Novel chain rules for one-shot entropic quantities via operational methods

Sayantan Chakraborty[1] Upendra Kapshikar[1]

[1] Centre for Quantum Technologies, National University of Singapore

Abstract

We introduce a new operational technique for deriving chain rules for general information theoretic quantities. This technique is very different from the popular (and in some cases fairly involved) methods like SDP formulation and operator algebra or norm interpolation. Instead, our framework considers a simple information transmission task and obtains lower and upper bounds for it. The lower bounds are obtained by leveraging a successive cancellation encoding and decoding technique. Pitting the upper and lower bounds against each other gives us the desired chain rule. As a demonstration of this technique, we derive chain rules for the *smooth max mutual information* and the *smooth-Hypothesis testing mutual information*.

1 Introduction

In 1948, Shannon [Sha48] pioneered the field of information theory by introducing two central problems; noiseless source coding and noisy channel coding. To that end, Shannon introduced the notions of Shannon entropy and mutual information, which characterise these two information processing tasks, respectively. Since then, these two quantities have found numerous applications in many other problems, both within information theory, as well as in cryptography and computer science in general. For a random variable $X \sim P_X$, its entropy $H(X)$ is defined as

$$H(X) = \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[\frac{1}{\log(P_X(x))} \right].$$

For a joint probability distribution P_{XY}, one can analogously define its entropy $H(XY)$;

$$H(XY) = \mathop{\mathbb{E}}_{x,y \leftarrow P_{XY}} \left[\frac{1}{\log(P_{XY}(x,y))} \right].$$

A *chain rule* for the entropy establishes a relationship between the joint entropy and the entropies of the individual variables:

$$H(XY) = H(X) + H(Y \mid X),$$

where

$$H(Y \mid X) \coloneqq \mathop{\mathbb{E}}_{x \leftarrow P_X} \left[H(Y \mid X = x) \right]$$

is the *conditional entropy* of the random variable Y given X.

Such decompositions of joint variable functionalities into individual functionalities are known to hold not only for the entropy function, but also for other useful quantities. For example, consider a tripartite probability distribution P_{XYZ}. Then a chain rule for the mutual information between the systems XY and Z can be written as:

$$I(XY : Z) = I(X : Z) + I(Y : Z \mid X).$$

Chain rules in general are very useful in the design and analysis of information processing protocols, particularly those where multiple parties are present [SW73, Ahl71, Ahl74, Lia72] . Chain rules for mutual information have been used in contexts other than information theoretic tasks, for example, in proving direct sum and direct product theorems in communication complexity, [Raz92, JRP03, JRS03, JRS05] to name a few (see [Jai21] for a more comprehensive list).

The information theoretic quantities mentioned above can be also be defined for more general objects such as quantum states. For a quantum state ρ^A, the *von Neumann entropy*, is defined as

$$H(A) \coloneqq -\operatorname{Tr}\left[\rho^A \log \rho^A\right].$$

Analogously, for a bipartite state ρ^{AB}, the quantum mutual information is defined as

$$I(A:B) \coloneqq H(A) + H(B) - H(AB).$$

However, the conditional entropy of the system A given B cannot be defined in a manner similar to that in the classical case. Thus, in this case one uses the chain rule itself to define the quantum conditional entropy:

$$H(A \mid B) \coloneqq H(AB) - H(B).$$

The chain rule for mutual information follows from the chain rule of H and the definition of I. Furthermore, Jain [Jai21] used these chain rules along with the existence of Nash Equilibrium for some suitably defined games to derive a chain rule for the *capacity* of classical-quantum and quantum channels.

The Shannon and von Neumann entropic quantities although useful in characterizing many important information processing tasks, are somewhat restricted. They are most useful in settings where many *independent* copies of the underlying resource are available. For example, in quantum source compression one exhibits an algorithm to compress the quantum state $\rho^{\otimes n}$ using only $nH(A)$ many qubits. For this, it is usually assumed that n copies of a quantum state ρ are available, in order to show that there exists a compression algorithm. Comparatively, a more natural framework is that of *one-shot* information theory which considers the setting where only *one* copy of the underlying resource is available. There exists a rich body of work which explores information theoretic questions in this setting with the aid of the smooth min and max entropy formalism. This formalism was introduced and developed by a series of papers [Ren05, RW04, RK05, Dat09, RR11a, TCR10, TCR09, RR11b] in the context of both information theoretic and cryptographic applications. The (conditional) smoothed min and max entropies ($H^{\varepsilon}_{\max}(A \mid B)$ and $H^{\varepsilon}_{\min}(A \mid B)$, respectively) are *robust* versions of the corresponding unsmoothed quantities. Here the parameter ε, referred to as the smoothing parameter, is used to specify the accuracy of the certain protocols. For example, the smooth min entropy $H^{\varepsilon}_{\min}(A \mid B)$ characterizes the number of (almost) random bits one can extract from the system A when an adversary is in possession of the system B. The parameter ε here denotes the requirement that the random bits produced in such an extraction should have a bias of at most ε (see [DBWR14, Dup10]). Similarly, the quantity $H^{\varepsilon}_{\max}(A \mid B)$ characterizes the number of entangled qubits required for *state merging* [HOW07, Ber09]. Thus, given their importance, a natural question is whether these quantities obey chain rules similar to their von Neumann counterparts. This question was investigated in the work of Vitanov et al. [VDTR13], where the authors provided several chain rules for the smooth min max entropies. It is worth pointing out that the chain rules that one gets for such quantities are only one sided chain rules, in that they are inequality expressions rather than an equality.

Example 1. In [VDTR13], Vitanov et al. showed the following chain rule for the smooth min entropy (ignoring additive log terms): Given a quantum state ρ^{ABC} and $\varepsilon, \varepsilon', \varepsilon'' > 0$ such that $\varepsilon > \varepsilon' + 2\varepsilon''$, it holds that:

$$H^{\varepsilon}_{\min}(AB|C) \geq H^{\varepsilon'}_{\min}(A|C) + H^{\varepsilon''}_{\min}(A|BC).$$

Dupuis further showed similar chain rules for the sandwiched Rényi α-entropies in [Dup15].

Although the smoothed min and max entropy formalism has proven to be very useful in the description of several quantum information processing tasks, it does not tell the whole story. The works of Anshu et al. [ADJ17, AJW19b, AJW19a] and Wang and Renner [WR12] highlight the importance of smooth max divergence, the smooth hypothesis testing divergence and their derivative quantities (see Appendix A.1 for the relevant definitions). Wang and Renner characterised the one-shot capacity of a classical-quantum channel $\mathcal{N}^{X \to B}$ in terms of the smooth hypothesis testing mutual information:

$$\max_{P_X} I^{\varepsilon}_{H}(X:B).$$

A similar characterisation for the entanglement assisted classical capacity of a channel $\mathcal{N}^{A\to B}$ was given in the work of Anshu et al. [AJW19a], who showed the assisted classical capacity of any quantum channel is given by

$$\max_{|\varphi\rangle^{RA}} I_H^{\varepsilon}(R:B)_{\mathbb{I}^R\otimes\mathcal{N}^{A\to B}(\varphi^{RA})}.$$

Another important quantity *smooth max mutual information* $I_{\max}^{\varepsilon}(A:B)$ gives an achievable quantum communication cost for the state redistribution problem [ADJ17] and state splitting [ADJ17, BCR11]. Unlike their smooth max min entropic counterparts, to the best of our knowledge, the existence of chain rules for these important information quantities have not received much attention. Our goal in this paper is to introduce techniques which will enable us to present chain rules for these quantities.

1.1 Our Contribution

The main results that we present in this work are as follows:

Theorem 1.1. [Informal] *Given any quantum state ρ^{ABC}, it holds that*

$$I_H^{\varepsilon}(AB:C)\geq I_H^{\varepsilon'}(A:C)+I_H^{\varepsilon''}(B:AC)-I_{\max}(A:B)+O\left(\log\varepsilon\right)$$

where, ε' and ε'' are $O(\varepsilon^2)$.

Theorem 1.2. [Informal] *Given any quantum state ρ^{ABC}, it holds that*

$$I_{\max}^{\varepsilon}(AB:C)\leq I_{\max}^{\varepsilon'}(A:C)+I_{\max}^{\varepsilon''}(B:AC)-I_H^{\varepsilon'''}(A:B)-O\left(\log\varepsilon\right)$$

where, ε', ε'' and ε''' are $O(\varepsilon^2)$.

Remark 1.3. Please refer to Appendix A.1 for the definition of I_H^{ε} and $I_{\max}^{\varepsilon}$ along with other important facts used in this paper. Also note that in Theorem 1.1, we use the unsmoothed max information $I_{\max}$. The reason for this will become clear later when we describe our proof technique.

Remark 1.4. We should mention that it is not at all obvious how to prove Theorem 1.1 using standard techniques in one shot information theory. One can suspect that due to a close connection between the smooth hypothesis testing divergence and the information spectrum divergence, it might be to arrive at a chain rule like Theorem 1.1. Indeed, exploiting the said relation one can prove the following statement (ignoring additive log factors):

$$\begin{aligned}I_H^{\varepsilon}\left(AB:C\right)_{\rho}\geq{}& I_H^{\varepsilon}\left(A:C\right)_{\rho}\\&+D_H^{\varepsilon}\left(\rho^{ABC}\parallel\rho^{B}\otimes\Pi_s^{AC}\rho^{AC}\Pi_s^{AC}\right)\end{aligned}$$

where Π_s^{AC} is the information spectrum projector. However, it is not clear how to remove this projector from the expression above to get the desired chain rules, since in general it does not commute with ρ^{AC}.

Remark 1.5. In [DKF$^+$], Dupuis et al. showed chain rules for the smooth hypothesis testing conditional entropy H_H^{ε} using a chain rule for the smooth hypothesis testing divergence between an arbitrary state ρ and a state σ which is invariant under some group action. However, it is not clear how this technique can be used to prove the chain rule claimed in Theorem 1.1.

Remark 1.6. Chakraborty et al. proved a weaker version of Theorem 1.2 in [CNS21]. To be precise, the authors in that paper proved the following bound:

$$I_{\max}^{\varepsilon}(AB:C)\leq I_{\max}^{\varepsilon'}(A:C)+I_{\max}^{\varepsilon''}(B:AC)-O\left(\log\varepsilon\right)$$

for any state ρ^{ABC}. We present a sharper version of this inequality in this paper.

