsilon). \]
This inequality is called the Alicki-Fannes inequality [35], and leads to
\[ |S(A|B)_\rho - S(A|B)_\sigma| \leq 4\epsilon(\epsilon) \log d_A. \] (89)
Note that the upper bound in (89) does not depend on \( d_B \). As a consequence, we have
\[ |I(A : B)_\rho - I(A : B)_\sigma| \leq |S(A)_\rho - S(A)_\sigma| + |S(A|B)_\rho - S(A|B)_\sigma| \leq 5\epsilon(\epsilon) \log d_A. \] (90)
The following lemma will be used for evaluating average errors.

**Lemma 35** Let \( \epsilon \in (0, \infty) \) be a constant, \( f : [0, \epsilon] \to \mathbb{R} \) be a monotonically nondecreasing function that satisfies \( f(\epsilon) < \infty \), and \( \{p_k\}_{k \in \mathbb{K}} \) be a probability distribution on a countable set \( \mathbb{K} \). Suppose \( \epsilon_k \in [0, \epsilon] \), and \( \sum_{k \in \mathbb{K}} p_k \epsilon_k \leq \epsilon \) for a given \( \epsilon \in (0, \epsilon] \). Then we have
\[ \sum_{k \in \mathbb{K}} p_k f(\epsilon_k) \leq f(\sqrt{\epsilon}) + f(\epsilon) \cdot \sqrt{\epsilon}. \] (91)

**Proof:** Define \( \mathbb{K}(\lambda) := \{k \in \mathbb{K} \mid \epsilon_k \leq \lambda\} \). For any \( t > 0 \), we have
\[ \sum_{k \in \mathbb{K} \setminus \mathbb{K}(\epsilon t)} p_k \leq \frac{1}{\epsilon t} \sum_{k \in \mathbb{K} \setminus \mathbb{K}(\epsilon t)} p_k \epsilon_k \leq \frac{1}{\epsilon t} \sum_{k \in \mathbb{K} \setminus \mathbb{K}(\epsilon t)} p_k \epsilon_k \leq \frac{1}{t}, \]
and consequently,
\[ \sum_{k \in \mathbb{K}} p_k f(\epsilon_k) = \sum_{k \in \mathbb{K}(\epsilon t)} p_k f(\epsilon_k) + \sum_{k \in \mathbb{K} \setminus \mathbb{K}(\epsilon t)} p_k f(\epsilon_k) \leq f(t \epsilon) + f(\epsilon) \cdot \sqrt{\epsilon}. \]
Choosing \( t = 1/\sqrt{\epsilon} \), we obtain (91). \( \Box \)

### D. Entanglement Measures

A function \( \hat{E} : S(\mathcal{H}^A \otimes \mathcal{H}^B) \to [0, \infty) \) is called an entanglement measure if it satisfies the following three properties [26]:

1. If \( \rho \) is a pure state on \( AB \), then \( \hat{E}(\rho) = S(\rho^A) \).
2. If \( \rho \) is a separable state on \( AB \), then \( \hat{E}(\rho) = 0 \).
3. \( \hat{E}(\rho) \) does not increase on average under LOCC, i.e., if an ensemble \( \{\rho_i, p_i\} \) (\( \rho_i \in S(\mathcal{H}^A \otimes \mathcal{H}^B) \)) is obtained from \( \rho \in S(\mathcal{H}^A \otimes \mathcal{H}^B) \) by an LOCC transformation between \( A \) and \( B \), then \( \sum_i p_i \hat{E}(\rho_i) \leq \hat{E}(\rho) \).

An entanglement measure \( \hat{E} \) is said to be asymptotically continuous, if there exists an \( n \)-independent nonnegative function \( \delta(\epsilon) \) that satisfies \( \lim_{n \to \infty} \delta(\epsilon) = 0 \), and it holds that
\[ \left| \hat{E}(\rho) - \hat{E}(\sigma) \right| \leq \delta(\epsilon) \log d_A d_B \]
for all \( \rho, \sigma \in S(\mathcal{H}^A \otimes \mathcal{H}^B) \) satisfying \( \|\rho - \sigma\|_1 \leq \epsilon \). Examples of asymptotically continuous entanglement measures are entanglement of formation [36, 37], the relative entropy of entanglement [38–40] and squashed entanglement [35, 41].

### Appendix B

**Equivalence of Conditions (4) and (5)**

We prove the equivalence of Conditions (4) and (5) by showing that
\[ \frac{d^n}{36(d^2n + 1)} (1 - F_{en}(M_n, K_n, L_n; U^{\otimes n}))^2 \leq 1 - F_{av}(M_n, K_n, L_n; U^{\otimes n}) \leq 8 \sqrt{1 - F_{en}(M_n, K_n, L_n; U^{\otimes n})}. \] (92)
(93)
In the following, we prove these inequalities for the case of \( n = 1 \). It is straightforward to generalize the proof for an arbitrary \( n \).

Define states \( \rho(M_1) \) and \( \rho(M_1, \psi) \) by
\[ \rho(M_1) := M_1(\Phi_d^{AB, K_1} A | \Phi_d^{AB, K_1} A B_1), \]
\[ \rho(M_1, \psi) := M_1(\psi^{AB, K_1} A B_1), \]
and define quantum operations \( M_1 : AB \to A B \) and \( M_1 : AB \to A_1 B_1 \) by
\[ \tilde{M}_1(\tau) = Tr_{A_1 B_1}[M_1(\tau^{AB} \Phi_{K_1}^{A B})], \]
\[ \tilde{M}_1(\tau) = Tr_{AB}[M_1(\tau^{AB} \Phi_{K_1}^{A B})], \]
respectively. Using (64) and
\[ \int_{\text{Haar}} p(d\psi) |\psi\rangle \langle \psi| = \pi_d^{A} \otimes \pi_d^{B}, \]
where the integral is taken with respect to the Haar measure on \( H^A \otimes H^B \), it is straightforward to verify that
\[ \int_{\text{Haar}} p(d\psi) F(\rho(M_1, \psi) A_1 B_1, | \Phi_{L_1} \rangle) = \int_{\text{Haar}} p(d\psi) F(M_1(\pi_d^{A} \otimes \pi_d^{B}), | \Phi_{L_1} \rangle), \]
\[ = F(M_1(\pi_d^{A} \otimes \pi_d^{B}), | \Phi_{L_1} \rangle) = F(\rho(M_1) A_1 B_1, | \Phi_{L_1} \rangle). \] (95)
From (74), (70), and (1), we have
\[ F_{av}(\tilde{M}_1; U) = \int_{\text{Haar}} p(d\psi) F(\tilde{M}_1(\psi), U | \psi\rangle) \]
\[ = \int_{\text{Haar}} p(d\psi) F(\rho(M_1, \psi) A B, U | \psi\rangle) \]
\[ \geq \int_{\text{Haar}} p(d\psi) F(\rho(M_1, \psi), (U | \psi\rangle)^{A B} | \Phi_{L_1} \rangle_{A_1 B_1}) \]
\[ = F_{av}(M_1, K_1, L_1; U). \] (97)
and
\[
\int_{\text{Haar}} p(d\psi) F(\rho(\mathcal{M}_1, \psi)^A B_1, |\Phi_{L_1}|) \\
\geq \int_{\text{Haar}} p(d\psi) F(\rho(\mathcal{M}_1, \psi), (U|\psi\rangle)^A B_1 |\Phi_{L_1}) \\
= F_{\text{av}}(\mathcal{M}_1, K_1, L_1; U).
\]  
(98)

Similarly, from (75), (70) and (3), we have
\[
F_{\text{en}}(\mathcal{M}_1; U) \\
= F(\rho(\mathcal{M}_1)^A B_1, |\Phi_{L_1}|) \\
\geq F(\rho(\mathcal{M}_1), |\Psi_U\rangle)^A B_1 |\Phi_{L_1}) \\
= F_{\text{en}}(\mathcal{M}_1, K_1, L_1; U).
\]  
(99)

and
\[
F(\rho(\mathcal{M}_1)^A B_1, |\Phi_{L_1}|) \\
\geq F(\rho(\mathcal{M}_1), |\Psi_U\rangle)^A B_1 |\Phi_{L_1}) \\
= F_{\text{en}}(\mathcal{M}_1, K_1, L_1; U).
\]  
(100)

Due to (76), we also have
\[
1 - F_{\text{en}}(\mathcal{M}_1; U) = \left(1 + \frac{1}{d^2}\right) (1 - F_{\text{av}}(\mathcal{M}_1; U)) \\
\geq 1 - F_{\text{av}}(\mathcal{M}_1; U).
\]  
(101)

Suppose we have
\[
\epsilon = 1 - F_{\text{av}}(\mathcal{M}_1, K_1, L_1; U).
\]
From (97), (102) and (99), we have
\[
\left(1 + \frac{1}{d^2}\right) \epsilon \geq \left(1 + \frac{1}{d^2}\right) (1 - F_{\text{av}}(\mathcal{M}_1; U)) \\
= 1 - F_{\text{en}}(\mathcal{M}_1; U) \\
= 1 - F(\rho(\mathcal{M}_1)^A B_1, |\Phi_{L_1}|).
\]
(102)

Due to (98) and (99), we also have
\[
\epsilon \geq 1 - F(\rho(\mathcal{M}_1)^A B_1, |\Phi_{L_1}|).
\]
(103)

Therefore, from (3) and Lemma 32, we obtain that
\[
1 - F_{\text{en}}(\mathcal{M}_1, K_1, L_1; U) \\
= 1 - F(\rho(\mathcal{M}_1), |\Psi_U\rangle)^A B_1 |\Phi_{L_1}) \\
\leq 6\sqrt{\epsilon} \sqrt{1 + \frac{1}{d^2}},
\]
which implies (22) for \(n = 1\).

To prove Inequality (93), let us define indicator functions \(\theta_1\) and \(\theta_2\) by
\[
\theta_1(\psi|x) := \begin{cases} 1 & \text{if } 1 - F(\rho(\mathcal{M}_1, \psi)^A B_1, U|\psi\rangle) \leq x \\ 0 & \text{otherwise} \end{cases}
\]
and
\[
\theta_2(\psi|x) := \begin{cases} 1 & \text{if } 1 - F(\rho(\mathcal{M}_1, \psi)^A B_1, |\Phi_{L_1}|) \leq x \\ 0 & \text{otherwise} \end{cases}
\]
respectively, for any positive number \(x\). It follows from Lemma 32 that for any \(\psi\) satisfying \(\theta_1(\psi|x) = \theta_2(\psi|x) = 1\), we have
\[
1 - F(\rho(\mathcal{M}_1, \psi), (U|\psi\rangle)^A B_1 |\Phi_{L_1}) \leq 6\sqrt{x}.
\]
Thus we obtain
\[
1 - F_{\text{av}}(\mathcal{M}_1, K_1, L_1; U) \\
= \int_{\text{Haar}} p(d\psi) \left(1 - F(\rho(\mathcal{M}_1, \psi), (U|\psi\rangle)^A B_1 |\Phi_{L_1})\right) \\
= \int_{\text{Haar}} p(d\psi) \theta_1(\psi|x)\theta_2(\psi|x) \\
\times \left(1 - F(\rho(\mathcal{M}_1, \psi), (U|\psi\rangle)^A B_1 |\Phi_{L_1})\right) \\
+ \int_{\text{Haar}} p(d\psi) (1 - \theta_1(\psi|x)\theta_2(\psi|x)) \\
\times \left(1 - F(\rho(\mathcal{M}_1, \psi), (U|\psi\rangle)^A B_1 |\Phi_{L_1})\right) \\
\leq 6\sqrt{x} \int_{\text{Haar}} p(d\psi) \theta_1(\psi|x)\theta_2(\psi|x) \\
+ \int_{\text{Haar}} p(d\psi) (1 - \theta_1(\psi|x)\theta_2(\psi|x)) \\
\leq 6\sqrt{x} + \int_{\text{Haar}} p(d\psi) (1 - \theta_1(\psi|x)\theta_2(\psi|x))
\]  
(104)

for any \(x > 0\).