2 Organisation of the paper

The paper is organised as follows: in Section 3.1 we present an overview of the main operational method that we use to prove Theorems 1.1 and 1.2. In this section we also show how an application of these ideas leads directly to the proof of Theorem 1.2. In Section 3.2 we explain why the ideas presented in Section 3.1 cannot be directly applied to prove Theorem 1.1. In this section we also present a weaker version of Theorem 1.1, called Proposition 3.1, which is a result akin to Theorem 1.1 but valid only for a specific subclass of quantum states, which we call IM-states (see Section 3.3). We introduce this proposition for sake of demonstrating the main ideas that eventually go into the proof of Theorem 1.1. In Section 3.3 we present an overview of our proof for Proposition 3.1, followed by the formal definitions and techniques in Sections 4 and 5. Finally, in Section 6, we present the full proof of Theorem 1.1. The definitions of information theoretic quantities and facts used throughout the paper can be found in the preliminaries section in Appendix A.1.

3 Overview of Techniques

In this section we present the main ideas that lead to the proofs of Theorems 1.1 and 1.2.

3.1 The Main Idea

The main techniques used thus far to prove chain rules for the smooth min and max entropies [VDTR13] and the Rényi α-entropies have involved the SDP formulations of these quantities or norm interpolation methods. While these techniques are extremely sophisticated and powerful, in this paper we take a much more simple operational approach to prove Theorems 1.1 and 1.2. The heart of our approach is the following observation:

Consider a situation where two parties Alice and Bob wish to perform a generic information processing task INFO_TASK using a resource state ρ^{AB} and communication. Suppose we are promised the following:

1. *Any* protocol which achieves INFO_TASK using the resource state ρ^{AB} requires Alice and Bob to communicate at least $C(A\to B)_{\rho^{AB}}$ number of bits, where the $C(\cdot)$ is a function of the state ρ^{AB}.
2. There exists a protocol $\mathcal{P}(A\to B)$ which achieves INFO_TASK using the state ρ^{AB}, with a communication cost $C(A\to B)_{\rho^{AB}}$.
3. Additionally, $\mathcal{P}(A\to B)$ ensures that at the end of the protocol the share A of the state ρ^{AB} belongs to Bob and both the classical and quantum correlations between the A and B remain intact.

One can consider a successive cancellation strategy for achieving INFO_TASK : Consider a situation where Alice and Bob wish to achieve INFO_TASK with the resource state $\rho^{A_1A_2B}$. Consider the following strategy:

1. Alice enacts the protocol $\mathcal{P}(A_1\to B)$ using only the marginal ρ^{A_1B}, at a communication cost of $C(A_1\to B)_{\rho^{A_1B}}$. At the end of $\mathcal{P}(A_1\to B)$, the A_1 share of the resource state $\rho^{A_1A_2B}$ belongs to Bob.
2. Alice can then enact the protocol $\mathcal{P}(A_2\to A_1B)$, with a communication cost $C(A_2\to A_1B)_{\rho^{A_1A_2B}}$.

The above protocol achieves INFO_TASK while using the resource state $\rho^{A_1A_2B}$ with a cumulative

$$C(A_1\to B)_{\rho^{A_1B}}+C(A_2\to A_1B)_{\rho^{A_1A_2B}}$$

of communication. Then, using the promised lower bound on the amount of communication required to achieve this task, we see that

$$C(A_1\to B)_{\rho^{A_1B}}+C(A_2\to A_1B)_{\rho^{A_1A_2B}}\geq C(A_1A_2\to B)_{\rho^{A_1A_2B}}.$$

This algorithmic technique of showing the existence of chain rules was first exploited by Chakraborty et al.[CNS21] to demonstrate chain rules for the smooth max mutual information. As mentioned previously, we present below an improved version of this result in Theorem 1.2. The proof idea is as follows :

Consider the task of *quantum state splitting*, in which a party (say Alice) holds the AM share of the pure quantum state $|\varphi\rangle^{RAM}$ at the beginning of the protocol, R being held by the referee. Alice is then required to send the M portion of the state to Bob, while trying to minimize the number of qubits communicated to Bob. It is known [BCR11] that for this problem, Alice needs to communicate at least

$$\frac{1}{2}I^{\varepsilon}_{\max}(R : M)$$

number of qubits to Bob. To show the chain rule, consider a pure stage $|\varphi\rangle^{RAM_1M_2}$. Then:

1. First Alice sends the M_1 system to Bob using state redistribution protocol [ADJ17]. At this point the global state is some $|\varphi'\rangle^{RAM_1M_2}$, which is ε close (in the purified distance) to the original state $|\varphi\rangle^{RAM_1M_2}$, with the M_1 system being in the possession of Bob.

2. To do this, Alice communicated $\frac{1}{2}I^{\varepsilon}_{\max}(R : M_1)$ qubits to Bob (suppressing the additive log terms).

3. Next, Alice sends the system M_2 to Bob. Note that, instead of using the vanilla state redistribution protocol once more (which would cost about $\frac{1}{2}I^{\varepsilon}_{\max}(RM_1 : M_2)$ qubits of communication), we can take advantage of the fact that Bob possesses some side information about the state, in particular, the register M_1 already in his possession. Anshu et al. presented a modified state redistribution protocol in [AJW18] which does precisely this, while reducing the quantum communication cost to

$$\frac{1}{2}\cdot\left(I^{\varepsilon}_{\max}(RM_1 : M_2) - I^{\varepsilon}_{H}(M_1 : M_2)\right).$$

Putting the achievable communication rate derived above against the lower bound shown by [BCR11] then gives us the chain rule:

$$I^{\varepsilon'}_{\max}(R : M_1M_2) \leq I^{\varepsilon}_{\max}(R : M_1) + I^{\varepsilon}_{\max}(RM_1 : M_2) - I^{\varepsilon}_{H}(M_1 : M_2)$$

where we have ignored the additive log terms and set ε' to reflect the total error made by the achievable strategy. The explicit computation of the error is easy and follows along similar lines of the calculation presented in [CNS21] with some minor tweaks. Hence we do not repeat it here. Instead, the rest of the paper is devoted to proving Theorem 1.1, which is technically much harder to prove.

3.2 Issues with I^{ε}_{H}

The idea presented in Section 3.1 can similarly be used to prove chain rules where the direction of the inequality is reversed. In that case one has to consider a task for which Alice and Bob wish to *maximise* the amount of communication, and there exists a known upper bound. However, this idea cannot be readily applied when trying to prove chain rules for I^{ε}_{H}. The difficulties are as follows:

Suppose we wish to derive a chain rule of the form

$$C(A_1A_2 \to B)_{\rho^{A_1A_2B}} \geq C(A_1 \to B) + C(A_2 \to A_1B)$$

we require the existence of an information processing task for which the *maximum* number of bits that can be transmitted is quantified by $C(A_1A_2 \to B)$. Note that this number is a function of a specific fixed state $\rho^{A_1A_2B}$. For I^{ε}_{H}, a natural task that one may consider for this purpose is entanglement assisted channel coding over some quantum channel $\mathcal{N}^{A\to B}$. However, as mentioned before, Anshu et al. showed in [AJW19a, AJW19b] that the maximum number of bits that can be sent using this channel is given by

$$\max_{|\varphi\rangle^{A_1A_2A}} I^{\varepsilon}_{H}(A_1A_2 : B)_{\mathbb{I}^{A_1A_2}\otimes\mathcal{N}^{A\to B}\left(\varphi^{A_1A_2A}\right)}.$$

Note that the above capacity expression is a function of the *channel* and *not* a fixed state. In particular, there is a maximization over state $|\phi\rangle$. This prevents us from directly importing our operational approach here. Note that this was not an issue in the case of $I^{\varepsilon}_{\max}$ since the task of state splitting is defined for a specific fixed state, and not a channel and neither did it involve any maximization. To remedy this situation, we need to do the following:

1. Given a state $\rho^{A_1A_2B}$ we need to exhibit a channel $\mathcal{N}^{A\to B}$ and pure state $|\varphi\rangle^{A_1A_2A}$ such that
$$\mathcal{N}^{A_1A_2\to B}\left(\varphi^{A_1A_2A}\right) = \rho^{A_1A_2B}.$$
2. Having exhibited this channel, we need to show that *any* protocol which uses $|\varphi\rangle^{A_1A_2A}$ as a shared entangled state and sending classical messages across $\mathcal{N}$, can send at most $I^{\varepsilon}_{H}(A_1A_2 : B)$ many bits (and error at most ε).

Before going to chain rules for an arbitrary state, we first show the following preposition, which includes some core ideas of our protocol.

Proposition 3.1. *Given a quantum state ρ^{ABC} such that*
$$\mathrm{Tr}_C\left[\rho^{ABC}\right] = \rho^A \otimes \rho^B,$$
it holds that
$$I^{\varepsilon}_H(AB : C) \geq I^{\varepsilon'}_H(A : C) + I^{\varepsilon''}_H(B : AC) + O\left(\log \varepsilon\right)$$
where both ε' and ε'' are $O(\varepsilon^2)$.

Note that when A and B marginals are in tensor, $I_{\max}(A : B)_\rho = 0$ and hence the above Preposition exactly recovers the chain rule we wanted. Throughout the paper we call such states (where marginals are independent) as IM-states (see Section 3.3 for details).

3.3 A Warm-up: Chain Rules for IM-States

In this section we introduce the proof techniques that we will use to prove Proposition 3.1. See Section 5 for a complete proof. We begin by defining a certain subfamily of tripartite states. The family that we will be interested in, will be such that its marginals on one of the pairs will be independent.
$$\mathcal{F}^{A_fB_fC}_{\mathsf{IM}} = \{\rho^{A_fB_fC} \in \mathcal{D}\left(\mathcal{H}_{A_fB_fC}\right) \mid \rho^{A_fB_f} = \rho^{A_f} \otimes \rho^{B_f}\}.$$
To mean that $\rho^{A_fB_fC} \in \mathcal{F}^{A_fB_fC}_{\mathsf{IM}}$, we will use the shorthand $\rho^{A_fB_fC}$ is an $(A_fB_fC) - \mathsf{IM}$ state. Note that the order of register A_fB_fC matters as the marginals only on A_f and B_f are independent. Recall that to prove chain rules for I^{ε}_H we need to fullfill the Requirements 1 and 2. Requirement 1 is not hard to satisfy for IM-states, as is shown by the following lemma:

Lemma 3.2. *For every $(A_fB_fC) - \mathsf{IM}$ state $\rho^{A_fB_fC}$ and purifications $\varphi_1^{A_fA}, \varphi_2^{B_fB}$ of ρ^{A_f}, ρ^{B_f} respectively, there exists a channel $\mathcal{N}^{AB\to C}$ such that the following holds:*
$$\mathbb{I}^{A_fB_f} \otimes \mathcal{N}^{AB\to C}\left(\varphi_1^{A_fA} \otimes \varphi_2^{B_fB}\right) = \rho^{A_fB_fC}.$$

Proof. Consider a purification $\rho^{A_fB_fCR}$ of $\rho^{A_fB_fC}$. Note that this is also a valid purification of the state $\rho^{A_fB_f} = \rho^{A_f} \otimes \rho^{B_f}$. Then, by the Uhlmann's Theorem (Fact A.6)there exists an isometry $V^{AB\to CR}$ such that
$$\mathbb{I}^{A_fB_f} \otimes V^{AB\to CR}\left(\varphi_1^{A_fA} \otimes \varphi_2^{B_fB}\right) = \rho^{A_fB_fCR}.$$

Define
$$\mathcal{N}^{AB\to C} := \mathrm{Tr}_R \circ V^{AB\to CR}.$$
Then it is easy to see that
$$\mathbb{I}^{A_fB_f} \otimes \mathcal{N}^{AB\to C}\left(\varphi_1^{A_fA} \otimes \varphi_2^{B_fB}\right) = \rho^{A_fB_fC}.$$
This concludes the proof. ∎

Remark 3.3. Given an $(A_f B_f C)-\mathsf{IM}$ state $\rho^{A_f B_f C}$, note that the channel which satisfies the conditions of Lemma 3.2 is not unique, but instead depends on the purifications $|\varphi\rangle_1^{A_f A}$ and $|\varphi\rangle_2^{B_f B}$. Nevertheless, we will always fix these purifications, and refer to *the* channel constructed in Lemma 3.2 as the IM-extended channel of $\left(\rho^{A_f B_f C}, \varphi_1^{A_f A}, \varphi_2^{B_f B}\right)$. When the registers are clear from the context, we will denote it as the IM-extended channel of $(\rho, \varphi_1, \varphi_2)$.

Requirement 2 is much harder to show. To prove that this requirement indeed holds, we use the following idea:

1. We first consider the set of *all* entanglement assisted protocols which use the channel $\mathcal{N}^{AB\to C}$ to send classical messages, with an error at most ε. We call this set $\mathcal{S}(\mathcal{N}, \varepsilon)$.

2. We divide this set into disjoint subsets $\mathcal{S}^{\sigma}(\mathcal{N}, \varepsilon)$, where each subset consists of all those protocols whose encoders create some *fixed* state σ^{AB} on the system which is input to the channel, when all other systems are traced out.

3. We then show that, for a fixed σ^{AB}, *any* protocol in the set $\mathcal{S}^{\sigma^{AB}}(\mathcal{N}, \varepsilon)$ can send at most

$$I_H^{\varepsilon}(A_f B_f : C)_{\mathbb{I}^{A_f B_f} \otimes \mathcal{N}(|\sigma\rangle^{ABA_f B_f})}$$

   number of bits through $\mathcal{N}$, where $|\sigma\rangle^{ABA_f B_f}$ is an arbitrary purification of σ^{AB}. We do this by using a slightly modified form of the converse shown by Anshu et al. [AJW19a].

4. For the case of IM-states, setting

$$\sigma^{AB} \leftarrow \varphi_1^{A_f} \otimes \varphi_2^{B_f}$$

   and

$$|\sigma\rangle^{ABA_f B_f} \leftarrow |\varphi_1\rangle^{AA_f} |\varphi_2\rangle^{BB_f}$$

   completes the argument.

We explore the above idea of partitioning the set of all protocols which make small error in Section 4. The precise definition of the terms $\mathcal{S}(\mathcal{N}, \varepsilon)$ and $\mathcal{S}^{\sigma}(\mathcal{N}, \varepsilon)$ can be found in Section 4.1. The proof of the upper bound for a fixed partition referred to in Point 3 can be found in Section 4.2.

To complete the proof of the chain rule we still need to show a successive coding strategy using the states $|\varphi_1\rangle^{AA_f} |\varphi_2\rangle^{BB_f}$ as a shared resource. To do this we use a standard successive cancellation style argument using Anshu et al.'s coding strategy for entanglement assisted classical message transmission [AJW19a]. Details can be found in Section 5 and Appendix C.

4 Partitioning the Space of Good Protocols

In this section we introduce the partitioning idea, referred to in Section 3.3 that is key to the proof of our chain rule. Towards that end, we clarify definition of an entanglement assisted classical communication protocol (over a noisy quantum channel) in Section 4.1, and go on to define the set of all such protocols which make a small amount of error. Then in Section 4.2 we introduce a way to partition this set and provide upper bounds on the rates of communication of the protocols that belong to a fixed partition.

4.1 The Setup for Entanglement Assisted Classical Communication

Definition 4.1. *A. A protocol $\mathcal{P}$ will be labelled by a tuple $\left(M, \mathcal{N}, \mathcal{E}, \mathcal{D}, |\varphi\rangle^{E_A E_B}\right)$. The average error of the protocol* Er *is given by the following expression:*

$$\begin{aligned}\mathsf{Er}(\mathcal{P}) &= \mathsf{Er}\left(M, \mathcal{N}, \mathcal{E}, \mathcal{D}, |\varphi\rangle^{E_A E_B}\right) \\ &= \left\| \mathcal{D} \circ \mathcal{N} \circ \mathcal{E} \ \left(\psi^{MM_A} \otimes \varphi^{E_A E_B}\right) - \psi^{MM_A} \right\|_1 .\end{aligned}$$

Figure 1: Setup for Entanglement assisted classical communication

Consider an entanglement assisted classical message transmission protocol over a channel $\mathcal{N}^{A\to B}$, which makes error at most ε. Any such protocol consists of the following objects:

1. A state
$$\psi^{MM_A} := \sum_{m\in M} \frac{1}{|M|} |m\rangle \langle m|^M \otimes |m\rangle \langle m|^{M_A}$$
held by the sender Alice.

2. Shared entanglement modelled by a pure state
$$|\varphi\rangle \langle\varphi|^{E_A E_B}.$$
where the E_A system is held by Alice and the E_B system is held by the receiver Bob.

3. An encoder $\mathcal{E}^{M_A E_A \to A}$ which takes as input the states in the M_A and E_A systems and produces a state on the register A, which is the input to the channel $\mathcal{N}^{A\to B}$.

4. A decoder $\mathcal{D}^{BE_B\to \widehat{M}}$, which acts on the register B (the output of the channel), as well as Bob's share of the entanglement, and produces a classical register $\widehat{M}$ which contains a guess for the message sent by Alice.

The protocol is said to make an (average)-error at most ε if
$$\left\|\mathcal{D}\circ\mathcal{N}\circ\mathcal{E}\ \left(\psi^{MM_A}\otimes\varphi^{E_A E_B}\right) - \psi^{MM_A}\right\|_1 \le \varepsilon \tag{1}$$

B. *Let $\mathcal{S}(\mathcal{N},\varepsilon)$ to be the set of all protocols $\mathcal{P}$ which makes an error at most ε while using channel the $\mathcal{N}$.*
$$\begin{aligned}\mathcal{S}(\mathcal{N},\varepsilon) =\Big\{&\mathcal{P} : \exists M,\mathcal{E},\mathcal{D} \textit{ such that } \mathcal{P} \textit{ is an}\\ &\left(M,\mathcal{N},\mathcal{E},\mathcal{D},|\varphi\rangle^{E_A E_B}\right) \textit{ and}\\ &\mathsf{Er}\left(M,\mathcal{N},\mathcal{E},\mathcal{D},|\varphi\rangle^{E_A E_B}\right)\le\varepsilon\Big\}.\end{aligned}$$

C. *We define $\mathcal{S}^{\rho^A}(\mathcal{N},\varepsilon)\subseteq\mathcal{S}(\mathcal{N},\varepsilon)$ to be the set of all those protocols $\mathcal{P}\in\mathcal{S}(\mathcal{N},\varepsilon)$ for which the state at the input to the channel is ρ^A.*
$$\mathrm{Tr}_{ME_B}\left[\mathcal{E}\ \left(\psi^{MM_A}\otimes\varphi^{E_A E_B}\right)\right] = \rho^A.$$

4.2 The Partitions and Corresponding Upper Bounds

Note that $\mathcal{S}^{\rho^A}(\mathcal{N},\varepsilon)$ partitions the set $\mathcal{S}(\mathcal{N},\varepsilon)$. That is, given $\mathcal{P}\in\mathcal{S}(\mathcal{N},\varepsilon)$, there exists a unique ρ^A such that $\mathcal{P}\in\mathcal{S}^{\rho^A}(\mathcal{N},\varepsilon)$. In [AJW19a] Anshu, Jain and Warsi showed that for any protocol $\mathcal{P}\in\mathcal{S}(\mathcal{N},\varepsilon)$ the number of messages M can be upper bounded by D_H^ε of certain states associated with the protocol. For our proof we need a slightly finer version of an analogous statement. In the following lemma, we note that such a statement remains valid over individual partitions as well. Our proof follows a similar strategy as theirs. We include it here for the sake of completeness.

Lemma 4.2. *Let $\mathcal{P}$ be an arbitrary $\left(M,\mathcal{N},\mathcal{E},\mathcal{D},|\varphi\rangle^{E_A E_B}\right)$ protocol in $\mathcal{S}^{\rho^A}(\mathcal{N},\varepsilon)$. Then,*
$$\log|M| \le \min_{\sigma^B} D_H^\varepsilon\left(\mathcal{N}(\tau^{AB'})\ ||\ \sigma^B\otimes\tau^{B'}\right),$$
where $|\tau\rangle^{AB'}$ is an arbitrary purification of the state ρ^A.