Suppose that we have
\[
\epsilon = 1 - F_{\text{en}}(\mathcal{M}_1, K_1, L_1; U).
\]  
From (100), (103) and (96), we have
\[
\epsilon \geq 1 - F_{\text{en}}(\mathcal{M}_1; U) \\
\geq 1 - F_{\text{av}}(\mathcal{M}_1; U) \\
= \int_{\text{Haar}} p(d\psi) \left(1 - F(\rho(\mathcal{M}_1, \psi)^A B_1, U|\psi\rangle)\right).
\]  
(105)

Due to (101) and (95), we also have
\[
\epsilon \geq 1 - F(\rho(\mathcal{M}_1)^A B_1, |\Phi_{L_1}|) \\
= \int_{\text{Haar}} p(d\psi) \left(1 - F(\rho(\mathcal{M}_1, \psi)^A B_1, |\Phi_{L_1}|)\right).
\]  
(106)

It follows from Inequality (105) that
\[
\epsilon \geq \int_{\text{Haar}} p(d\psi) \left(1 - \theta_1(\psi|x)\right) \left(1 - F(\rho(\mathcal{M}_1, \psi)^A B_1, U|\psi\rangle)\right) \\
\geq x \int_{\text{Haar}} p(d\psi) \left(1 - \theta_1(\psi|x)\right),
\]
and from (106) that
\[
\epsilon \geq \int_{\text{Haar}} p(d\psi) \left(1 - \theta_2(\psi|x)\right) \left(1 - F(\rho(\mathcal{M}_1, \psi)^A B_1, U|\psi\rangle)\right) \\
\geq x \int_{\text{Haar}} p(d\psi) \left(1 - \theta_2(\psi|x)\right).
Thus we obtain
\[
\frac{2x}{x} \geq \int_{\text{Haar}} p(d\psi)(1 - \theta_1(\psi|x)) + \int_{\text{Haar}} p(d\psi)(1 - \theta_2(\psi|x)) \\
\geq \int_{\text{Haar}} p(d\psi) (1 - \theta_1(\psi|x) + \theta_1(\psi|x))(1 - \theta_2(\psi|x)) \\
= \int_{\text{Haar}} p(d\psi) (1 - \theta_1(\psi|x)\theta_2(\psi|x)).
\]
Combining this with [104] and substituting \( x = \epsilon^2/3 \), we arrive at
\[
1 - F_{\text{av}}(M_1, K_1, L_1; U) \leq 8 \sqrt{\epsilon},
\]
which completes the proof of Inequality (93).

**APPENDIX C**

**PROOFS OF LEMMA 11 AND INEQUALITIES 22**

**A. Proof of Lemma 11**

Let \( V : AA_0 \rightarrow A' E_0 \) be an isometry such that the Naimark extension of \( \{M_k\}_{k \in \mathbb{K}} \) is given by \( M_k = |k\rangle \langle k| \otimes V \), and let
\[
\Psi' : A' E_0 : = \sum_{k \in \mathbb{K}} |k\rangle \langle k| E_0 ((\Psi^{\otimes n})^\dagger \otimes \rho^{AA_0}) V^\dagger |k\rangle \langle k| E_0.
\]
We have
\[
H(\{p_k\}_{k \in \mathbb{K}}) = S(\Psi') = S(A' E_0) = S(A'|E_0) = \sum_{k \in \mathbb{K}} p_k S(A') = \Delta S(A')_{\text{ave}},
\]
where the second line follows due to the von Neumann entropy nondecreasing under dephasing operations. Hence we obtain the first inequality in (19). The second inequality is due to \( \Delta S(G)_{\text{ave}} \geq 0 \), which follows from the concavity of the von Neumann entropy.

As for the third inequality, we first prove that there exists a nondecreasing function \( \tilde{n}(x) \), satisfying \( \lim_{x \rightarrow 0} \tilde{n}(x) = 0 \), such that
\[
\Delta S(A')_{\text{ave}} - \Delta S(G)_{\text{ave}} \geq I(\hat{B}C : G)_{\text{ave}} - n\tilde{n}(x) \log (d_B d_C).
\]
Define
\[
\epsilon_k := \| |(\Psi^{\otimes n})^\dagger \hat{B}C - \Psi_k^\dagger \hat{B}C \|_1.
\]
Using (88), we have
\[
\Delta S(A')_{k} := n S(A) + S(A_0) - S(A')_{\Psi_k} \\
= S(\hat{B}C)_{\Psi^{\otimes n}} + S(A_0) - S(\hat{B}C)_{\Psi_k} \\
\geq S(\hat{B}C)_{\Psi_k} + S(A_0) - S(\hat{B}C)_{\Psi_k} \\
- n\epsilon_k \log (d_B d_C) \\
= S(A_0) - S(G) - n\epsilon_k \log (d_B d_C) \\
= S(G) - n\epsilon_k, \\
= \Delta S(G) - n\epsilon_k \log (d_B d_C),
\]
where we defined \( \Delta S(G)_{k} := S(G) - S(A')_{\Psi_k} \). In the fifth line, we used the fact that \( \rho \) is a pure state on \( AA_0 G \). Averaging over \( k \), we obtain
\[
\Delta S(A')_{\text{ave}} - \Delta S(G)_{\text{ave}} \geq \sum_{k \in \mathbb{K}} p_k I(\hat{B}C : G)_{\Psi_k} - n\epsilon_k \log (d_B d_C).
\]
Applying Lemma 35 for \( c = 2 \) together with \( \epsilon_k \leq 2 \) and \( \sum_{k \in \mathbb{K}} p_k \epsilon_k \leq \epsilon < 1/8 < 2^2 \) yields
\[
\Delta S(A')_{\text{ave}} - \Delta S(G)_{\text{ave}} \geq I(\hat{B}C : G)_{\text{ave}} - n\tilde{n}(x) \log (d_B d_C),
\]
where we defined
\[
\tilde{n}(x) := \eta(x) + n\epsilon_k \log (d_B d_C).
\]
Second we prove that there exists a function \( \xi(x) \), satisfying \( \lim_{x \rightarrow 0} \xi(x) = 0 \), such that we have
\[
I(\hat{B}C : G)_{\text{ave}} \geq nM_{\text{R}_{AA_0}^{AB}}(\Psi^{AB}) - n\xi(nc) \log (d_A d_B d_C).
\]
This simply follows from the results in (13) (see Theorem 15 and Inequalities (69) and (71) therein). Defining
\[
\tilde{x}(x) := \xi(x) + \tilde{n}(x),
\]
and noting \( \tilde{n}(nc) \leq \tilde{n}(nc) \), we obtain the last inequality in (19).

**B. Proof of Inequalities 22**

The following theorem is essentially the same, but technically different from what is proved in [21]. We give a rigorous proof for completeness.

**Theorem 36** Let \( N \) be state merging of \( \Psi \) with error \( \epsilon \in (0, 1/4] \). Entanglement cost and classical communication cost of \( N \) are bounded below as
\[
\log K - \log L \geq S(B|A) - \eta'(2\sqrt{x}) \log (d_B d_L), \\
C \geq I(B : R) - 5\eta'(2\sqrt{x}) \log d_R,
\]
where
\[
\eta'(x) := \frac{5}{2} \eta(x) + \eta(2\sqrt{x}) + \eta(2\sqrt{x}).
\]

**Proof:** Without loss of generality, we assume that the protocol \( N \) consists of (i) Bob’s measurement described by \( \{N_{B_0}^{B_0 \rightarrow B_1}\} \), (ii) communication of \( l \) from Bob to Alice, and (iii) Alice’s operation described by a CPTP map \( O_l : AA_0 \rightarrow AA_0 A_1 \). The final state is given by
\[
\rho(N) = \sum_l p_l \hat{O}_l,
\]
where
\[
p_l := \| N_l |\Psi_K \rangle \rangle^2, \quad |\Psi_l \rangle := \rho^{-1/2} N_l |\Psi \rangle \Phi_K \rangle
\]
and \( \hat{O}_l := O_l(|\Psi_l \rangle \rangle^2 \Phi_	ext{L}^{A_1 B_1}) \). From (20) and (73), we have
\[
\| \rho(N) - \Psi^{AA_0} \otimes \Phi_L^{A_1 B_1} \|_1 \leq 2\sqrt{\epsilon}.
\]
Define
\[ \epsilon_l := \| \hat{\Psi} - \Psi^A_{ABR} \otimes \Phi_L^{A_1B_1} \|_1 \] (112)
and
\[ f_l := F(\hat{\Psi}_l, \Psi^A_{ABR} \otimes \Phi_L^{A_1B_1}) \]
for each \( l \). Due to the convexity of the square function, Inequality (73), Equalities (64), (110) and Inequality (20), we have
\[ \left( \sum_p p_l \epsilon_l^2 \right) \leq \sum_l p_l \epsilon_l^2 \leq 4 \sum_l p_l (1 - f_l) = 4 - 4F(\rho(N), \| \Psi^A_{ABR} \otimes \Phi_L^{A_1B_1} \|) \leq 4\epsilon, \]
which yields
\[ \sum_l p_l \epsilon_l \leq 2\sqrt{\epsilon}. \]

Consider the following protocol, which is as a whole equivalent to the protocol described above.

1) Bob performs a CPTP map \( \mathcal{E}_1 : BB_0 \rightarrow B_1C \) defined by \( \mathcal{E}_1(\tau) = \sum_p |l\rangle\langle l| \otimes \tau_{lN}^{C}. \) The state after this operation is \( \hat{\Psi}_l = \sum_p p_l |l\rangle\langle l| \otimes \hat{\Psi}_l^{AA_0B_1R}. \)

2) Bob transmits system \( C \) to Alice.

3) Alice performs a CPTP map \( \mathcal{E}_2 : CA_0 \rightarrow A_0A_1 \) defined as \( \mathcal{E}_2(\tau) = \sum_p \mathcal{O}_l |l\rangle\langle l| \otimes \tau_{lC}^{AA_0}. \) The state after the operation is \( \mathcal{E}_2(\hat{\Psi}_l) = \rho(N). \)

By the chain rule and the data processing inequality, we have
\[ 2S(A) + 2\log K = I(AA_0 : BB_0) + I(AA_0 : C) \leq I(AA_0 : B_1) \leq I(AA_0 : B_1R). \]

Due to Inequality (111) and (90), we also have
\[ I(AA_0 : B_1R) \rho(N) \leq I(AA_0 : B_1R) \psi + S(B_1R) \psi^L \log (d_R L) = S(R) \psi^L + S(AR) \psi - S(A) \psi \log (d_R L). \]

From (111), (112), (90) and Lemma 35, we obtain
\[ \sum_p p_l [\eta(\epsilon_l) + \eta(2\sqrt{\epsilon})] \log (d_R L) \leq 2 \sum_p p_l [\eta(2\sqrt{\epsilon}) + \eta(2 \cdot \sqrt{\epsilon})] \log (d_R L). \]

From (113), (114) and (115), we obtain
\[ \log K - \log L \geq S(R) \psi - S(A) \psi - \eta(2\sqrt{\epsilon}) \log (d_R L) \]
for the entanglement cost. As for the classical communication cost, from (111) and (90), we have
\[ 2S(R) \psi = (AA_0 : R) \psi \leq (AA_0 : R) \rho(N) + 5\eta(2\sqrt{\epsilon}) \log (d_R L) \leq (AA_0 : R) \psi^L + S(C) \psi^L - \eta(2\sqrt{\epsilon}) \log (d_R L). \]

Here, the fifth line follows from the fact that Bob’s measurement does not change the average reduced state of \( AA_0R. \) Thus we obtain
\[ \sum \mathcal{O}_l |l\rangle\langle l| \otimes \tau_{lC}^{AA_0}. \]

which concludes the proof.

**Appendix D**

**Proof of Lemma 18, 19 and 20**

**A. Proof of Lemma 18**

We prove that an \( M \)-induced map is \( (3\kappa + 2\nu) \)-decoupling between \( A' R_A \) and \( R_B \) if it is \( \zeta \)-oblivious and \( \nu \)-Markovianizing from \( R_A B R_B \), which implies Lemma 18.

Due to Equalities (61), (71), (28) and (24), we have
\[ \| \Phi^{R_A}_{M} - \pi^{R_A}_{d} \|_1 \]
\[ = \| \Phi^{R_A}_{M} \otimes \Phi^{B R_B}_{d} - \pi^{R_A}_{d} \otimes \Phi^{B R_B}_{d} \|_1 \]
\[ = \| \tilde{U}^{R A}_{M} R_B (\Phi^{R_A}_{M} \otimes \Phi^{B R_B}_{d}) \tilde{U}^{R A}_{M} R_B \|_1 \]
\[ = \| \Phi^{R A B R_B}_{M} - \Phi^{R A B R_B}_{d} \|_1. \]

Thus the condition of \( \zeta \)-obliviousness is equivalent to
\[ \| \Phi^{R A B R_B}_{M} - \Phi^{R A B R_B}_{d} \|_1 \leq \zeta, \]
which implies
\[ \| \Phi^{R A B R_B}_{M} - \pi^{R_A}_{d} \otimes \pi^{R_B}_{d} \|_1 \leq \zeta \]
and
\[ \left\| \Psi^R_M - \pi^{R_A}_d \right\|_1 \leq \varsigma, \quad \left\| \Psi^B_M - \pi^{R_B}_d \right\|_1 \leq \varsigma, \]  
(116)
due to (69). From (59) and (61), it follows that
\[ \left\| \Psi^R_M \otimes \Psi^R_M - \pi^{R_A}_d \otimes \pi^{R_B}_d \right\|_1 \leq \xi, \]
\[ \left\| \Psi^R_M \otimes \Psi^R_M - \pi^{R_A}_d \otimes \pi^{R_B}_d \right\|_1 \leq \xi, \]
\[ \left\| \Psi^R_M \otimes \Psi^R_M - \pi^{R_A}_d \otimes \pi^{R_B}_d \right\|_1 \leq \xi, \]
and
\[ \left\| \Psi^R_M \otimes \Psi^R_M - \pi^{R_A}_d \otimes \pi^{R_B}_d \right\|_1 \leq \xi, \]
which completes the proof.

Using the relation
\[ (\Psi^{RA}_{M})^2 \]
\[ = \left( U^{RA}_{M} \pi^{R_A}_d \otimes \Phi_0 \langle \Phi_d | B_{RB}^B \rangle U^{RA}_{M} \right)^2 \]
\[ = \frac{1}{\sqrt{d}} U^{RA}_{M} \pi^{R_A}_d \langle \Phi_d | B_{RB}^B \rangle U^{RA}_{M} \]

we have
\[ \Xi^{RA}_{M} (\Psi^{RA}_{M}) = \Psi^{RA}_{M}, \]
whence leads to
\[ \Psi^{A'R}_{M} (\Psi^{RA}_{M}) = \Xi^{RA}_{M} (\Psi^{RA}_{M}), \]
It is straightforward to verify that a map \( \mathcal{R}' : R_A \to R_AR_B \)
defined by
\[ \mathcal{R}'(\tau) = \Xi(\tau^{R_A} \otimes \pi^{R_B}_d) \quad (\forall \tau \in \mathcal{S}(\mathcal{H}^{R_A})) \]
is CPT as well. Therefore, from (68), (59) and (61), we have
\[ \left\| \Psi^{A'R}_{M} (\Psi^{RA}_{M}) - \mathcal{R}'(\Psi^{RA}_{M}) \right\|_1 \leq \varsigma + \mu, \]
where the last line follows from the assumption and (116).

C. Proof of Lemma 20

Suppose that an \( M_k \)-induced map is \( \mu_k \)-decoupling between
\( A'R_A \) if it is \( \varsigma \)-oblivious and \( \mu \)-decoupling between
\( A'R_A \) and \( R_B \), which implies Lemma 19. Let \( \Xi' : R_AR_B \to R_AR_B \)
be a linear CPT map defined by
\[ \Xi'(\tau) = d^2 \cdot (\Psi^{RA}_{M})^2 (\tau^{R_AR_B} \otimes IB) (\Psi^{RA}_{M})^2 \]
This is indeed CPT since we have
\[ \text{Tr}[\Xi'(\tau)] = d^2 \cdot \text{Tr}[\tau^{R_AR_B} (\Psi^{R_AR_B})] \]
\[ = d^2 \cdot \text{Tr}[\tau^{R_AR_B} (\pi^{R_A}_d \otimes \pi^{R_B}_d)] \]
\[ = \text{Tr}[\tau]. \]

B. Proof of Lemma 19

We prove that an \( M \)-induced map is \( (\varsigma + \mu) \)-Markovianizing
from \( A'R_A \) if it is \( \varsigma \)-oblivious and \( \mu \)-decoupling between
\( A'R_A \) and \( R_B \), which implies Lemma 19. Let \( \Xi : R_AR_B \to R_AR_B \)
be a linear CPT map defined by
\[ \Xi(\tau) = d^2 \cdot (\Psi^{RA}_{M})^2 (\tau^{R_AR_B} \otimes IB) (\Psi^{RA}_{M})^2 \]
This is indeed CPT since we have
\[ \text{Tr}[\Xi(\tau)] = d^2 \cdot \text{Tr}[\tau^{R_AR_B} (\Psi^{R_AR_B})] \]
\[ = d^2 \cdot \text{Tr}[\tau^{R_AR_B} (\pi^{R_A}_d \otimes \pi^{R_B}_d)] \]
\[ = \text{Tr}[\tau]. \]
and
\[ \left\| (\Psi')^{\hat{B}B} - (\Psi^p)^{\hat{B}B} \otimes \pi_d^{B} \right\|_1 \leq 2\sqrt{\mu} \] (123)
due to (69).

Let \( \{\lambda_i\}_{i=1}^{\dim \hat{B}} \) be the eigenvalues of \( (\Psi^p)^{\hat{B}} \) sorted in decreasing order, let
\[ (\Psi^p)^{\hat{B}} = \sum_{i=1}^{\dim \hat{B}} \lambda^i \ket{i}ket{\bar{i}} \]
be the eigen decomposition of \( (\Psi^p)^{\hat{B}} \), and define a linear operator \( \Pi \) on \( \mathcal{H}^{\hat{B}} \) by
\[ \Pi := \sum_{i=1}^{\dim \hat{B}A_0} \ket{i}ket{\bar{i}}. \]
Due to Lemma 34 and (123), we have
\[ \text{Tr}[\Pi(\Psi^p)^{\hat{B}}] = \text{Tr}[(\Pi \otimes I^B)((\Psi^p)^{\hat{B}} \otimes \pi_d^{B})] \geq 1 - 2\sqrt{\mu}. \]
(124)

Note that, by definition, we have
\[ \text{rank}[(\Psi')^{\hat{B}B}] \leq \dim B_0 \times \dim B. \]
Define
\[ |\tilde{\Psi}^p\rangle^{A'R_AB} := \frac{\Pi^{\hat{B}}|\Psi^p\rangle}{\|\Pi^{\hat{B}}|\Psi^p\rangle\|.} \]
From (124) and the gentle measurement lemma, we obtain
\[ \|\tilde{\Psi}^p|\Psi^p\rangle - |\Psi^p\rangle\|_1 \leq 2\sqrt{2\sqrt{\mu}}. \] (125)
From (69), (61), (122), (125) and \( \mu \in (0, 1] \), we see that
\[
\begin{align*}
\|\Psi'|\Psi'\rangle - (\Psi^p)^{A'R_AB} \otimes \phi_d^{BRB} & \|_1 \\
\leq \|\Psi'|\Psi'\rangle - (\Psi^p)^{A'R_AB} \otimes \phi_d^{BRB} & \|_1 \\
& + \|\Psi^p)^{A'R_AB} \otimes \phi_d^{BRB} & \|_1 \\
& \leq 2\sqrt{\mu} + 2\sqrt{2\sqrt{\mu}} \leq 2(1 + \sqrt{2})\sqrt{\mu} \\
& \leq 5\sqrt{\mu},
\end{align*}
\]
which implies (30). From (121), \( \Psi^k = \Psi^p \), (125) and (69), we also have
\[ \|\Psi^{kA} - \pi_d^{A} \|_1 \leq 2\sqrt{2\sqrt{\mu}} \leq 3\sqrt{\mu}. \]

\section*{APPENDIX E PROOFS OF LEMMA 22 AND 23

\subsection*{A. Settings}

As we described in Section VII, we assume without loss of generality that the protocol \( \mathcal{M} \) proceeds as follows. (See also Remark at the end of this appendix.)

I-1. Alice performs a measurement \( \{M_k^{AA_0} \rightarrow A'\}_{k \in K} \). The probability of obtaining measurement outcome \( k \) is given by \( p_k = \|M_k|\Psi\rangle\|_{\phi_{res}}\|^2 \), and the state after the measurement is \( |\Psi_k\rangle = p_k^{1/2} M_k|\Psi\rangle\|\phi_{res}\).
basis \( \{ |l \} \)l, and obtains outcome \( l \) with probability \( p_{l|k} \).
The state after the measurement is
\[
|\Psi_{kl}\rangle := F^{-1/2}|l\rangle B |\Psi_{k}\rangle.
\]
(127)

**II-5.** Bob communicates \( l \) to Alice.
**II-6.** Alice performs \( V_{kl} \). The state becomes
\[
|\hat{\Psi}_{kl}\rangle := A^{A_1A_2A_RB_{B_1}} := V_{kl} |\Psi_{kl}\rangle.
\]
(128)

**II-7.** Alice discards \( A_E \).

**Remark.** In the description of \( M \) by I-1~5, we assume that all information about the outcome of Alice’s measurement, represented by \( k \in \mathbb{K} \), is communicated to Bob. In a general protocol, however, not all information about the measurement outcome need to be communicated. In such cases, the measurement outcomes are represented as \((k_1, k_2) \in \mathbb{K}_1 \times \mathbb{K}_2\) by two countable sets \( \mathbb{K}_1 \) and \( \mathbb{K}_2 \). The \( \mathbb{K}_1 \) part of the outcome is communicated to Bob, whereas the \( \mathbb{K}_2 \) part is kept on Alice’s register until she performs the last operation. We show that such protocols can also be described by II-1~7 as follows. Let \( V : A \rightarrow A’A_1E_1, A_2E_2 \) be an isometry such that the Naimark extension of Alice’s measurement is given by \( M_{k_1k_2} := \langle k_1 | A_2E_2, k_2 | A_2E_2 \rangle \hat{V} \) and let \( V_{k_1k_2} : A’ \rightarrow A_1A_E \) be an isometry such that the Stinespring dilatation of Alice’s last operation is given by \( O_{k_1k_2}(\tau) = \text{Tr}_{A_E} [V_{k_1k_2}\tau V_{k_1k_2}^\dagger] \). The procedure II-1~7 then gives a description of the general protocol by the following correspondence:
\[
\begin{align*}
k &\rightarrow k_1 \\
A’ &\rightarrow A’A_1E_1 \\
M_k^{A_0A_2A_0A_2} &\rightarrow M_{k_1k_2}^{A_0A_2A_0A_2} = \langle k_1 | A_2E_2 \rangle \hat{V} \\
V_{kl} &\rightarrow V_{k_1k_2l} = \sum_{k_2 \in \mathbb{K}_2} | k_2 \rangle | A_2E_2 \rangle \otimes V_{k_1k_2l} \\
A_E &\rightarrow A_1E_1, A_2E_2.
\end{align*}
\]