Proof. See Appendix B ☐

The following corollary follows immediately by setting $\sigma^B = \mathcal{N}(\rho^A)$:

Corollary 4.3. *Given the setting of Lemma 4.2, we see that for a channel $\mathcal{N}^{A\to B}$ and for all protocols $\mathcal{P} \in \mathcal{S}^{\rho^A}(\mathcal{N}, \varepsilon)$, it holds that*

$$\begin{aligned}\log|M| &\leq D_H^{\varepsilon}\left(\mathcal{N}(\tau^{AB'}) \mid\mid \tau^{B'} \otimes \mathcal{N}(\rho^A)^B\right) \\ &= I_H^{\varepsilon}\left(B : B'\right)_{\mathcal{N}(\tau^{AB'})}.\end{aligned}$$

5 Proof of Proposition 3.1

In this section we present the proof of Proposition 3.1, which we restate below as a theorem:

Theorem 5.1. *Given an* IM*-state* $\rho^{A_f B_f C}$*, it holds that*

$$\begin{aligned}I_H^{O(\varepsilon)}(A_f B_f : C) \ &\geq I_H^{O(\varepsilon^2)}(A_f : C) + I_H^{O(\varepsilon^2)}(B_f : A_f C) \\ &\quad + O(\log \varepsilon).\end{aligned}$$

Proof. First, define $\mathcal{N}^{AB\to C}$ to be the IM-extended channel of the triple

$$\left(\rho^{A_f B_f C}, \varphi_1^{A_f A}, \varphi_2^{B_f B}\right)$$

where $\varphi_1^{A_f A}$ and $\varphi_2^{B_f B}$ are purifications of the states ρ^{A_f} and ρ^{B_f} respectively. The existence of this channel is guaranteed by Lemma 3.2. Recall from Section 4.1 that

$$\mathcal{S}(\mathcal{N}, \delta)$$

denoted the set of all those protocol $\mathcal{P}$ which make an (average) error at most δ while sending classical messages through the channel $\mathcal{N}^{AB\to C}$, with the help of shared entanglement. Note that now our channel takes as input the states in the bipartite system AB and sends the output to the system C. Thus, the description of any protocol $\mathcal{P}$ for this channel will be given by the tuple:

$$\left(M, \mathcal{N}, \mathcal{E}^{M_{AB} E_{AB} \to AB}, \mathcal{D}^{E_C C \to \widehat{M}_{AB}}, |\varphi\rangle^{E_{AB} E_C}\right)$$

It is important to note that the above description does not treat the two systems AB as belonging to two different senders. This allows us to bound the rate at which *any* protocol can send classical messages through $\mathcal{N}$, and not just those protocols which treat A and B as belonging to two spatially separated senders. Now, recall that

$$\mathcal{S}^{\varphi_1^A \otimes \varphi_2^B}(\mathcal{N}, \delta)$$

denotes that subset of $\mathcal{S}(\mathcal{N}, \delta)$ which contains the protocols $\mathcal{P}$ for which the state created by the encoder $\mathcal{E}^{M_{AB} E_{AB} \to AB}$ on the system AB is

$$\varphi_1^A \otimes \varphi_2^B.$$

Then, by Lemma 4.2 and Corollary 4.3, we know that, for *all* protocols in the set $\mathcal{S}^{\varphi_1^A \otimes \varphi_2^B}(\mathcal{N}, \delta)$ it holds that

$$\begin{aligned}\log M &\leq D_H^{\delta}\left(\mathcal{N}^{AB\to C}\left(\varphi_1^{AA_f} \otimes \varphi_2^{BB_f}\right) \ \Big|\Big| \right. \\ &\qquad \left. \varphi_1^{A_f} \otimes \varphi_2^{B_f} \otimes \mathcal{N}\left(\varphi_1^A \otimes \varphi_2^B\right)\right) \\ &= D_H^{\delta}\left(\rho^{A_f B_f C} \mid\mid \rho^{A_f} \otimes \rho^{B_f} \otimes \mathcal{N}\left(\varphi_1^A \otimes \varphi_2^B\right)\right) \\ &= D_H^{\delta}\left(\rho^{A_f B_f C} \mid\mid \rho^{A_f B_f} \otimes \rho^C\right) \\ &= I_H^{\delta}(A_f B_f : C)\end{aligned}$$

We will now exhibit a protocol for sending classical information with entanglement assistance through the channel $\mathcal{N}$ which achieves the rate

$$I_H^{\varepsilon'}(A_f : C) + I_H^{\varepsilon''}(B_f : A_f C) + O(\log \varepsilon)$$

and also satisfies the invariant that the state the encoder of this protocol creates at the input to the channel, averaged over all messages, is

$$\varphi_1^A \otimes \varphi_2^B$$

This protocol makes an error $O(\sqrt{\varepsilon})$. Thus, by setting

$$\delta \leftarrow O(\sqrt{\varepsilon})$$

and noticing that this protocol belongs to the set

$$\mathcal{S}^{\varphi_1^A \otimes \varphi_2^B}(\mathcal{N}, O(\sqrt{\varepsilon})),$$

we will conclude that

$$\begin{aligned} I_H^{O(\sqrt{\varepsilon})}(A_f B_f : C) \ &\geq I_H^{O(\varepsilon)}(A_f : C) + I_H^{O(\varepsilon)}(B_f : A_f C) \\ &+ O(\log \varepsilon). \end{aligned}$$

Replacing $\varepsilon \leftarrow \sqrt{\varepsilon}$, gives the expression in the required form.

The Protocol

To describe the protocol, it will be easier to consider two senders Alice and Bob who have access to the systems A and B of the channel $\mathcal{N}$ respectively. We also refer to the receiver as Charlie. Set

$$\begin{aligned} R_1 &\leftarrow I_H^{\varepsilon'}(A_f : C) + O(\log \varepsilon) \\ R_2 &\leftarrow I_H^{\varepsilon''}(B_f : A_f C) + O(\log \varepsilon) \end{aligned}$$

Resources:

1. Alice possesses a set of messages $[M]$ of size
$$M \leftarrow 2^{R_1}.$$
2. She also shares 2^{R_1} copies of the state $|\varphi_1\rangle^{AA_f}$ with Charlie:
$$|\varphi_1\rangle^{E_{A_1} E_{C_1}} |\varphi_1\rangle^{E_{A_2} E_{C_2}} \ldots |\varphi_1\rangle^{E_{A_{2^{R_1}}} E_{C_{2^{R_1}}}}$$
where
$$A \equiv E_{A_i} \text{ and } A_f \equiv E_{C_i}$$
for all i.
3. Bob possesses a set of messages $[N]$ of size
$$N \leftarrow 2^{R_2}.$$
4. He also shares 2^{R_2} copies of the state $|\varphi_2\rangle^{BB_f}$ with Charlie:
$$|\varphi_2\rangle^{F_{B_1} F_{C_1}} |\varphi_2\rangle^{F_{B_2} F_{C_2}} \ldots |\varphi_2\rangle^{F_{B_{2^{R_2}}} F_{C_{2^{R_2}}}}$$
where
$$B \equiv F_{B_j} \text{ and } B_f \equiv F_{C_j}$$
for all j.

To describe the protocol in terms of the notation that we defined in Fig. 1, consider the following assignments:

$$\psi^{MM_{AB}} \leftarrow \sum_{\substack{i\in[M]\\ j\in[N]}} \frac{1}{MN} \left|m,n\right\rangle \left\langle m,n\right|^{MN} \otimes \left|m,n\right\rangle \left\langle m,n\right|^{M_A N_B}.$$

and

$$\left|\varphi\right\rangle^{E_{AB}E_C} \leftarrow \left(\bigotimes_{i\in[M]} \left|\varphi_1\right\rangle^{E_{A_i}E_{C_i}}\right) \otimes \left(\bigotimes_{j\in[N]} \left|\varphi_2\right\rangle^{F_{B_j}F_{C_j}}\right).$$

One should also note that the encoder of the protocol $\mathcal{E}$ acts on the systems:

$$M_A N_B E_{A_1} \ldots E_{A_M} F_{B_1} \ldots F_{B_N} \to A$$

and the decoder $\mathcal{D}$ acts on

$$C E_{C_1} \ldots E_{C_M} F_{C_1} \ldots F_{C_N} \to \widehat{M}\widehat{N}.$$

We will now give a brief and informal overview of the design of the encoder $\mathcal{E}$ and the decoder $\mathcal{D}$. A detailed description along with the error analysis is provided in Appendix C.2 and C.3.

1. To send the message $m \in [M]$, Alice inputs the contents of the register E_{A_m} into the system A.
2. To send the message $n \in [N]$, Bob inputs the contents of the register F_{B_n} into the system B.
3. To decode, Charlie first disregards the input from Bob as noise and decodes only for Alice.
4. Having successfully decoded Alice's message, Charlie then uses this as side information to decode Bob's message at a higher rate.

It is not hard to see that, for any the channel $\mathcal{N}$ from Alice to Charlie, while averaging over Bob's input can be considered to be:

$$\mathcal{N}_0^{A\to C}(\cdot) := \mathcal{N}^{AB\to C}\left(\cdot \otimes \varphi_2^B\right).$$

With an analysis similar to $\mathcal{N}_0$, we can show that the rate of communication for $\mathcal{N}_1$ is

$$I_H^{\varepsilon'}(A_f : C) + O(\log \varepsilon). \tag{1}$$

Since Anshu-Jain-Warsi protocol decodes correct m with high probability, we can assume that Charlie knows m while decoding Bob's message. In other words, Charlie will decode Bob's message conditioned on Alice's message being m. As usual, the actual state in the protocol may differ from the conditioned state, but gentle measurement lemma guarantees that these states are not far, and the $\mathcal{L}_1$ distance between them can be consumed in the overall error of the protocol. Since, we assume that Alice's message was m, we define a new channel for analysis

$$\mathcal{N}_1^{B\to CE_{C_m}}(\cdot) := \mathcal{N}^{AB\to C}(\varphi_1^{AE_{C_m}} \otimes \cdot).$$

Since the system $E_{C_m} \equiv A_f$ for all m, it holds that, conditioned on correct decoding for Alice, the rate at which Bob can communicate with Charlie is given by

$$I_H^{\varepsilon''}(B_f : CA_f) + O(\log \varepsilon). \tag{2}$$

Thus, the total rate of protocol is given by adding expressions (1) and (2), which is equal to $R_1 + R_2$ by our choice. This completes our proof sketch that there exists a strategy for achieving rates $R_1 + R_2$. $\square$

6 Chain Rules for General Quantum States

In this section we will introduce the ideas required to prove Theorem 1.1. We formally restate the theorem below:

Theorem 6.1. *Given $\varepsilon > 0$ and a tripartite state $\rho^{A_f B_f C}$, it holds that*

$$I_H^{\varepsilon}(A_f B_f : C) \geq I_H^{O(\varepsilon^4)}(A_f : C) + I_H^{O(\varepsilon^4)}(B_f : A_f C) - I_{\max}(A_f : B_f) - \log(1 - O(\varepsilon^{1/4})) - O(1).$$

First, we will need the concept of *quantum rejection sampling* as introduced in [JRS05].