**B. Proof of Lemma 22**

We prove that the measurement \( M \) is \( 4\sqrt{\epsilon} \)-oblivious and \( 8\sqrt{\epsilon} \)-decoupling between \( A’R_A \) and \( R_B \), which implies Lemma 22 combined with Lemma 19 from (25) and Equality (67), we have
\[
1 - \epsilon \leq \sum_{kl} p_{kl} F \left( |\hat{\Psi}_{kl}\rangle, |\Phi_{d}^{AR}A_B R_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B \right)
\]
\[
= \sum_{kl} p_{kl} F \left( |\hat{\Psi}_{kl}\rangle, |\Phi_{d}^{AR}A_B R_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B \right)
\]
for some states \( \phi_{kl} \), which leads to
\[
\sum_{kl} p_{kl} \epsilon_{kl} \leq \epsilon
\]
for
\[
\epsilon_{kl} := 1 - F \left( |\hat{\Psi}_{kl}\rangle, |\Phi_{d}^{AR}A_B R_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B \right).
\]
(129)
Due to Lemma 35 and \( \epsilon_{kl} \in [0, 1] \), we have
\[
\sum_{kl} p_{kl} \sqrt{\epsilon_{kl}} \leq \sqrt{\epsilon} + \sqrt{\epsilon} \leq 2\sqrt{\epsilon}.
\]
(130)
Therefore, by using (73), (69) and (60), we obtain
\[
4\sqrt{\epsilon} \geq 2 \sum_{kl} p_{kl} \sqrt{\epsilon_{kl}}
\]
\[
\geq \sum_{kl} p_{kl} \left| |\hat{\Psi}_{kl}\rangle - |\Phi_{d}^{AR}A_B R_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B \right|_1
\]
\[
\geq \sum_{kl} p_{kl} \left| |\hat{\Psi}_{kl}\rangle - |\Phi_{d}^{AR}A_B R_B A_1 A_E |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B \right|_1
\]
\[
= \sum_{kl} p_{kl} \left| |\Psi_{kl}^{AR}A_B R_B | V_{kl}^\dagger |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B \right|_1
\]
\[
\geq \sum_{kl} p_{kl} \left| |\Phi_{d}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1
\]
\[
= \sum_{kl} p_{kl} \left| |\Phi_{d}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1
\]
\[
\geq \sum_{kl} p_{kl} \left| |\Psi_{kl}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1.
\]
(131)
where we defined
\[
\psi_{k}^{AR} := \sum_{l} p_{kl} |V_{kl}^\dagger (|\Phi_{d}^{AR}A_B |\Phi_{d}^{AR}A_B \rangle V_{kl} \otimes \tau_{d}^{RB})|_1.
\]
(132)
Hence, from (59), (61) and (68), we obtain
\[
\sum_{k} p_{k} \left| |\Phi_{d}^{AR}A_B R_B | - \psi_{k}^{AR} | A_1 A_E \right|_1
\]
\[
\leq \sum_{k} p_{k} \left| \psi_{k}^{AR} | A_1 A_E \right|_1 + \sum_{k} p_{k} \left| \psi_{k}^{AR} \right|_1
\]
\[
\leq 3 \sum_{k} p_{k} \left| \psi_{k}^{AR} | A_1 A_E \right|_1 + \sum_{k} p_{k} \left| \psi_{k}^{AR} | A_1 A_E \right|_1
\]
\[
\leq 12 \sqrt{\epsilon},
\]
Thus Alice’s measurement is \( 12 \sqrt{\epsilon} \)-decoupling between \( A’R_A \) and \( R_B \). From (131), (69), (132), (71), (28), (29) and (61), we also have
\[
4\sqrt{\epsilon} \geq \sum_{kl} p_{kl} \left| |\Phi_{d}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1
\]
\[
= \sum_{kl} p_{kl} \left| |\Phi_{d}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1
\]
\[
= \sum_{kl} p_{kl} \left| |\Phi_{d}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1
\]
\[
= \sum_{kl} p_{kl} \left| |\Phi_{d}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1
\]
\[
= \sum_{kl} p_{kl} \left| |\Phi_{d}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1
\]
\[
= \sum_{kl} p_{kl} \left| |\Phi_{d}^{AR}A_B R_B | - |\Phi_{d}^{AR}A_B | A_1 A_E \right|_1
\]
which implies that Alice’s measurement is \( 4\sqrt{\epsilon} \)-oblivious. ■
C. Proof of Lemma 23

From (129), (70) and (128), we have

$$1 - \epsilon_{kl} \leq F(\tilde{\Psi}_{kl}^{BR_B}, \Phi_{d}^{BR_B}) = F(\Psi_{kl}^{BR_B}, \Phi_{d}^{BR_B}).$$

By using (64), we have

$$1 - \sum_{l} p_{l} |k| \epsilon_{kl} \leq F(\sum_{l} p_{l} |k| \tilde{\Psi}_{kl}^{BR_B}, \Phi_{d}^{BR_B}) = F(\tilde{\Psi}^{BR_B}, \Phi_{d}^{BR_B})$$

because of (127). Due to Equality (67), there exist pure states $|\Psi_{k}^{P}\rangle$ of $A' R_k B_1 B_E (k \in \mathcal{K})$ such that

$$1 - \sum_{l} p_{l} |k| \epsilon_{kl} \leq F(\Psi_{k}^{P}, (\Psi_{k}^{P})^{A' R_k A B_1 B_E} \otimes \Phi_{d}^{BR_B})$$

for each $k$, which leads to

$$\sqrt{2} \sum_{k} p_{k} \left| \Psi_{k}^{P} - (\Psi_{k}^{P})^{A' R_k A B_1 B_E} \otimes \Phi_{d}^{BR_B} \right|^{1}$$

due to (73). Thus we have

$$\sum_{k} p_{k} \left| \left| \Psi_{k}^{P} - (\Psi_{k}^{P})^{A' R_k A B_1 B_E} \otimes \Phi_{d}^{BR_B} \right|^{1} \leq 2 \sqrt{2} \epsilon,$$

where the last line follows from the concavity of the square root function and Inequality (130). This completes the proof of Lemma 23.

APPENDIX F

PROOF OF INEQUALITY (36)

In this Appendix, we describe an evaluation of the total error, which have appeared in Section VI-A in the proof of the direct part of Theorem 5.

From Lemma 21, we have

$$\left| \Psi^{P}(\Psi^{P})^{A^{2} A_{B} B} \otimes (\Phi_{d}^{\otimes n})^{B R_{B}} \right|^{1} \leq 5 \sqrt{2} \epsilon,$$

and

$$\left| \left| (\tilde{\Psi}_{d}^{P})^{R_{A}} - (\pi_{d}^{\otimes n})^{R_{A}} \right|^{1} \right| \leq 3 \sqrt{2} \epsilon,$$

(corresponding to (33) and (34), respectively. Let $A_{B'}$ be Alice’s register which is identical to $B'$, and $N': A' B' \rightarrow A' A_{B'}$ be state merging of $\left| \tilde{\Psi}_{d}^{P} \right|^{A' R_{A} B'}$. Define the merging error $\epsilon_{\text{merg}}$ by

$$\epsilon_{\text{merg}} := \left| \left| N'((\tilde{\Psi}_{d}^{P})^{A' R_{A} B'}) - (\tilde{\Psi}_{d}^{P})^{A' R_{A} A_{B'}} \right|^{1} \right|.$$  

From (133) and (135), we have

$$\left| \left| \Psi^{P}(\Psi^{P})^{A_{B} B} \otimes (\Phi_{d}^{\otimes n})^{B R_{B}} \right|^{1} \right| \leq 5 \sqrt{2} \epsilon,$$

which leads to

$$\left| \left| N'((\tilde{\Psi}_{d}^{P})^{A' R_{A} B'}) - (\tilde{\Psi}_{d}^{P})^{A' R_{A} A_{B'}} \right|^{1} \right| \leq 5 \sqrt{2} \epsilon,$$

By definition, we have $(\tilde{\Psi}_{d}^{P})^{R_{A}} = (\tilde{\Psi}_{d}^{P})^{R_{A}}$. Therefore, due to Inequality (134) and Lemma 31 there exists a quantum operation $O: A' A_{B'} \rightarrow A$ such that

$$\left| \left| O((\tilde{\Psi}_{d}^{P})^{A' R_{A} A_{B'}}) - (\tilde{\Psi}_{d}^{P})^{R_{A}} \right|^{1} \right| \leq 2 \sqrt{3} \sqrt{2} \epsilon.$$  

Define a quantum operation $N': A' B' \rightarrow A$ by $N' := O \circ N$. From (45), (136) and (138), we obtain

$$\left| \left| N'((\tilde{\Psi}_{d}^{P})^{A' R_{A} B'}) - N'((\tilde{\Psi}_{d}^{P})^{A' R_{A} A_{B'}}) \right|^{1} \leq \epsilon_{\text{merg}},$$

and

$$\left| \left| N'((\tilde{\Psi}_{d}^{P})^{A' R_{A} B'}) \otimes (\Phi_{d}^{\otimes n})^{B R_{B}} \right|^{1} \right| \leq 5 \sqrt{2} \epsilon.$$

From (59), (61), (137), (138) and (139), we see that

$$\left| \left| \right|\right| N'((\tilde{\Psi}_{d}^{P})^{A' R_{A} B'}) - (\tilde{\Psi}_{d}^{P})^{A' R_{A}} \otimes (\tilde{\Psi}_{d}^{P})^{B R_{B}} \right|^{1} \right| \right|\right| \leq \left| \left| \right| N'((\tilde{\Psi}_{d}^{P})^{A' R_{A} B'}) \otimes (\tilde{\Psi}_{d}^{P})^{B R_{B}} \right|^{1} \right| \leq 5 \sqrt{2} \epsilon.$$  

Due to (20), (21) and (73), we have

$$\epsilon_{\text{merg}} \leq 2 \left( 2 \sqrt{2} \sqrt{2} - \frac{1}{d} \right)^{2} \leq 4 \cdot 2^{-n} \epsilon.$$  

Thus we obtain (36).

APPENDIX G

PROOF OF THE CONVERSE PART

Fix arbitrary $n \in \mathbb{N}$ and $\epsilon > 0$ such that

$$16 \cdot 12 \sqrt{\epsilon} < n \leq \frac{1}{4 \cdot 12 \sqrt{\epsilon}},$$

and let $M_{n}$ be a $(2, n, \epsilon)$-protocol for implementing $U$ with the entanglement cost $n E$, the classical communication cost $n C_f$ and the backward classical communication cost $n C_b$. We assume here for simplicity that $K_n$ and $L_n$ is bounded above as

$$\log K_n, \log L_n \leq n \log \kappa$$

(141)
with a constant $\kappa > 0$. As we prove below, the following inequalities hold for any such $M_n$:
\begin{align}
C_f &\geq M(U^\dagger) - \xi(12\sqrt{e} \cdot n) \log d, \quad (142) \\
n \log d + \log K_n - \sum_{k \in K} pk S(A') \psi_k \\
&\geq nM(U^\dagger) - n\hat{\xi}(12\sqrt{e} \cdot n) \log d, \quad (143) \\
\frac{1}{n} (\log K_n - \log L_n) \\
&\geq M(U^\dagger) - \epsilon(12\sqrt{e} \cdot n) \log (dk), \quad (144) \\
C_b &\geq M(U^\dagger) - \epsilon(12\sqrt{e} \cdot n) \log (dk\cdot 2^{C_B}). \quad (145)
\end{align}

Here, $\hat{\xi}$ is a function defined by (108), and $\xi_1, \xi_2$ are nonnegative functions that are independent of $n$ and $d$, and satisfy $\lim_{x \to 0} \xi_1(x) = \lim_{x \to \infty} \xi_2(x) = 0$. The converse part of Theorem 2 immediately follows by substituting $\epsilon_n$ to $\epsilon$ in Inequalities (142), (143) and (145), and by taking the limit of $n \to \infty$. Note that Assumption (9) implies $\lim_{n \to \infty} 12\sqrt{e} \cdot n = 0$.

Let us prove Inequalities (142)–(145). From Lemma 22 Alice’s measurement in $M_n$ is $4\sqrt{e}$-oblivious and $12\sqrt{e}$-Markovianizing from $A'R_A$. Hence Conditions 1) and 2) in Definition 9 are satisfied by the correspondence described by (38). Thus we can apply Lemma 11 to obtain the above four inequalities.

Inequality (142) follows from $nC_f \geq H(\{p_k\}_{k \in K})$ and Inequality (19). Inequality (143) follows from Inequality (19) on $\Delta S(A'_{AV})$. We prove Inequalities (144) and (145) in the following subsections.

A. Proof of Inequality (144)

Let $\tilde{\mathcal{M}}_{n,k} : A'B_B \to \tilde{A}_A B_1 B_2$ be a CPTP map that describes the procedure 4–7, presented in Appendix E–A, averaged over the measurement outcome $l$. The final state is given by
\begin{equation}
\rho(\mathcal{M}_n, U^\dagger) = \sum_{k \in K} p_k \tilde{\mathcal{M}}_{n,k}(\psi_k^l). \quad (146)
\end{equation}

Define
\begin{equation}
\epsilon_{k,1} := \left\| \tilde{\mathcal{M}}_{n,k}(\psi_k^l) - \left( \Phi_d \otimes \Phi_l \right)^{\tilde{A}_A} \otimes \left( \Phi_d \otimes \Phi_l \right)^{B_B} \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1. \quad (147)
\end{equation}

and
\begin{equation}
f_{k,1} := F(\tilde{\mathcal{M}}_{n,k}(\psi_k^l), \left( \Phi_d \otimes \Phi_l \right)^{\tilde{A}_A} \otimes \left( \Phi_d \otimes \Phi_l \right)^{B_B} \otimes \Phi_{L_n}^{A_1 B_1})
\end{equation}

for $k \in K$. Due to the convexity of the square function, Inequality (71), Equalities (64), (146) and Inequality (37), we have
\begin{align}
\left( \sum_{k \in K} p_k \epsilon_{k,1} \right)^2 &\leq \sum_{k \in K} p_k^2 \epsilon_{k,1}^2 \leq 4 \sum_{k \in K} p_k (1 - f_{k,1}) \\
&= 4 - 4F(\rho(\mathcal{M}_n, U^\dagger), \left( \Phi_d \otimes \Phi_l \right)^{\tilde{A}_A} \otimes \left( \Phi_d \otimes \Phi_l \right)^{B_B} \otimes \Phi_{L_n}^{A_1 B_1}) \\
&\leq 4\epsilon,
\end{align}

which yields
\begin{equation}
\sum_{k \in K} p_k \epsilon_{k,1} \leq 2\sqrt{\epsilon}. \quad (148)
\end{equation}

From Lemma 23 there exist pure states $|\Psi_k^{p,h}\rangle^{\tilde{A}_A B_1 B_2}$ such that
\begin{equation}
\sum_k p_k \left\| \psi_k - (\Phi_d \otimes \Phi_l)^{\tilde{A}_A B_1 B_2} \otimes \Phi_{L_n}^B \right\|_1 \leq 4\sqrt{\epsilon}.
\end{equation}

Defining
\begin{equation}
\epsilon_{k,2} := \left\| \psi_k' - (\Phi_d \otimes \Phi_l)^{\tilde{A}_A B_1 B_2} \otimes \Phi_{L_n}^B \right\|_1, \quad (149)
\end{equation}

we obtain
\begin{equation}
\sum_{k \in K} p_k \epsilon_{k,2} \leq 4\sqrt{\epsilon}. \quad (150)
\end{equation}

From (148) and (150), we have
\begin{equation}
\sum_{k \in K} p_k \epsilon_k \leq 2\sqrt{\epsilon} + 4\sqrt{\epsilon} \leq 6\sqrt{\epsilon} \quad (151)
\end{equation}

for
\begin{equation}
\epsilon_k := \epsilon_{k,1} + \epsilon_{k,2}. \quad (152)
\end{equation}

Defining $K(\lambda) := \{k \in K | \epsilon_k \leq \lambda \}$, this leads to
\begin{equation}
\sum_{k \in K \setminus K(\lambda)} p_k \epsilon_k \leq \sum_{k \in K \setminus K(\lambda)} \frac{1}{\lambda} \sum_{k \in K \setminus K(\lambda)} p_k \epsilon_k \leq \frac{6\sqrt{\epsilon}}{\lambda} \quad (153)
\end{equation}

for any $\lambda > 0$.

From (149) and (68), we have
\begin{equation}
\left\| \tilde{\mathcal{M}}_{n,k}(\psi_k^l) - \tilde{\mathcal{M}}_{n,k}(\psi_k^{p,h})^{\tilde{A}_A B_1 B_2} \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1 \leq \epsilon_{k,2} \quad (154)
\end{equation}

for each $k$, which implies
\begin{equation}
\left\| (\psi_k^l)^{A'} - (\psi_k^{p,h})^{A'} \right\|_1 = \left\| (\psi_k^l)^{A'} - (\psi_k^{p,h})^{A'} \right\|_1 \leq \epsilon_{k,2} \leq \epsilon_k \quad (155)
\end{equation}

due to (69) and (126). By (61), (59), (154), (147) and (152), we see that
\begin{align}
&\left\| \tilde{\mathcal{M}}_{n,k}(\psi_k^l)^{\tilde{A}_A B_1 B_2} - \left( \Phi_d \otimes \Phi_l \right)^{\tilde{A}_A B_1 B_2} \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1 \\
&= \left\| \tilde{\mathcal{M}}_{n,k}(\psi_k^l)^{\tilde{A}_A B_1 B_2} - \left( \Phi_d \otimes \Phi_l \right)^{\tilde{A}_A B_1 B_2} \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1 \\
&\leq \left\| \tilde{\mathcal{M}}_{n,k}(\psi_k^l) - \tilde{\mathcal{M}}_{n,k}(\psi_k^{p,h})^{\tilde{A}_A B_1 B_2} \otimes \Phi_{L_n}^{A_1 B_1} \right\|_1 \\
&\leq \epsilon_k, \quad (156)
\end{align}

which leads to
\begin{equation}
\left\| (\psi_k^l)^{R_A} - (\pi_{\Phi_l}^{\tilde{A}_A} R_A) \right\|_1 \leq \epsilon_k. \quad (157)
\end{equation}

by (69). Therefore, due to Lemma 31 there exists a quantum operation $\mathcal{T}_{n,k} : A' \to A'B_B$, where $A'B_B$ is a quantum system which is identical to $B_B$, such that
\begin{equation}
\left\| (\psi_k^l)^{A'R_A A'B_B} - \mathcal{T}_{n,k}(\psi_k^{p,h}) \right\|_1 \leq 2\sqrt{\epsilon_k}. \quad (158)
\end{equation}

Define
\begin{equation}
\tilde{\mathcal{M}}_{n,k} := \mathcal{T}_{n,k} \circ \tilde{\mathcal{M}}_{n,k}. \quad (159)
\end{equation}
Owing to (59), (159), (68), (61) and Inequalities (156), (158), we have

\[
\left\| \hat{M}_{n,k} \left( \Psi_{k}^{p} \right) A \hat{R}_{A} A B_{1} B_{E} A_{1} B_{1} \right\|_{1} \leq \left\| \hat{M}_{n,k} \left( \Psi_{k}^{p} \right) A \hat{R}_{A} A B_{1} B_{E} A_{1} B_{1} \right\|_{1}
\]

Hence \( \hat{M}_{n,k} \) is a state merging of \( \Psi_{k}^{p} A \hat{R}_{A} (B_{1} B_{E}) \) with the error \( \epsilon_{k} + 2 \sqrt{\epsilon_{k}} \) and the entanglement cost \(-\log L_{n}\) (see Definition 12 and Inequality (73)).

Note that we have \( \epsilon_{k} + 2 \sqrt{\epsilon_{k}} \leq \sum_{n \in \mathbb{N}} \frac{\log \left( \psi_{n} \right) \log \left( \psi_{n} \right)}{\eta_{0}(\epsilon_{k})} \).

which corresponds to (26), and Assumption (141). Therefore, we apply Theorem 36 Inequalities (155), (157) and (87) to obtain

\[
\log L_{n} + \eta' \left( 2 \sqrt{\epsilon_{k} + 2 \sqrt{\epsilon_{k}}} \right) \log (d \kappa)
\]

for each \( k \) such that \( \epsilon_{k} + 2 \sqrt{\epsilon_{k}} \leq 1/4 \), where \( \eta' \) is a function defined by (109). Substituting \( \chi^{*} := (9 - 4 \sqrt{5})/4 \) to \( \lambda \) in (153), and noting that \( \epsilon_{k} + 2 \sqrt{\epsilon_{k}} \leq 1/4 \) if and only if \( \epsilon_{k} \in \mathbb{K}(\lambda^{*}) \), we also have

\[
\sum_{k: \epsilon_{k} + 2 \sqrt{\epsilon_{k}} > 1/4} p_{k} \leq \sum_{k \in \mathbb{K} \setminus \mathbb{K}(\lambda^{*})} p_{k} \leq \frac{4 \cdot 6 \sqrt{\tau}}{9 - 4 \sqrt{5}} < 432 \sqrt{\tau} (162)
\]

Defining

\[
\eta''(\epsilon) := \eta' \left( 2 \sqrt{\epsilon} + 2 \sqrt{\epsilon} \right) + 2 \eta(\epsilon),
\]

Inequalities (161) and (162) yield

\[
- \log L_{n}
\]

Let us define a function \( \eta_{1} \) by

\[
\eta_{1}(x) := \eta'' \left( \sqrt{x/2} \right) + \eta''(4 \cdot \sqrt{x/2} + 36x),
\]

which satisfies \( \lim_{x \to 0} \eta_{1}(x) = 0 \). From Inequality (164), Lemma 35 and (151), we have

\[
- \log L_{n} \geq n \log d - \sum_{k \in \mathbb{K}} p_{k} S(A') \eta_{k} - \eta_{1} \left( 12 \sqrt{\tau} \right) \log (d \kappa),
\]

where we used the fact that we have \( \epsilon_{k} \leq 4 \) from (152) and \( 6 \sqrt{\tau} = 16 = 4^{2} \) from (140). Combining this with Inequality (143), and defining \( \xi_{1}(x) := \xi(x) + \eta_{1}(x) \), we obtain Inequality (144).

B. Proof of Inequality (145)

To prove Inequality (145), note that \( \hat{M}_{n,k} \) is a state merging of \( \Psi_{k}^{p} A \hat{R}_{A} (B_{1} B_{E}) \) with the error \( \epsilon_{k} + 2 \sqrt{\epsilon_{k}} \) and the classical communication cost \( nC_{b} \). Therefore, from Theorem 36 Inequalities (157) and (88), we have

\[
nC_{b} + 5n \eta'' \left( 2 \sqrt{\epsilon_{k} + 2 \sqrt{\epsilon_{k}}} \right) \log (d \kappa)
\]

for each \( k \in \mathbb{K}(\lambda^{*}) \). From (149), (87), (126), (152) and

\[
\log \dim \mathcal{H}_{B_{E}} = nC_{b}, \quad \Psi_{k}^{B_{0}} = (\pi_{d}^{\otimes n})^{B} \otimes \Psi_{k}^{B_{0}},
\]
we have
\[ S(B_1 B\delta) \eta_k^n = S(B_1 B\delta B) \eta_k^n \otimes \eta_k^{2n} - S(B) \eta_k^{2n} \]
\[ \leq S(B_1 B\delta B) \eta_k^n - n \log d - n \eta_k^2 \log (d k \cdot 2^{C_k}) \]
\[ = S(B B\delta) \eta_k^n - n \log d - n \eta_k^2 \log (d k \cdot 2^{C_k}) \]
\[ = S(B) \eta_k^n - n \eta_k^2 \log (d k \cdot 2^{C_k}) \]
\[ \geq S(B) \eta_k^n - n \eta_k^2 \log (d k \cdot 2^{C_k}). \]  

(167)

Using (149), (160), (126) and (152), we also have
\[ S(A') \eta_k^n \leq S(A') \eta_k^n + n \eta_k^2 \log (d k \cdot 2^{C_k}) \]
\[ \leq S(A') \eta_k^n + n \eta_k^2 \log (d k \cdot 2^{C_k}). \]  

(168)

From (166), (167), (168) and (163), we see that
\[ n C_b \]
\[ \geq n \log d - S(A') \eta_k^n + S(B_0) \eta_k^n \]
\[ - n \left( 5 \eta_k^2 (2 \sqrt{\epsilon} + 2 \sqrt{\epsilon}) + 3 \eta_k^2 \right) \log (d k \cdot 2^{C_k}) \]
\[ \geq n \log d - S(A') \eta_k^n + S(B_0) \eta_k^n - 5 n \eta_k^2 \log (d k \cdot 2^{C_k}) \]
for each \( k \in K(\lambda^*). \) Thus we have
\[ n C_b \geq \sum_{k \in K(\lambda^*)} p_k n C_b \]
\[ \geq \sum_{k \in K(\lambda^*)} \left( n \log d - S(A') \eta_k^n + S(B_0) \eta_k^n \right) \]
\[ - 5 n \left( \eta_k^2 (2 \sqrt{\epsilon} + 2 \sqrt{\epsilon}) + \eta_k^2 \right) \log (d k \cdot 2^{C_k}) \]
\[ \geq \sum_{k \in K(\lambda^*)} \left( n \log d - S(A') \eta_k^n + S(B_0) \eta_k^n \right) \]
\[ - 5 n \eta_k^2 \log (d k \cdot 2^{C_k}) \]
\[ - \sum_{k \in K(\lambda^*)} \left( n \log d - S(A') \eta_k^n + S(B_0) \eta_k^n \right). \]  

(169)

Noting that \( \dim H_{B_0} = k \) and
\[ \sum_{k \in K(\lambda^*)} \left( n \log d - S(A') \eta_k^n + S(B_0) \eta_k^n \right) \]
\[ \leq \sum_{k \in K(\lambda^*)} \left( n \log d + \log k \right) \leq 432 \sqrt{\epsilon} \cdot n \log (d k) \]
from (162). Inequality (169) leads to
\[ n C_b \geq \sum_{k \in K(\lambda^*)} \left( n \log d - S(A') \eta_k^n + S(B_0) \eta_k^n \right) \]
\[ - 5 n \eta_k^2 (2 \sqrt{\epsilon} + 2 \sqrt{\epsilon}) \log (d k \cdot 2^{C_k}) \]
where \( \eta_1 \) is a function defined by (165). Thus, from Inequality (19) on \( \Delta S(A')_{av} - \Delta S(G)_{av}, \) we have
\[ n C_b \geq n M(\Upsilon^+) \]
\[ - n \left( \xi(12 \sqrt{\epsilon} \cdot n) + 5 \eta_1 (12 \sqrt{\epsilon}) \right) \log (d k \cdot 2^{C_k}). \]  

(170)

Defining \( \xi_2(x) := \xi(x) + 5 \eta_1(x), \) we obtain Inequality (145).