6.1 Quantum Rejection Sampling

The rejection sampling problem can be framed as follows:

1. Consider two distributions P_X and Q_X over some alphabet $\mathcal{X}$, with the assumption that $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$.
2. Alice has access to iid samples from the distribution Q_X.
3. The task is for Alice is to output a letter X_{output}, using only the samples from Q_X and her own private coins, such that $X_{\text{output}} \sim P_X$.
4. We also require that Alice uses as few iid samples from Q_X as possible.

Figure 2: Classical Rejection Sampling

It can be shown that Alice can achieve the above task with $2^{D_{\max}(P_X \,\|\, Q_X)}$ many samples on expectation. In this paper we will require the quantum analog of this problem, which can be stated as follows:

1. Consider two quantum states ρ^A and σ^A such that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$.
2. Alice is provided multiple independent copies of the state σ^A along with ancilla registers as workspace.
3. The task is for her to produce the state ρ^A, while using as few copies of σ^A as possible.

Figure 3: Quantum Rejection Sampling

It can be shown [JRS05] that the above task can be achieved with $2^{D_{\max}(\rho^A \,\|\, \sigma^A)}$ many copies of the state σ^A, on expectation.

Remark 6.2. In fact, if Alice can tolerate some error in the state that she outputs, in the sense that she creates a state $\rho'^A \overset{\varepsilon}{\approx} \rho^A$, then the task can be achieved with $\frac{1}{\varepsilon} \cdot 2^{D^{\varepsilon}_{\max}(\rho^A \,\|\, \sigma^A)}$ copies of σ^A on expectation. However, due to the nature of our protocol, we will require the exact version of this protocol, which requires more copies of ρ^A to work.

The way the protocol works is as follows:

1. Alice possesses multiple iid copies of σ^A.
2. By definition, it holds that
$$\rho^A \leq 2^{D^{\varepsilon}_{\max}(\rho \,\|\, \sigma)} \sigma^A$$
which implies that there exists a quantum state τ^A such that
$$\sigma^A = \frac{1}{2^{D_{\max}}} \rho^A + \left(1 - \frac{1}{2^{D_{\max}}}\right) \tau^A$$

where in the above we used $D_{\max}$ as a shorthand for $D_{\max}(\rho^A \mid\mid \sigma^A)$.

3. Alice uses two registers R and Q to produce a certain purification of σ^A. Here, Q will be a single qubit register:

$$|\sigma\rangle^{ARQ} := \sqrt{\frac{1}{2^{D_{\max}}}}\,|\rho\rangle^{AR}|0\rangle^{Q} + \sqrt{\left(1-\frac{1}{2^{D_{\max}}}\right)}\,|\tau\rangle^{AR}|1\rangle^{Q}$$

   where $|\rho\rangle^{AR}$ and $|\tau\rangle^{AR}$ are purifications of ρ^A and τ^A.

4. Alice performs this purification for a large number of copies of σ^A.

5. Alice then measures the Q register in the computational basis. She gets 0 with probability $1/2^{D_{\max}}$.

6. Discarding the system R completes the protocol.

It is easy to see that the protocol detailed above requires $2^{D_{\max}}$ many copies of σ^A to succeed on expectation. We will require this idea in the following sections.

6.2 The Channel $\mathcal{N}^{AB\to C}$ for General States

The Quantum Rejection Sampling protocol gives us a hint as to how we might go about defining a channel $\mathcal{N}^{AB\to C}$ and some state $|\phi\rangle^{A_fB_fAB}$ such that

$$\mathcal{N}^{AB\to C}\left(\phi^{A_fB_fAB}\right) = \rho^{A_fB_fC}$$

for some fixed $\rho^{A_fB_fC}$. The idea is as follows:

1. Consider the marginals ρ^{A_f} and ρ^{B_f} of $\rho^{A_fB_fC}$ and their purifications $|\varphi_1\rangle^{AA_f}$ and $|\varphi_2\rangle^{BB_f}$ as before.

2. Let Alice have access to the A and B systems of multiple iid copies of $|\varphi_1\rangle^{AA_f}|\varphi_2\rangle^{BB_f}$.

3. Recall that by definition,

$$\rho^{A_fB_f} \le 2^{I_{\max}(A_f:B_f)}\rho^{A_f}\otimes\rho^{B_f}$$

   where

$$I_{\max}(A_f:B_f) = D_{\max}(\rho^{A_fB_f} \mid\mid \rho^{A_f}\otimes\rho^{B_f}).$$

   We will use the shorthand Imax to refer to $I_{\max}(A_f:B_f)$ from here onward.

4. Consider then, the purification $|\varphi\rangle^{AA_fBB_fQ}$ of $\rho^{A_f}\otimes\rho^{B_f}$:

$$|\varphi\rangle^{AA_fBB_fQ} := \sqrt{\frac{1}{2^{\text{Imax}}}}\,|\phi\rangle^{AA_fBB_f}|0\rangle^{Q} + \sqrt{\left(1-\frac{1}{2^{\text{Imax}}}\right)}\,|\tau\rangle^{AA_fBB_f}|1\rangle^{Q}$$

   where $|\phi\rangle^{AA_fBB_f}$ is some purification of $\rho^{A_fB_f}$.

5. Since Alice possesses the A and B systems of the state $|\varphi_1\rangle^{AA_f}|\varphi_2\rangle^{BB_f}$, she can use the Uhlmann isometry $W^{AB\to ABQ}$ to create the state $|\varphi\rangle^{AA_fBB_fQ}$. She does for many copies of the states that she possesses.

6. Now Alice measures the Q register for each copy of the state $|\varphi\rangle^{AA_fBB_fQ}$. On expectation she receives a 0 outcome after 2^{Imax} many measurements.

7. Now, as before, consider an arbitrary purification $|\rho\rangle^{A_fB_fCE}$ of $\rho^{A_fB_fC}$. Then, there exists an Uhlmann isometry $V^{AB\to CE}$ such that

$$V^{AB\to CE}|\phi\rangle^{AA_fBB_f} = |\rho\rangle^{A_fB_fCE}$$

8. Composing the trace out operation on the system E with V gives us the channel $\mathcal{N}^{AB\to C}$.

The above discussion implies the following lemma:

Lemma 6.3. *Given any quantum state $\rho^{A_fB_fC}$, and an arbitrary purification $|\phi\rangle^{A_fB_fAB}$ of the state $\rho^{A_fB_f}$, there exists a channel $\mathcal{N}^{AB\to C}$ such that*

$$\mathbb{I}^{A_fB_f}\otimes\mathcal{N}^{AB\to C}\left(\phi^{A_fB_fAB}\right) = \rho^{A_fB_fC}.$$

6.3 Towards a Proof of Theorem 6.1

In this section we informally describe our strategy to prove Theorem 6.1. First, we fix a purification $|\phi\rangle^{AA_fBB_f}$ of $\rho^{A_fB_f}$ and consider the channel $\mathcal{N}^{AB\to C}$ given by Lemma 6.3. We construct an entanglement assisted communication protocol for this channel using the shared entangled states $|\varphi_1\rangle^{AA_f}$ and $|\varphi_2\rangle^{BB_f}$, which are purifications of ρ^{A_f} and ρ^{B_f} respectively. To do this we make use of two things:

1. The Quantum Rejection Sampling Algorithm.
2. A completely dephasing channel $\mathcal{P}^{X_A\to X_B}$ from Alice to Bob, as an extra resource, which can send $I_{\max}(A_f : B_f) + \log\frac{1}{\delta}$ many bits noiselessly. Here X_A and X_B denote Alice and Bob's classical registers respectively, and both are of size $\frac{1}{\delta}2^{\text{Imax}}$.

The rational behind integrating rejection sampling is simple. Instead of using $[M]$ copies $|\varphi_1\rangle$, we share C_0M copies where C_0 is suitably chosen. The whole protocol is now viewed as having M blocks of size C_0. The Rejection sampling then takes us to a candidate index (say b^*) with certain desired properties (with high probability). The number of such coordinates having index b^* is thus M (one in each block). Restricted over these coordinates, the protocol now has a behavior similar to that of $\mathsf{IM}-\mathtt{state}$ protocol. A similar modification is done for the second step of successive cancellation as well, where C_0N copies of $|\varphi_2\rangle$ are shared. The detailed protocol is as follows:

Table 1: Protocol Modified quantum assisted classical communication (with blocks)

1. We arrange Alice's set of messages as $[M]\times[N]$, where

$$\log M = I_H^{\varepsilon}(A_f : C) + O(\log\varepsilon)$$
$$\log N = I_H^{\varepsilon}(B_f : A_fC) + O(\log\varepsilon).$$

2. Alice shares $M\times\frac{1}{\delta}\cdot 2^{\text{Imax}}$ many copies of the state $|\varphi_1\rangle^{A_fA}$ with Bob. She divides these into M blocks, each of size $\frac{2^{\text{Imax}}}{\delta}$. Blocks are index by $m\in[M]$ and the elements inside a block are further indexed by $b\in\mathcal{B} := \left[\frac{2^{\text{Imax}}}{\delta}\right]$. The corresponding states are therefore represented as

$$\bigotimes_{i\in[M]}\bigotimes_{b\in\mathcal{B}}|\varphi_1\rangle^{A_{f_{b,i}}A_{b,i}}$$

3. Similarly Alice shares $N\times\frac{1}{\delta}\cdot 2^{\text{Imax}}$ many copies of the state $|\varphi_2\rangle^{BB_f}$ with Bob. She divides these into N blocks of size $\frac{1}{\delta}\cdot 2^{\text{Imax}}$ as well. The blocks are indexed analogously. The states then are

$$\bigotimes_{j\in[N]}\bigotimes_{b\in\mathcal{B}}|\varphi_2\rangle^{B_{f_{b,j}}B_{b,j}}$$

4. To send a message (m,n) Alice picks the m-th block of $|\varphi_1\rangle$'s

$$\bigotimes_{b\in\mathcal{B}}|\varphi_1\rangle^{A_{f_{b,m}}A_{b,m}}$$

and the n-th block of $|\varphi_2\rangle$'s

$$\bigotimes_{b\in\mathcal{B}}|\varphi_2\rangle^{B_{f_{b,n}}B_{b,n}}.$$

5. For each $b\in\mathcal{B}$, Alice applies the isometry $W^{AB\to ABQ}$ such that

$$W|\varphi_1\rangle^{A_{f_{b,m}}A_{b,m}}|\varphi_2\rangle^{B_{f_{b,n}}B_{b,n}} = |\varphi\rangle_b^{A_{b,m}A_{f_{b,m}}B_{b,n}B_{f_{b,n}}Q_b}$$

where $|\varphi\rangle_b^{A_{b,m}A_{f_{b,m}}B_{b,n}B_{f_{b,n}}Q_b}$ indicates the b-th copy of $|\varphi\rangle^{AA_fBB_fQ}$.