C. On the Convergence Speed of the Error

We prove that the converse part of Theorem 5 holds even when we drop Condition (9), if Conjecture 28 is true. First, as we proved in Appendix E of [13] (see Remark therein), the function \( \xi(n e) \) in (107) can be replaced by another function \( \xi_2(e) \), which is independent of \( n \) and satisfies \( \lim_{\epsilon \to 0} \xi_2(\epsilon) = 0 \). Consequently, functions \( \xi_1(12 \sqrt{\epsilon} \cdot n) \) and \( \xi_2(12 \sqrt{\epsilon} \cdot n) \) in Inequalities (142) - (145) can be replaced by different functions \( \xi_2(\epsilon), \xi_2^1(\epsilon) \) and \( \xi_2^2(\epsilon), \) respectively, which do not depend on \( n \) and vanish in the limit of \( \epsilon \to 0 \). Inequalities (142) - (145) then hold for any \( n \) and \( \epsilon \), which implies that the converse part holds without additional assumption (9).

Appendix H

Proof of Theorem 26

We prove Theorem 26 after introducing a theorem regarding the cost of randomness for destroying correlations in a bipartite quantum state.

A. Decoupling

The following lemma is obtained as a corollary of Proposition 2 in [42], except an evaluation of the convergence speed of the error.

Lemma 37 Let \( \pi^A \) be the maximally mixed state on \( \mathcal{H}^A \), and suppose a bipartite state \( \rho^{AB} \in \mathcal{S}(\mathcal{H}^A \otimes \mathcal{H}^B) \) satisfies \( \rho^A = \pi^A \). There exists a constant \( c > 0 \) that satisfies the following properties for any \( R > I(A : B)_\rho \), sufficiently small \( \delta > 0 \) and sufficiently large \( n \). That is, for an arbitrary ensemble of unitaries on \( (\mathcal{H}^A)^{\otimes n} \) satisfying
\[ \forall |\phi\rangle \in (\mathcal{H}^A)^{\otimes n} : \int_V p(dV) |\phi\rangle \langle \phi| V^\dagger = (\pi^A)^{\otimes n}, \]  

(171)

there exists a set of unitaries \( \{ V_k \}_{k=1}^{2^n R} \) on the support of \( p(dV) \), such that a random unitary operation \( \mathcal{V}_n \) on \( \mathcal{S}(\mathcal{H}^A) \) defined by
\[ \mathcal{V}_n(\cdot) = \frac{1}{2^n R} \sum_{k=1}^{2^n R} V_k(\cdot) V_k^\dagger \]  

(172)

satisfies
\[ \| \mathcal{V}_n((\rho^{AB})^{\otimes n}) - (\pi^A)^{\otimes n} \otimes (\rho^B)^{\otimes n} \|_1 \]
\[ \leq 2^{-n \delta} + 14 \exp \left( -\frac{\sigma^2 c^2}{4} \right). \]  

(173)

Proof: The proof is basically the same as that of Proposition 2 in [42]. Fix an arbitrary \( \delta > 0 \). Let \( \mathcal{H}^A_{n, \delta} \subset (\mathcal{H}^{AB})^{\otimes n} \) and \( \mathcal{H}^B_{n, \delta} \subset (\mathcal{H}^{AB})^{\otimes n} \) be the \( \delta \)-weakly typical subspaces with respect to \( (\rho^{AB})^{\otimes n} \) and \( (\rho^B)^{\otimes n} \), and let \( \Pi^A_{n, \delta} \) and \( \Pi^B_{n, \delta} \) be the projection onto those subspaces, respectively. There exists a \( \delta \)-independent constant \( c > 0 \) such that we have
\[ \text{Tr}[(\Pi^A_{n, \delta} (\rho^{AB})^{\otimes n})] \geq 1 - \exp(-c\delta^2 n) \]
\[ \text{Tr}[(\Pi^B_{n, \delta} (\rho^B)^{\otimes n})] \geq 1 - \exp(-c\delta^2 n) \]  

(174)
for any $\delta > 0$ and $n$ [43]. Define

$$
\tilde{D}_{n,\delta} := \text{Tr}[\Pi_{n,\delta}^{A} \Pi_{n,\delta}^{B} (\rho^{AB})^{\otimes n} \Pi_{n,\delta}^{A} \Pi_{n,\delta}^{B}].
$$

Due to Lemma 33, we have

$$
\left\| (\rho^{AB})^{\otimes n} - \tilde{\rho}_{n,\delta}^{AB} \right\|_{1} \leq 5 \exp \left(-\frac{c\delta^{2}n}{4}\right),
$$

in addition to

$$
\tilde{D}_{n,\delta} \geq 1 - 2 \exp \left(-\frac{c\delta^{2}n}{2}\right).
$$

Let $\Pi_{n,\delta}^{B}$ be the projection onto the subspace of $H_{n,\delta}^{B}$ spanned by the eigenvectors of $\tilde{\rho}_{n,\delta}^{B}$ corresponding to the eigenvalues not smaller than $\lambda^{*} := 2^{-n(S(\rho^{B})+\delta)} \cdot \exp \left(-\frac{c\delta^{2}n}{2}\right)$.

Define

$$
D_{n,\delta} := \text{Tr}[\Pi_{n,\delta}^{A} \tilde{\rho}_{n,\delta}^{AB}],
$$

$$
\rho_{n,\delta}^{AB} := \frac{\Pi_{n,\delta}^{B} \tilde{\rho}_{n,\delta}^{AB} \Pi_{n,\delta}^{B}}{D_{n,\delta}}.
$$

We then have

$$
D_{n,\delta} = 1 - \text{Tr}[(I - \Pi_{n,\delta}^{A}) \rho_{n,\delta}^{A}],
$$

$$
\geq 1 - \lambda^{*} \times \text{rank}[\rho_{n,\delta}^{A}],
$$

$$
\geq 1 - 2^{-nc\delta^{2} \log e/2},
$$

$$
\geq 1 - \exp \left(-\frac{c\delta^{2}n}{2}\right),
$$

where we used the fact that

$$
\text{rank}[\rho_{n,\delta}^{AB}] \leq \text{dim} \mathcal{H}_{n,\delta}^{B} \leq 2^{n(S(\rho^{B})+\delta)}.
$$

Therefore, due to the gentle measurement lemma, we have

$$
\left\| \rho_{n,\delta}^{AB} - \tilde{\rho}_{n,\delta}^{AB} \right\|_{1} \leq 2 \exp \left(-\frac{c\delta^{2}n}{4}\right).
$$

From (175), (179) and the triangle inequality, we obtain

$$
\left\| (\rho^{AB})^{\otimes n} - \tilde{\rho}_{n,\delta}^{AB} \right\|_{1} \leq 7 \exp \left(-\frac{c\delta^{2}n}{4}\right).
$$

From Definitions (174), (177) and Inequalities (176), (178), the maximum eigenvalue $\lambda^{+}$ of $\tilde{\rho}_{n,\delta}^{AB}$ is bounded as

$$
\lambda^{+} \leq \frac{2^{-n(S(\rho^{A})+\delta)} \cdot D_{n,\delta} \lambda^{+}}{1 - 3 \exp \left(-\frac{c\delta^{2}n}{2}\right)} \leq 2^{-n(S(\rho^{A})-2\delta)}
$$

for sufficiently large $n$. By definition, we also have

$$
\frac{1}{\lambda^{+}} \tilde{\rho}_{n,\delta}^{A}(V) \leq (\Pi^{A})^{\otimes n} \Pi_{n,\delta}^{A}.
$$

Let $\{p(dV), V\}$ be an ensemble of unitaries on $(H^{A})^{\otimes n}$ that satisfies (171), and define

$$
\tilde{\rho}_{n,\delta}^{AB}(V) := V \tilde{\rho}_{n,\delta}^{AB} V^{\dagger}.
$$

As an ensemble average, we have

$$
\rho_{n,\delta}^{AB} := \mathbb{E}[\tilde{\rho}_{n,\delta}^{AB}(V)] = (\pi^{A})^{\otimes n} \otimes \rho_{n,\delta}^{B}.
$$

Inequality (180) then implies

$$
\left\| \rho_{n,\delta}^{AB} - (\pi^{A})^{\otimes n} \otimes (\rho^{B})^{\otimes n} \right\|_{1} \leq \left\| \rho_{n,\delta}^{AB} - (\rho^{AB})^{\otimes n} \right\|_{1} \leq 7 \exp \left(-\frac{c\delta^{2}n}{4}\right),
$$

where the second line follows from the monotonicity of the trace distance. Due to (177) and (178), the minimum nonzero eigenvalue $\lambda^{-}$ of (184) is bounded as

$$
\lambda^{-} \geq \lambda^{*} - \frac{\lambda^{+}}{d_{A}^{A} d_{A}^{B}} \geq d_{A}^{-n} \cdot \lambda^{*},
$$

which leads to

$$
\lambda := \frac{\lambda^{-}}{\lambda^{+}} \geq 2^{-n[\log d_{A} + S(\rho^{B}) - S(\rho^{A} + 3\delta)]} \cdot \exp \left(-\frac{c\delta^{2}n}{2}\right)
$$

$$
= 2^{-n[I(A:B)_{\rho} + 3\delta]} \cdot \exp \left(-\frac{c\delta^{2}n}{2}\right).
$$

Suppose $V_{1}, \cdots, V_{N}$ are unitaries that are randomly and independently chosen from an ensemble $\{p(dV), V\}$. Due to (182) and the operator Chernoff bound (Lemma 3 in [42]), we have

$$
\text{Pr} \left\{ \frac{1}{N} \sum_{i=1}^{N} \tilde{\rho}_{n,\delta}^{AB}(V_{i}) \notin \left( 1 - \epsilon_{1} \right) \rho_{n,\delta}^{AB}, \left( 1 + \epsilon_{1} \right) \rho_{n,\delta}^{AB} \right\} \leq \text{Pr} \left\{ \frac{1}{N} \sum_{i=1}^{N} \tilde{\rho}_{n,\delta}^{AB}(V_{i}) \notin \left( 1 - \epsilon_{1} \right) \rho_{n,\delta}^{AB}, \left( 1 + \epsilon_{1} \right) \rho_{n,\delta}^{AB} \right\}
$$

$$
\leq 2 d_{A}^{n} d_{B}^{n} \exp \left(-\frac{N \epsilon_{1}^{2}}{2}\right)
$$

for any $\epsilon_{1} \in (0, 1]$, which implies that

$$
\text{Pr} \left\{ \left\| \frac{1}{2nR} \sum_{i=1}^{2nR} \rho_{n,\delta}^{AB}(V_{i}) - \tilde{\rho}_{n,\delta}^{AB} \right\|_{1} \leq 2 \epsilon_{1} \right\}
$$

$$
\geq 1 - 2 d_{A}^{n} d_{B}^{n} \exp \left(-\frac{2nR \lambda^{2}_{+}}{2}\right)
$$

for an arbitrary $R > 0$. Substituting $\epsilon_{1} = 2^{-1-n\delta}$, we obtain

$$
\text{Pr} \left\{ \left\| \frac{1}{2nR} \sum_{i=1}^{2nR} \rho_{n,\delta}^{AB}(V_{i}) - \tilde{\rho}_{n,\delta}^{AB} \right\|_{1} \leq 2^{-n\delta} \right\}
$$

$$
\geq 1 - 2 d_{A}^{n} d_{B}^{n} \exp \left(-\frac{2n(R-2\delta) \lambda^{+}_{+}}{8}\right).
$$

Therefore, if $R$ satisfies

$$
R > I(A:B)_{\rho} + 5\delta + \frac{1}{2} c\delta^{2} \log e,
$$

and if $n$ is sufficiently large so that Inequality (183) holds and the R.H.S. in (186) is greater than 0, there exists a set of unitaries $\{V_{i}\}_{i=1}^{2nR}$ such that

$$
\left\| \frac{1}{2nR} \sum_{i=1}^{2nR} \rho_{n,\delta}^{AB}(V_{i}) - \tilde{\rho}_{n,\delta}^{AB} \right\|_{1} \leq 2^{-n\delta}.
$$
Using unitaries in the set, construct a random unitary operation \( \mathcal{V}_n \) on \( \bar{A} \) as (172).

The total error is evaluated as follows. From (180), (183) and the monotonicity of the trace distance, we have

\[
\| \mathcal{V}_n((\rho^{AB})^\otimes n) - \frac{1}{2^{nR}} \sum_{i=1}^{2^n} \rho^{\bar{A}\bar{B}}_n \delta(V_i) \|_1 \leq 7 \exp \left( -\frac{c\delta^2 n}{4} \right). 
\]  

(189)

Due to (183), (188), (189) and the triangle inequality, we obtain

\[
\| \mathcal{V}_n((\rho^{AB})^\otimes n) - (\pi^A)^\otimes n \otimes (\rho^B)^\otimes n \|_1 
\leq 2^{-n\delta} + 14 \exp \left( -\frac{c\delta^2 n}{4} \right)
\]

for any \( R > I(A : B)_\rho, \) \( \delta \in (0,1] \) satisfying (187) and sufficiently large \( n \). Thus we obtain (173). \( \blacksquare \)

B. Proof of Theorem 26

We prove Theorem 26 by showing that \( M(U) \leq K(U) \) holds for any generalized Clifford operator \( U \), which implies \( M(U) = K(U) \) due to Lemma 24. From Equality (46), we have

\[
M(U) = \left| \left| \sum_{i=1}^{2^n} \rho^{\bar{A}\bar{B}}_n \delta(V_i) \right| \right|_1 
\leq 7 \exp \left( -\frac{c\delta^2 n}{4} \right).
\]

(181)

Without loss of generality, we assume that the basis \( \{ |t_i \rangle \}_{i=1}^{d-1} \), by which the generalized Pauli operators are defined as (51), is the Schmidt basis of \( |\Phi_d\rangle^{AR_A} \) and \( |\Phi_d\rangle^{BR_B} \). That is, we assume that

\[
|\Phi_d\rangle^{AR_A} = \frac{1}{\sqrt{d}} \sum_{t=1}^{d} |t\rangle |t\rangle^{R_A}, \quad |\Phi_d\rangle^{BR_B} = \frac{1}{\sqrt{d}} \sum_{t=1}^{d} |t\rangle |t\rangle^{R_B}.
\]

A simple calculation then leads to

\[
s_{pq}^{AB} |\Phi_d\rangle^{BR_B} = \langle \sigma^{T}_{pq} \rangle^{B} |\Phi_d\rangle^{BR_B} = \exp \left( 2\pi i pq / d \right) \cdot s_{pq}^{B} |\Phi_d\rangle^{BR_B},
\]

where the superscript \( T \) denotes the transposition with respect to the basis \( \{ |t_i \rangle \}_{i=1}^{d} \) defined by

\[
X^T = \sum_{t,t'=1}^{d} \langle t| X |t' \rangle \cdot |t'\rangle |t\rangle
\]

for \( X \in \mathcal{L}(\mathcal{H}_B^2) \). Therefore, for some phase \( \theta'_{pq} \in \mathbb{R} \) we have

\[
s_{pq}^{AB} |\Psi_U\rangle^{AB} |\Phi_d\rangle^{BR_B} = \langle \sigma^{A}_{pq} \rangle^{A} \langle \sigma^{A}_{pq} \rangle^{A} |\Psi_U\rangle^{AB} |\Phi_d\rangle^{BR_B} = \exp \left( 2\pi i \theta'_{pq} / d \right) \cdot U^{AB} \langle \sigma^{A}_{pq} \rangle^{A} |\Phi_d\rangle^{BR_B}
\]

and the consideration of unitaries

\[
\left\{ \left| 1/d^{2n} \right|, \sigma_{pq}^{AB} \right\}_{pq \in \{1,\cdots,d\}^{2n}}
\]

on \( R_B^{2n} = \bar{R}_B \). Because of Schur’s lemma, the ensemble satisfies

\[
\frac{1}{d^{2n}} \sum_{pq} \sigma_{pq}^{AB} |\phi\rangle \langle \phi|^{R_B} \sigma_{pq}^{AB} = (\pi_d^{AB})^\otimes n
\]

for any \( |\phi\rangle \in \mathcal{H}_B^d \). Therefore, due to Lemma 37 there exists a subset \( \{ p_k q_k \}_{k=1}^{2^n} \subset \{1,\cdots,d\}^{2n} \) such that

\[
\frac{1}{2^{nR}} \sum_{k=1}^{2^n} \sigma_{p_k q_k}^{AB} (\mathcal{V}_U^{\otimes n})^{\bar{A}\bar{R}_A R_B} \sigma_{p_k q_k}^{AB} = (\pi_d^{AB})^\otimes n
\]

where

\[
\epsilon_n := 2^{-n\delta} + 16 \exp \left( -\frac{c\delta^2 n}{4} \right).
\]

(191)

We then have

\[
\| \mathcal{V}_n((\mathcal{V}_U^{\otimes n})^{\bar{A}\bar{R}_A R_B} - (\mathcal{V}_U^{\otimes n})^{\bar{A}\bar{R}_A}) \|_1 \leq \epsilon_n
\]

from (190). Note that \( \mathcal{V}_U^{R_B} = \pi_d^{R_B} \). Due to Lemma 19 and the proof thereof (see Appendix D-B), it follows that the state \( \mathcal{V}_n((\mathcal{V}_U^{\otimes n})^{\bar{A}\bar{R}_A R_B}) \) is \( \epsilon_n \)-recoverable from \( AR_A \). Therefore, from Definition 9 and Theorem 11 in [13], we obtain \( R \geq M(U) \). Note that the error vanishes exponentially to \( n \) due to (191). \( \blacksquare \)
APPENDIX I
PROOF OF LEMMA 29 AND 30

A. Proof of Lemma 29

The proof is based on an idea which is used in [44] to prove that information causality is satisfied in quantum mechanics.

For Inequality (52), define \( L^0 = K^0 = \emptyset \), and denote system \( B \) before the first step by \( B_0 \). By the data processing inequality and the chain rule, we have

\[
I(\vec{X} : B_{\gamma}, L', K') \\
\leq I(\vec{X} : B_{\gamma-1}, L'^{-1}, K' ) \\
+ I(\vec{X} : \tilde{K}_{\gamma} | B_{\gamma-1}, L'^{-1}, K'^{-1}) \\
= I(\vec{X} : B_{\gamma-1}, L'^{-1}, K'^{-1}) \\
+ I(\vec{X}, B_{\gamma-1}, L'^{-1}, K'^{-1} : \tilde{K}_{\gamma}) \\
- I(\tilde{B}_{\gamma-1}, L'^{-1}, K'^{-1} : \tilde{K}_{\gamma}) \\
\leq I(\vec{X} : B_{\gamma-1}, L'^{-1}, K'^{-1}) + H(\tilde{K}_{\gamma}) \\
+ I(\vec{X} : B_{\gamma-1}, L'^{-1}, K'^{-1}) + \log |\tilde{K}_{\gamma}|
\]

for \( \gamma = 1, \ldots , \Gamma \). Hence we obtain

\[
I(\vec{X} : B_{T'}, L', K') \\
= \sum_{\gamma = 1}^{\Gamma} \left[ I(\vec{X} : B_{\gamma}, L', K') - I(\vec{X} : B_{\gamma-1}, L'^{-1}, K'^{-1}) \right] \\
\leq \sum_{\gamma = 1}^{\Gamma} \log |\tilde{K}_{\gamma}| = C_{\text{tot}}.
\]

For Inequality (53), observe that

\[
H(\vec{X}) - H(\vec{X} | \vec{X}') = I(\vec{X} : B_{T'}) \\
\leq I(\vec{X} : B_{T'}, L', K') \leq C_{\text{tot}}.
\]

(192)

due to the data processing inequality. By definition, we have

\[
H(\vec{X}) = nR.
\]

(193)

Fano’s inequality [45] implies

\[
H(\vec{X} | \vec{X}') \leq h(P_e) + nRP_e,
\]

(194)

Substituting (193) and (194) to (192), we obtain (53). \( \blacksquare \)

B. Proof of Lemma 30

The proof of the first statement is based on a protocol proposed in [46]. Let \( B_1 \) and \( B_2 \) be \( d \)-dimensional quantum systems, and let \( \{\sigma_A^d\}_{i=1}^{d^2} \) be the set of generalized Pauli operators on \( \mathcal{H}_A^d \). Define states

\[
|\Phi_A^d\rangle_{AB_1} := \sigma_A^d |\Phi_d^d\rangle_{AB_1} = (\sigma_A^T)_{B_1} |\Phi_d^d\rangle_{AB_1}
\]

and

\[
|\Psi_{U,i}^d\rangle_{AB_1BB_2} := U^{AB_1} |\Phi_{U,i}^d\rangle_{AB_1} |\Phi_d^d\rangle_{BB_2} \\
= (\sigma_A^T)_{B_1} |\Psi_{U}^d\rangle_{AB_1BB_2}
\]

(195)

for \( i = 1, \ldots , d^2 \). Let us introduce notations

\[
\vec{i} := i_1 \cdots i_n \in \{1, \ldots , d^2\}^n, \\
\sigma_T := \sigma_{i_1} \otimes \cdots \otimes \sigma_{i_n}, \\
|\Phi_A^d\rangle_{\vec{A}B_1} := |\Phi_{\vec{i}}^d\rangle_{AB_1} \otimes \cdots \otimes |\Phi_{i_n}^d\rangle_{AB_1}, \\
|\Psi_{U,i}^d\rangle_{\vec{A}B_1BB_2} := |\Psi_{U,i}^d\rangle_{AB_1BB_2} \otimes \cdots \otimes |\Psi_{U,i}^d\rangle_{AB_1BB_2},
\]

and

\[
\rho(U_i, \vec{i}) := U_i(|\Phi_A^d\rangle_{\vec{A}B_1} |\Phi_d^d\rangle_{\vec{B}B_2}).
\]

(196)

From (195) and (196), we have

\[
\rho(U_i, \vec{i}) = (\sigma_{\vec{i}}^T_{B_1}) B_1 \rho(U_i) (\sigma_{\vec{i}}^T_{B_1}) B_1
\]

and

\[
|\Psi_{U,i}^d\rangle_{\vec{A}B_1BB_2} := (\sigma_{\vec{i}}^T_{B_1}) B_1 |\Psi_{U}^d\rangle_{\vec{A}B_1BB_2},
\]

where we defined

\[
\sigma_{\vec{i}}^T := (\sigma_{\vec{i}}^T)^{B_1}.
\]

Therefore, due to the unitary invariance of the fidelity (see Equality (72)), Condition (54) implies

\[
F \left( \rho(U_i, \vec{i}), |\Psi_{U,i}^d\rangle_{\vec{A}B_1BB_2} \right) \geq 1 - \epsilon,
\]

which leads to

\[
F \left( \rho(U_i, \vec{i}), B_1 B_2, |\Psi_{U,i}^d\rangle_{\vec{A}B_1BB_2} \right) \geq 1 - \epsilon
\]

(197)

for any \( \vec{i} \) by taking the partial trace.