6. Alice then measures the registers $Q_1 Q_2 \dots Q_{\frac{1}{\delta}2^{\mathrm{Imax}}}$ in a *random order*, and stops the first time the measurement succeeds. By Claim D.1 she gets at least one 0 outcome with probability at least $1-e^{-1/\delta}$. If none of the measurements succeed, Alice aborts.
7. Suppose the index on which the measurement succeeded is b^*. Then, by Appendix D, the distribution of b^* is uniform. Alice then sends the index b^* through the noiseless completely dephasing channel to Bob.
8. Alice also puts the contents of the system $A_{b^*,m}B_{b^*,n}$ into the system AB, i.e., the systems which are input to the channel. She initialises the $A_{b^*,m}B_{b^*,n}$ registers with some junk state.
9. Bob can then simply pick out the b^*-th element in every message block and repeat the successive cancellation decoding procedure as given in Appendix C.2 (see Claim C.9 for details). To be precise, Bob performs his measurements on the collective states

$$\begin{aligned}
&\bigotimes_{\substack{i\in[M],j\in[N]\\ i\neq m,j\neq n}} \varphi_1^{A_{f_{b^*,i}}} \otimes \varphi_2^{B_{f_{b^*,j}}} \bigotimes \mathcal{N}\left(\phi_{b^*}^{ABA_{f_{b^*,m}}B_{f_{b^*,n}}}\right) \\
&= \bigotimes_{\substack{i\in[M],j\in[N]\\ i\neq m,j\neq n}} \rho^{A_{f_{b^*,i}}} \otimes \rho^{B_{f_{b^*,j}}} \bigotimes \rho^{CA_{f_{b^*,m}}B_{f_{b^*,n}}}
\end{aligned}$$

where

$$\rho^{CA_{f_{b^*,m}}B_{f_{b^*,n}}} \equiv \rho^{A_f B_f C}.$$

The above protocol gives us an achievable strategy to send $I_H^{\varepsilon}(A_f : C) + I_H^{\varepsilon}(B_f : A_f C)$ many bits via the channel $\mathcal{N}^{AB\to C} \otimes \mathcal{P}^{X_A\to X_B}$, while making an overall decoding error of at most $28\sqrt{\varepsilon}$ (see Claim C.9 for details). To complete proof we must find a suitable upper bound as given by Corollary 4.3. Before using Corollary 4.3 however, we should point out some subtle issues:

1. The proof of Corollary 4.3 does not consider encoders which can abort the protocol. However, since the quantum rejection sampling procedure can fail with some non-zero probability, we must ensure that Corollary 4.3 can be suitably adapted to this case. In Claim D.2 we extend the proof of Corollary 4.3 to the case when the encoder can toss its own private coins and may abort the protocol with some probability. We show that the results of Corollary 4.3 essentially remain unchanged even in this case.

2. Recall that Corollary 4.3 provides an upper bound on the number of bits any protocol can send through a channel as a function of the state that the encoder of the protocol creates on the input register of the channel, averaged over all messages. For the protocol that we presented in this section, we must find this averaged input state on the system ABX_A. By the arguments in Item 1 above, we are only interested in the case when Alice does not abort. Conditioned on Alice not aborting, the state created on the input system AB of the channel is ρ^{AB}. Note that this state is independent of the index b^* on which the measurement succeeded. Also note that, by Item 7, the distribution on the system X_A, which is input to the completely dephasing channel $\mathcal{P}^{X_A\to X_B}$ is uniform over the size of X_A. Therefore, the state, averaged over all other systems, on the input registers ABX_A of the channel $\mathcal{N}\otimes\mathcal{P}$ is

$$\phi^{AB} \otimes \frac{\mathbb{I}^{X_A}}{\frac{1}{\delta}2^{\mathrm{Imax}}}.$$

where ϕ^{AB} is the marginal of the state $|\phi\rangle^{A_f B_f AB}$.

For ease of notation, let us refer to Protocol 1 as $\mathcal{P}_{\text{ACHIEVABLE}}$. Let us denote by E the event that the quantum rejection sampling phase succeeds, and define

$$\mathcal{P}_{\text{ACHIEVABLE}}|_E$$

be the execution of Protocol 1 conditioned on the event that the quantum rejection sampling phase succeeded. Then, by the discussion above, one can see that

$$\mathcal{P}_{\text{ACHIEVABLE}}|_E \in \mathcal{S}^{\rho^{AB}\otimes\frac{\mathbb{I}^{X_A}}{\frac{1}{\delta}2^{\mathrm{Imax}}}}\left(\mathcal{N}^{AB\to C}\otimes\mathcal{P}^{X_A\to X_B}, 28\sqrt{\varepsilon}\right).$$

Then, by the arguments of Claim D.2, we see that the maximum number of bits that can be transmitted by Protocol 1 is

$$D_{\text{FINAL}} \coloneqq D_H^{28\sqrt{\varepsilon}}\left(\mathcal{N}^{AB\to C}\otimes\mathcal{P}^{X_A\to X_B}\left(\phi^{A_fB_fAB}\otimes\Phi^{RX_A}\right)\,||\,\mathcal{N}(\phi^{AB})\otimes\mathcal{P}\left(\pi^{X_A}\right)\otimes\phi^{A_fB_f}\otimes\pi^R\right)$$

where Φ^{RX_A} is a maximally entangled state on the system X_A and $R\cong X_A$ is the system purifying the maximally mixed state on the system X_A. We use the notation π^{X_A} to denote the maximally mixed state on X_A. Recall that we can make the above statement by Claim D.2 and Corollary 4.3, and the fact that the converse given by those results is true for any arbitrary purification of $\phi^{AB}\otimes\pi^{X_A}$. Note that

$$\mathcal{P}^{X_A\to X_B}\left(\Phi^{RX_A}\right)=\frac{1}{|X_A|}\sum_x|x\rangle\langle x|^R\otimes|x\rangle\langle x|^{X_B}.$$

and

$$\mathcal{P}^{X_A\to X_B}\left(\pi^{X_A}\right)=\pi^{X_B}.$$

Then, by Claim D.4, we can see that

$$D_{\text{FINAL}}\leq D_H^{\sqrt{28}\varepsilon^{1/4}}\left(\mathcal{N}^{AB\to C}\left(\phi^{ABA_fB_f}\right)\,||\,\mathcal{N}^{AB\to C}\left(\phi^{AB}\right)\otimes\phi^{A_fB_f}\right)+\log|X_A|-\log(1-O(\varepsilon^{1/4})).$$

Now recall that

$$\begin{aligned}
&\mathcal{N}^{AB\to C}\left(\phi^{ABA_fB_f}\right)=\rho^{A_fB_fC}\\
&\mathcal{N}^{AB\to C}\left(\phi^{AB}\right)=\rho^C\\
&\phi^{A_fB_f}=\rho^{A_fB_f}\\
&\log|X_A|=I_{\max}(A_f:B_f)_{\rho^{A_fB_f}}+\log\frac{1}{\delta}.
\end{aligned}$$

Collating all these arguments together, we see that

$$\begin{aligned}
I_H^{\varepsilon}(A_F:C)+I_H^{\varepsilon}(B_F:A_FC)&\leq D_{\text{FINAL}}\\
&\leq D_H^{\sqrt{28}\varepsilon^{1/4}}(\rho^{A_fB_fC}\,||\,\rho^{A_fB_f}\otimes\rho^C)+I_{\max}(A_f:B_f)+\log(1-O(\varepsilon^{1/4}))+\log\frac{1}{\delta}\\
&=I_H^{\sqrt{28}\varepsilon^{1/4}}(A_fB_f:C)+I_{\max}(A_f:B_f)+\log(1-O(\varepsilon^{1/4}))+\log\frac{1}{\delta}.
\end{aligned}$$

Finally, we rearrange terms in the above inequality, while setting $\varepsilon\leftarrow\varepsilon^4$ and using the fact that $\log\frac{1}{\delta}=O(1)$. This concludes the proof of Theorem 6.1.

References

[ADJ17] Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. Quantum communication using coherent rejection sampling. *Phys. Rev. Lett.*, 119:120506, Sep 2017.

[Ahl71] R. Ahlswede. Multi-way communication channels. In *Proceedings of 2nd International Symposium on Information Theory (ISIT)*, pages 23–52, 1971.

[Ahl74] R. Ahlswede. The capacity region of a channel with two senders and two receivers. *The annals of probability*, 2(5):805–814, 1974.

[AJW18] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. A one-shot achievability result for quantum state redistribution. *IEEE Transactions on Information Theory*, 64(3):1425–1435, 2018.

[AJW19a] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. Building blocks for communication over noisy quantum networks. *IEEE Transactions on Information Theory*, 65(2):1287–1306, 2019.

[AJW19b] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. On the near-optimality of one-shot classical communication over quantum channels. *Journal of Mathematical Physics*, 60(1):012204, 2019.

[BCR11] Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse shannon theorem based on one-shot information theory. *Communications in Mathematical Physics*, 306(3):579, Aug 2011.

[Ber09] Mario Berta. Single-shot quantum state merging. *ArXiv: 0912.4495*, 2009.

[CNS21] Sayantan Chakraborty, Aditya Nema, and Pranab Sen. One-shot inner bounds for sending private classical information over a quantum mac. In *2021 IEEE Information Theory Workshop (ITW)*, pages 1–6, 2021.

[Dat09] Nilanjana Datta. Min- and max-relative entropies and a new entanglement monotone. *IEEE Transactions on Information Theory*, 55(6):2816–2826, 2009.

[DBWR14] Frédéric Dupuis, Mario Berta, Jürg Wullschleger, and Renato Renner. One-shot decoupling. *Communications in Mathematical Physics*, 328(1):251–284, May 2014.

[DKF$^+$] F. Dupuis, L. Krãmer, P. Faist, J. M. Renes, and R. Renner. *Generalized Entropies*, pages 134–153. XVIIth International Congress on Mathematical Physics.