Due to Schur’s lemma, we have

\[
\frac{1}{d^2} \sum_{i=1}^{d^2} \sigma_{\vec{i}}^A |\Phi_d^d\rangle_{AB_1} |\Phi_d^d\rangle_{AB_1} = \pi_d^A \otimes \pi_d^{B_1}.
\]

Thus the average state of \( |\Psi_{U,i}^d\rangle \) with respect to the uniform distribution \( p_i = 1/d^2 \) \( (i = 1, \ldots , d^2) \) is given by

\[
\tilde{\Psi}_{U,i}^{AB_1BB_2} := \frac{1}{d^2} \sum_{i=1}^{d^2} \Psi_{U,i}^{AB_1BB_2} \\
= \pi_d^{B_1} \otimes U^{AB_1} (\sigma_d^A \otimes \Phi_d^{B_2}) U^{\dagger AB_1}.
\]

The reduced state of \( \tilde{\Psi}_{U} \) on \( B_1 B_2 \) is

\[
\tilde{\Psi}_{U,1}^{AB_1BB_2} = \pi_d^{B_1} \otimes \text{Tr}_A[U^{AB_1} (\pi_d^A \otimes \Phi_d^{B_2}) U^{\dagger AB_1}] \\
= \pi_d^{B_1} \otimes \text{Tr}_A[\tau_A^{B_1} (\Phi_d^{B_1} \otimes \Phi_d^{B_2}) U^{\dagger AB_1}] \\
= \pi_d^{B_1} \otimes \sum_{s=0}^{d^2-1} c_s^2 F_s |\Phi_d^d\rangle_{AB_1BB_2}
\]

(198)

where \( c_s \) and \( F_s \) are defined by (10). The von Neumann entropy of states \( \tilde{\Psi}_{U,i}^{AB_1BB_2} \) and \( \tilde{\Psi}_{U,1}^{AB_1BB_2} \) are then

\[
S(B_1BB_2)_{\tilde{\Psi}_{U,i}} = S(A)_{\tilde{\Psi}_{U,i}} = \log d
\]

and

\[
S(B_1BB_2)_{\tilde{\Psi}_{U}} = \log d + K(U),
\]

respectively, where the latter follows from (12) and the orthonormality of \( \{ F_s |\Phi_d^d\rangle_{AB_1BB_2} \}_{s=0}^{d^2-1} \). Thus the Holevo information (47), (48), corresponding to the signal states \( \{ \tilde{\Psi}_{U,i}^{AB_1BB_2} \}_{i=1}^{d^2} \)
and the uniform distribution \( p_i = 1/d^2 \) (\( i = 1, \cdots, d^2 \)), is given by

\[
\chi(\{p_i, \Psi^{B_1B_2B_3}_{U,i}\}) = S(B_1B_2B_3)_{\Psi_U} - \frac{1}{d^2} \sum_{i=1}^{d^2} S(B_1B_2B_3)_{\Psi_{U,i}} = K(U).
\]

Due to the Holevo-Schumacher-Westmoreland theorem [47, 48], for any \( \epsilon > 0 \) and sufficiently large \( n \), there exists a subset \( C_n \subset \{1, \cdots, d^2\}^n \) of cardinality \( n(K(U) - \epsilon) \), such that all elements in the set

\[
\{\Psi^{B_1B_2B_3}_{U,i}\}_{i \in C_n}
\]

are distinguishable up to a small error \( \epsilon \). That is, there exists a measurement on \( B_1B_2B_3 \), described by a set of measurement operators \( \{D_i\}_{i \in C_n} \), such that we have

\[
P_e := \frac{1}{|C_n|} \sum_{i \in C_n} \left(1 - \text{Tr} \left[ D_i \Psi^{B_1B_2B_3}_{U,i} \right] \right) \leq \epsilon
\]

Consider the following protocol in which Alice transmits \( n(K(U) - \epsilon) \) bits of classical message to Bob by \( U_n \), assisted by shared entanglement:

1. Alice and Bob initially share \( |\Phi_+^{BB}| \otimes |\Phi_+^{BB}| \), where \( B_1 \) and \( B_2 \) are additional quantum registers that Bob has.
2. To send a message \( k \in \{1, \cdots, 2^nR\} \) where \( R = K(U) - \epsilon \), Alice chooses \( k \)-th element \( \vec{k} = i_1 \cdots i_n \) in \( C_n \), and applies \( \sigma_{\vec{k}} \) on \( \hat{A} \).
3. Alice and Bob apply \( U_n \).
4. Bob performs a measurement on \( B_1B_2B_3 \) described by \( \{D_i\}_{i \in C_n} \).

The state after Step 3) is equal to \( \rho(U_n, \vec{k}) \overline{AB}B_1B_2 \) for each \( k \), which satisfies

\[
\| \Psi^{B_1B_2B_3}_{U,i} - \Psi^{B_1B_2B_3}_{U,i}^\dagger \|_1 \leq \sqrt{\epsilon}
\]
due to (197) and (73). Therefore, due to (58), the average error in transmitting the message is bounded as above

\[
P_e' := \frac{1}{|C_n|} \sum_{i \in C_n} \left(1 - \text{Tr} \left[ D_i \Psi^{B_1B_2B_3}_{U,i} \right] \right) = \frac{1}{|C_n|} \sum_{i \in C_n} \left(1 - \text{Tr} \left[ D_i \Psi^{B_1B_2B_3}_{U,i} \right] \right) + \frac{1}{|C_n|} \sum_{i \in C_n} \text{Tr} \left[ D_i \Psi^{B_1B_2B_3}_{U,i} - \Psi^{B_1B_2B_3}_{U,i} \right] \leq \epsilon + 4\sqrt{\epsilon} \leq 5\sqrt{\epsilon},
\]

which completes the proof. \( \blacksquare \)

REFERENCES

[1] J. Eisert, K. Jacobs, P. Papadopoulos, and M. Plenio, “Optimal local implementation of nonlocal quantum gates,” Phys. Rev. A, vol. 62, p. 052317, 2000.
[2] J. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein, “Entangling operations and their implementation using a small amount of entanglement,” Phys. Rev. Lett., vol. 86, p. 544, 2001.
[3] B. Groisman and B. Reznik, “Implementing nonlocal gates with nonmaximally entangled states,” Phys. Rev. A, vol. 71, p. 032322, 2005.
[4] L. Chen and Y.-X. Chen, “Probabilistic implementation of a nonlocal operation using a nonmaximally entangled state,” Phys. Rev. A, vol. 71, p. 054302, 2005.
[5] M.-Y. Ye, Y.-S. Zhang, and G.-C. Guo, “Efficient implementation of controlled rotations by using entanglement,” Phys. Rev. A, vol. 73, p. 032337, 2006.
[6] D. W. Berry, “Implementation of multipartite unitary operations with limited resources,” Phys. Rev. A, vol. 75, p. 032349, 2007.
[7] N. B. Zhao and A. M. Wang, “Local implementation of nonlocal operations with block forms,” Phys. Rev. A, vol. 78, p. 014305, 2008.
[8] L. Yu, R. B. Griffiths, and S. M. Cohen, “Efficient implementation of bipartite nonlocal unitary gates using prior entanglement and classical communication,” Phys. Rev. A, vol. 81, p. 062315, 2010.
[9] S. M. Cohen, “Optimizing local protocols for implementing bipartite nonlocal unitary gates using prior entanglement and classical communication,” Phys. Rev. A, vol. 81, p. 062316, 2010.
[10] A. Soeda, P. Turner, and M. Murao, “Entanglement cost of implementing controlled-unitary operations,” Phys. Rev. Lett., vol. 107, p. 180501, 2011.
[11] D. Stahlke and R. Griffiths, “Entanglement requirements for implementing bipartite unitary operations,” Phys. Rev. A, vol. 84, p. 032316, 2011.
[12] E. Wakakuwa, A. Soeda, and M. Murao, “Markovianizing cost of tripartite quantum states,” IEEE Trans. Inf. Theory, vol. 63, no. 2, pp. 1280–1298, 2017.
[13] ——, “The cost of randomness for converting a tripartite quantum state to be approximately recoverable,” IEEE Trans. Inf. Theory, vol. PP, no. 99, pp. 1–1, 2017.
[14] M. A. Nielsen, “A simple formula for the average gate fidelity of a quantum dynamical operation,” Phys. Lett. A, vol. 303, no. 4, pp. 249–252, 2002.
[15] H.-K. Lo and S. Popescu, “Concentrating entanglement by local actions: Beyond mean values,” Phys. Rev. A, vol. 63, no. 2, p. 022301, 2001.
[16] P. Hayden, R. Josza, 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.
[17] M. Nielsen, C. Dawson, J. Dodd, A. Gilchrist, D. Mortimer, T. Osborne, M. Brenner, A. Harrow, and A. Hines, “Quantum dynamics as a physical resource,” Phys. Rev. A, vol. 67, p. 052301, 2003.
[18] J. Oppenheim and B. Reznik, “Probabilistic and information-theoretic interpretation of quantum evolutions,” Phys. Rev. A, vol. 70, p. 022312, 2004.
[19] W. Dür, P. Sekatski, and M. Skotiniotis, “Deterministic superreplication of one-parameter unitary transformations,” Phys. Rev. Lett., vol. 114, no. 12, p. 120504, 2015.
[20] G. Chiribella, Y. Yang, and C. Huang, “Universal superreplication of unitary gates,” Phys. Rev. Lett., vol. 114, no. 12, p. 120504, 2015.
[21] M. Horodecki, J. Oppenheim, and A. Winter, “Partial quantum information,” Nature, vol. 436, pp. 673–676, 2005.
[22] ——, “Quantum state merging and negative information,” Comm. Math. Phys., vol. 269, pp. 107–136, 2007.
[23] E. Walukauwa, A. Soeda, and M. Murao, “A four-round locc protocol outperforms all two-round protocols in reducing the entanglement cost for a distributed quantum information processing,” e-print arXiv:1608.07461, 2016.
[24] A. Nayak and J. Salzman, “On communication over an entanglement-assisted quantum channel,” Proc. of the 34th Ann. ACM Symp. on Theor. of Comp., pp. 698–704, 2002.
[25] A. Montanaro and A. Winter, “A lower bound on entanglement-assisted quantum communication complexity,” in Proc. of the 34th Int. Conf. on Autom., Lang. and Prog. (ICALP'07), 2007, pp. 122–133.
[26] M. B. Plenio and S. Virmani, “An introduction to entanglement measures,” Quant. Inf. Comput., vol. 7, no. 1–5, 2007.
[27] A. E. Gamal and Y.-H. Kim, Network Information Theory. Cambridge University Press, 2011.
[28] E. Wakakuwa and M. Murao, “Asymptotic compressibility of entanglement and classical communication in distributed quantum computation,” e-print arXiv:1310.3991.
[29] E. Wakakuwa, A. Soeda, and M. Murao, “A coding theorem for bipartite unitaries in distributed quantum computation,” in Proc. of 2015 IEEE Int. Symp. on Inf. Theory (ISIT), 2015, pp. 705–709.
[30] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information. Cambridge University Press, 2000.
[31] M. Hayashi, Quantum Information: An Introduction. Springer, 2006.
[32] M. Wilde, Quantum Information Theory. Cambridge University Press, 2013.
[33] A. Uhlmann, “The transition probability” in the state space of a *-algebra,” Rep. Math. Phys., vol. 9, pp. 273–279, 1976.
[34] M. Fannes, “A continuity property of the relative entropy density for spin lattice systems,” *Comm. Math. Phys.*, vol. 31, pp. 291–294, 1973.

[35] R. Alicki and M. Fannes, “Continuity of quantum conditional information,” *J. Phys. A: Math. Gen.*, vol. 37.5, pp. L55–L57, 2004.

[36] W. K. Wooters, “Entanglement of formation of an arbitrary state of two qubits,” *Phys. Rev. Lett.*, vol. 80, p. 10, 1998.

[37] M. A. Nielsen, “Continuity bounds for entanglement,” *Phys. Rev. A*, vol. 61, p. 064301, 2000.

[38] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, “Quantifying entanglement,” *Phys. Rev. Lett.*, vol. 78, p. 2275, 1997.

[39] V. Vedral and M. B. Plenio, “Entanglement measures and purification procedures,” *Phys. Rev. A*, vol. 57, p. 1619, 1998.

[40] M. J. Donald and M. Horodecki, “Continuity of relative entropy of entanglement,” *Phys. Lett. A*, vol. 264, pp. 257–260, 1999.

[41] M. Christandl and A. Winter, “‘squashed entanglement’: An additive entanglement measure,” *J. Math. Phys.*, vol. 45.3, pp. 829–840, 2004.

[42] B. Grossman, S. Popescu, and A. Winter, “Quantum, classical, and total amount of correlations in a quantum state,” *Phys. Rev. A*, vol. 72, p. 032317, 2005.

[43] R. Ahlswede, “A method of coding and an application to arbitrarily varying channels,” *J. Comb., Info. and Syst. Sciences*, vol. 5, pp. 10–35, 1980.

[44] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, “Information causality as a physical principle,” *Nature*, vol. 461, pp. 1101–1104, 2009.

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory (2nd ed).* Wiley-Interscience, 2005.

[46] D. W. Berry, “Lower bounds for communication capacities of two-qudit unitary operations,” *Phys. Rev. A*, vol. 76, p. 062302, 2007.

[47] B. Schumacher and M. D. Westmoreland, “Sending classical information via noisy quantum channel,” *Phys. Rev. A*, vol. 56, p. 131, 1997.

[48] A. S. Holevo, “The capacity of the quantum channel with general signal states,” *IEEE Trans. Inf. Theory*, vol. 44, pp. 269–273, 1998.