[Dup10] Frédéric Dupuis. The decoupling approach to quantum information theory. *ArXiv: 1004.1641*, 2010.

[Dup15] Frédéric Dupuis. Chain rules for quantum rényi entropies. *Journal of Mathematical Physics*, 56(2):022203, 2015.

[HOW07] Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269(1):107–136, Jan 2007.

[Jai21] Rahul Jain. Chain-rules for channel capacity. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 262–267, 2021.

[JRP03] R. Jain, J. Radhakrishnan, and Sen P. A lower bound for the bounded round quantum communication complexity of set disjointness. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 220–229, 2003.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Automata, Languages and Programming*, pages 300–315, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[JRS05] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, CCC '05, page 285–296, USA, 2005. IEEE Computer Society.

[Lia72] H. Liao. *Multiple access channels*. PhD thesis, University of Hawai, 1972.

[Raz92] A.A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[Ren05] Renato Renner. *Communication in Networks for Coordinating Behavior*. PhD thesis, 2005. Available onine: https://arxiv.org/abs/quant-ph/0512258.

[RK05] Renato Renner and Robert König. Universally composable privacy amplification against quantum adversaries. In Joe Kilian, editor, *Theory of Cryptography*, pages 407–425, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[RR11a] Joseph M. Renes and Renato Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Transactions on Information Theory*, 57(11):7377–7385, 2011.

[RR11b] Joseph M. Renes and Renato Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Transactions on Information Theory*, 57(11):7377–7385, 2011.

[RW04] R. Renner and S. Wolf. Smooth renyi entropy and applications. In *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, pages 233–, 2004.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[SW73] David Slepian and Jack Keil Wolf†. A coding theorem for multiple access channels with correlated sources. *The Bell System Technical Journal*, 52(7):1037–1076, 1973.

[TCR09] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, 2009.

[TCR10] Marco Tomamichel, Roger Colbeck, and Renato Renner. Duality between smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 56(9):4674–4681, 2010.

[Uhl76] A. Uhlmann. The "transition probability" in the state space of a *-algebra. *Rep. Math. Phys.*, 9:273–279, 1976.

[VDTR13] Alexander Vitanov, Frédéric Dupuis, Marco Tomamichel, and Renato Renner. Chain rules for smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 59(5):2603–2612, 2013.

[WR12] Ligong Wang and Renato Renner. One-shot classical-quantum capacity and hypothesis testing. *Phys. Rev. Lett.*, 108:200501, May 2012.

A Preliminaries

A.1 Definitions

Definition A.1. (Smooth Hypothesis Testing Relative Entropy) *The smooth min-relative entropy D_H^{ε} between two states ρ and σ is defined via the equation below:*

$$2^{-D_H^{\varepsilon}(\rho \,||\, \sigma)} := \min_{\substack{0 \leq \Pi \leq \mathbb{I} \\ \operatorname{Tr}[\Pi\, \rho] \geq 1-\varepsilon}} \operatorname{Tr}(\Pi\, \sigma)$$

Using the usual correspondence between entropy and mutual information, one can define smooth Hypothesis testing mutual information in a state ρ;

$$I_H^{\varepsilon}(A:B)_{\rho} = D_H^{\varepsilon}\left(\rho^{AB} \| \rho^A \otimes \rho^B\right).$$

Given the context of our work, we will be mostly interested in smooth Hypothesis testing mutual information of a particular state associated with a channel.

Definition A.2 (An optimal tester for $(I, \varepsilon, \rho, \mathcal{N})$). *Let $\mathcal{N}^{A \to B}$ be a channel and ρ^{AC} be a pure state. Then,*

$$\begin{aligned}
& I_H^{\varepsilon}(B:C)_{\mathcal{N}(\rho^{AC})} \\
& = D_H^{\varepsilon}\left(\mathcal{N}\left(\rho^{AC}\right) \| \mathcal{N}\left(\rho^{A}\right) \otimes \rho^{C}\right) \\
& = -\log \min_{\substack{0 \leq \Pi \leq \mathbb{I}^{BC} \\ \operatorname{Tr}\left[\Pi\left(\mathcal{N}(\rho^{AC})\right)\right] \geq 1-\varepsilon}} \operatorname{Tr}\left[\Pi\, \left(\mathcal{N}\left(\rho^{A}\right) \otimes \rho^{C}\right)\right].
\end{aligned}$$

A Π that achieves the optimum in the above equation will be referred to as an optimal tester for $(I, \varepsilon, \rho, \mathcal{N})$.

Thus, from definition, it follows that, if Π is an optimal tester for $(I, \varepsilon, \rho, \mathcal{N})$ then,

$$2^{-I_H^{\varepsilon}(B:C)_{\mathcal{N}(\rho^{AC})}} = \operatorname{Tr}\left[\Pi\left(\mathcal{N}\left(\rho^{A}\right) \otimes \rho^{C}\right)\right] \tag{3}$$

$$\operatorname{Tr}\left[\Pi\left(\mathcal{N}(\rho^{AC})\right)\right] \geq 1-\varepsilon. \tag{4}$$

Definition A.3 (Max Relative Entropy). *Given quantum states ρ and σ such that* $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$*, the max relative entropy $D_{\max}(\rho \,||\, \sigma)$ is defined as*

$$D_{\max} := \inf \left\{\lambda \mid \rho \leq 2^{\lambda}\sigma\right\}.$$

Again, using the usual correspondence between entropy and mutual information, one can define the max mutual information with respect to a state ρ^{AB} as:

$$I_{\max}(A:B)_{\rho^{AB}} := D_{\max}(\rho^{AB} \,||\, \rho^A \otimes \rho^B)$$

Definition A.4 (Smooth Max Relative Entropy). *Given quantum states ρ and σ such that* $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$*, let $\mathcal{B}^{\varepsilon}(\rho)$ be the ε ball around the state ρ in the sense that*

$$\mathcal{B}^{\varepsilon}(\rho) := \{\tau \geq 0 \mid \|\tau - \rho\| \leq \varepsilon, \operatorname{Tr}[\tau] \leq 1\}$$

Then the smooth max relative entropy $D_{\max}^{\varepsilon}(\rho \,||\, \sigma)$ is defined as

$$D_{\max}^{\varepsilon}(\rho \,||\, \sigma) := \inf_{\rho' \in \mathcal{B}^{\varepsilon}(\rho)} D_{\max}(\rho' \,||\, \sigma).$$

Similarly, the smooth max mutual information with respect to a state ρ^{AB} is defined as:

$$I_{\max}^{\varepsilon}(A:B)_{\rho^{AB}} := \inf_{\rho'^{AB} \in \mathcal{B}^{\varepsilon}(\rho^{AB})} I_{\max}(A:B)_{\rho'^{AB}}.$$

A.2 Facts

Fact A.5 (Gentle Measurement Lemma). *Let ρ be a state and $\{\Lambda_i\}_i$ be a POVM such that there exists an i_0 with*

$$\operatorname{Tr}\left(\Pi_{i_0}\rho\right) \geq 1-\varepsilon.$$

Let

$$\rho' = \sum_i \sqrt{\Lambda_i}\rho\sqrt{\Lambda_i} \otimes |i\rangle\langle i|$$

be the post measurement state. Then,

$$\|\rho \otimes |i_0\rangle\langle i_0| - \rho'\|_1 \leq 3\sqrt{\varepsilon}.$$

Fact A.6 (Uhlmann's Theorem [Uhl76]). *Let $\rho^A \in \mathcal{D}(\mathcal{H}_A)$ be a state and let $\rho^{AB} \in \mathcal{D}(\mathcal{H}_{AB}),\ \rho^{AC} \in \mathcal{D}\left(\mathcal{H}_{AC}\right)$ be purifications of ρ_A. Then there exists an isometry $V^{C\to B}$ (from a subspace of $\mathcal{H}_C$ to a subspace of $\mathcal{H}_B$) such that,*

$$\mathbb{I}_A \otimes V^{C\to B}\left(\rho^{AC}\right) = \rho^{AB}.$$

Fact A.7 (Closeness [WR12, AJW19b, Fact 9]). *Let $\phi^{MM'}$ be a quantum state that satisfies the conditions*

$$\begin{aligned}
&\phi^M = \frac{\mathbb{I}}{|M|}\\
&\operatorname{Tr}\left[\sum_m |m\rangle\langle m|^M \otimes |m\rangle\langle m|^{M'} \phi^{MM'}\right] \geq 1-\varepsilon
\end{aligned}$$

Then for any quantum state $\sigma^{M'}$, it holds that

$$D_H^\varepsilon(\phi^{MM'} \,||\, \phi^M \otimes \sigma^{M'}) \geq \log|M|.$$

B Proof of Lemma 4.2

Proof. Recall that the the state after Alice encodes is given by,

$$\rho^{ME_BA} = \mathcal{E}\ \left(\psi^{MM_A} \otimes \varphi^{E_AE_B}\right).$$

where ψ and φ are as per protocol 1. Consider an arbitrary purification $|\hat{\tau}\rangle^{ME_BAF}$ of ρ^{ME_BA} and define

$$\rho^{ME_BBF} := \mathcal{N}^{A\to B}\left(|\hat{\tau}\rangle\langle\hat{\tau}|^{ME_BAF}\right) \tag{5}$$

$$\phi^{MM'} := \mathbb{I}_M \otimes \mathcal{D}\left(\rho^{MBE_B}\right). \tag{6}$$

Since $\mathcal{P} \in \mathcal{S}^{\rho^A}(\mathcal{N},\varepsilon)$, it holds that

$$\begin{aligned}
&\operatorname{Tr}_{ME_B}\left(\rho^{ME_BA}\right) = \rho^A\\
&\phi^M = \rho^M = \psi^M = \frac{\mathbb{I}}{M}\\
&\mathsf{Er}(\mathcal{P}) = \left\|\phi^{MM'} - \frac{1}{M}\sum_m |m\rangle\langle m|^M \otimes |m\rangle\langle m|^{M'}\right\|_1\\
&\qquad \leq \varepsilon.
\end{aligned}$$

Consider the projector

$$\Pi^{MM'} := \sum_m |m\rangle\langle m|^M \otimes |m\rangle\langle m|^{M'}.$$

It is then easy to see that

$$\operatorname{Tr}\left[\Pi^{MM'}\phi^{MM'}\right] \geq 1-\varepsilon.$$

Thus, since $\phi^{MM'}$ satisfies all the conditions of Fact A.7, we see that

$$\log|M| \leq D_H^{\varepsilon}(\phi^{MM'} \mid\mid \phi^M \otimes \mathcal{D}\left(\sigma^B \otimes \rho^{E_B}\right)).$$

Each of the following inequality is a straightforward application of the data processing inequality.

$$\begin{aligned}\log|M| &\leq D_H^{\varepsilon}\left(\phi^{MM'} \mid\mid \phi^M \otimes \mathcal{D}\left(\sigma^B \otimes \rho^{E_B}\right)\right)\\ &\leq D_H^{\varepsilon}\left(\rho^{MBE_B} \mid\mid \rho^M \otimes \sigma^B \otimes \rho^{E_B}\right)\end{aligned}$$

... From Eq. (6) and $\rho^M = \phi^M$

$$\begin{aligned}&= D_H^{\varepsilon}\left(\rho^{MBE_B} \mid\mid \rho^{ME_B} \otimes \sigma^B\right)\\ &\leq D_H^{\varepsilon}\left(\rho^{MBE_BF} \mid\mid \rho^{ME_BF} \otimes \sigma^B\right)\\ &= D_H^{\varepsilon}\left(\mathcal{N}(\hat{\tau}^{MAE_BF}) \mid\mid \hat{\tau}^{ME_BF} \otimes \sigma^B\right).\end{aligned}$$

The first equality follows from $\rho^{ME_B} = \rho^M \otimes \rho^B$; whereas the last equality follows from $\hat{\tau} = \mathcal{N}(\rho)$ and noting that $\mathcal{N}$ does not act on any of the registers ME_BF. Thus, $\rho^{ME_BF} = \hat{\tau}^{ME_BF}$.
Now, one can consider any purification $\tau^{AB'}$ of ρ^A. By Uhlamann's theorem, there exists an isometry $V^{ME_BF\to B'}$ such that $V^{ME_BF\to B'}(\hat{\tau}) = \tau$. It then follows that

$$\begin{aligned}&D_H^{\varepsilon}\left(\mathcal{N}(\tau^{AB'}) \mid\mid \tau^{B'} \otimes \sigma^B\right)\\ =\ &D_H^{\varepsilon}\left(\mathcal{N}(\hat{\tau}^{MAE_BF}) \mid\mid \hat{\tau}^{ME_BF} \otimes \sigma^B\right) \geq \log|M|.\end{aligned}$$

This concludes the proof. □

C Achievable Strategies

C.1 The Anshu-Jain-Warsi Protocol

In this section we will recall the one-shot entanglement assisted classical message transmission protocol due to Anshu, Jain and Warsi [AJW19a], which we abbreviate as the AJW protocol. The protocol proceeds as follows:

1. We are given a point to point channel $\mathcal{N}^{A\to B}$ and a starting state ψ^{MM_A} held by the sender Alice;

$$\psi^{MM_A} = \frac{1}{2^R}\sum_{m\in[2^R]} |m\rangle\langle m|^M \otimes |m\rangle\langle m|^{M_A}.$$

2. Sender Alice and receiver Bob share 2^R copies of some pure state $|\varphi\rangle^{E_AE_B}$ as follows,

$$|\varphi\rangle^{E_{A_1}E_{B_1}}|\varphi\rangle^{E_{A_2}E_{B_2}}\ldots|\varphi\rangle^{E_{A_{2^R}}E_{B_{2^R}}}$$

   where the systems E_{A_i} belong to Alice and E_{B_i} belong to Bob.

3. Let Alice prepare some junk state $|\text{JUNK}\rangle^A$, where the system A is isomorphic to E_A.

4. The classical message that Alice wants to send is stored in the register M_A. Suppose Alice wants to send message m. Then Alice swaps systems E_{A_m} and A, followed by the action of $\mathcal{N}^{A\to B}$. To be precise, Alice acts the controlled unitary

$$\sum_{m\in[2^R]} |m\rangle\langle m|^{M_A} \otimes \mathrm{SWAP}^{E_{A_m}A}$$

   on the systems $M_AE_{A_1}E_{A_2}\ldots E_{A_{2^R}}A$. And then applies the channel $\mathcal{N}^{A\to B}$.

5. Let Π be an optimal tester for $\left(I,\varepsilon,\varphi^{AE_B},\mathcal{N}\right)$ (with $\varphi^{AE_B} \equiv \varphi^{E_AE_B}$). Consider a set (indexed by m) of projectors acting jointly on the registers $BE_{B_1}E_{B_2}\dots E_{B_{2R}}$;

$$\Lambda_m = \mathbb{I}^{E_{B_1}} \otimes \cdots \otimes \mathbb{I}^{E_{B_{m-1}}} \otimes \Pi^{BE_{B_m}} \otimes \mathbb{I}^{E_{B_{m+1}}} \otimes \cdots \otimes \mathbb{I}^{E_{B_{2R}}}.$$

Furthermore, using $\{\Lambda_m\}_{m\in[2^R]}$ define a POVM as follows:

$$\Omega_m = \left(\sum_i \Lambda_i\right)^{-\frac{1}{2}} \Lambda_m \left(\sum_i \Lambda_i\right)^{-\frac{1}{2}}.$$

This is the standard *PGM* construction of Ω_m out of Λ_m. To decode, Bob simply measures with the POVM $\{\Omega_m\}_m$. The output of the POVM is represented by $\widehat{M}$ and the state at the end of protocol is denoted by Θ_{END}.

The above protocol has error at most ε [AJW19a], stated by the fact below.

Fact C.1. *[AJW19a] For any*

$$R \leq I_H^{\varepsilon}(E_B : B)_{\mathcal{N}\left(|\varphi\rangle\langle\varphi|^{AE_B}\right)} - 2\log\left(\frac{1}{\varepsilon}\right),$$

where

$$|\varphi\rangle^{AE_B} \equiv |\varphi\rangle^{E_AE_B},$$

we have,

$$\Pr\left(\widehat{M} \neq m | M = m\right)_{\Theta_{\mathsf{END}}} \leq 16\varepsilon.$$

More formally, for all $i \in [m]$, let

$$\Theta_m = \bigotimes_{i\neq m} \varphi^{E_{B_i}} \otimes \mathcal{N}^{A\to B}\left(\varphi^{AE_{B_m}}\right).$$

Then, $\mathrm{Tr}\left(\Omega_m\Theta_m\right) \geq 1 - 16\varepsilon$.

Before going forward, we will need one additional observation about this protocol, which we state in the claim below. The proof of the claim is fairly straightforward, and follows from construction. We include it for the sake of completeness.

Claim C.2. *The state produced by the protocol above on register A (the input register for the channel), averaged over all messages, is φ^A.*

Proof. Firstly, recall that Alice will act her encoder on the M_A register of the the state

$$\psi^{MM_A} = \frac{1}{2^R}\sum_{m\in[2^R]} |m\rangle\langle m|^M \otimes |m\rangle\langle m|^{M_A}$$

and the systems $E_{A_1}\dots E_{A_{2R}}$ of the shared states

$$|\varphi\rangle^{E_{A_1}E_{B_1}} |\varphi\rangle^{E_{A_2}E_{B_2}} \dots |\varphi\rangle^{E_{A_{2R}}E_{B_{2R}}}.$$

It is easy to see that after the encoding, the global state on all systems is as follows:

$$\frac{1}{2^R}\sum_{m\in[2^R]} |m\rangle\langle m|^M \otimes |m\rangle\langle m|^{M_A} \otimes \left(|\varphi\rangle\langle\varphi|^{AE_{B_m}}\right) \otimes \bigotimes_{i\neq m} |\varphi\rangle\langle\varphi|^{E_{A_i}E_{B_i}} \otimes |\mathrm{JUNK}\rangle\langle\mathrm{JUNK}|^{E_{A_m}}$$

Tracing out all the registers except A, we see that the marginal on register A is φ^A. This proves the claim. ∎

C.2 A Multi-Party Generalisation

Consider the following scenario: Given a channel $\mathcal{N}^{AB\to C}$, and pure states $|\varphi_1\rangle^{E_A E_C}$ and $|\varphi_2\rangle^{F_B F_C}$. Let the sender Alice and receiver Charlie share 2^{R_1} copies of $|\varphi_1\rangle$ as

$$|\varphi_1\rangle^{E_{A_i} E_{C_i}}, \text{ where } i \in \left[2^{R_1}\right].$$

where Alice possesses the registers E_{A_i} and Charlie possesses the systems E_{C_i}. Similarly, the second sender Bob and receiver Charlie share 2^{R_2} copies of the state $|\varphi_2\rangle^{F_B F_C}$ as

$$|\varphi_2\rangle^{F_{B_i} F_{C_i}} \text{ where } i \in \left[2^{R_2}\right].$$

where Bob possesses the systems F_{B_i} and Charlie the systems F_{C_i}. To send the message pair $(m,n) \in [2^{R_1}] \times [2^{R_2}]$, Alice and Bob do the following protocol:

1. Alice prepares a junk state in the system A, as $|\text{JUNK}\rangle^A$ and similarly Bob prepares $|\text{JUNK}\rangle^B$.
2. Alice swaps the contents of A with E_{A_m} and Bob swaps the contents of B with F_{B_n}. These operations can be expressed by the following controlled unitary maps:

$$\mathsf{Enc}_A = \sum_{m\in[2^{R_1}]} |m\rangle\langle m|^{M_A} \otimes \text{SWAP}^{E_{A_m} A}$$
$$\mathsf{Enc}_B = \sum_{n\in[2^{R_2}]} |n\rangle\langle n|^{M_B} \otimes \text{SWAP}^{F_{B_n} B}$$

3. The senders then send the systems AB through the channel.

C.3 The Decoding Procedure

Charlie performs the decoding in two phases:

1. First where Charlie decodes Alice's message assuming *nothing* about Bob's message. In this step, Charlie outputs a candidate $\hat{m}$, for Alice's message. To do this, he uses a POVM $\{\Omega_{1,m}\}_m$.
2. In the second step, Charlie outputs a candidate message $\hat{n}$ for Bob's message, assuming that Alice had sent $\hat{m}$. For the decoder, we need to define two POVMs, one each for outputting $\hat{m}$ and $\hat{n}$. He does this using a POVM $\{\Omega_{2,n}\}_n$.
3. The POVMs $\{\Omega_{1,m}\}$ and $\{\Omega_{2,n}\}$ are defined explicitly later.

Figure 4: Multiparty Decoding

Claim C.3. *For any*

$$R_1 \le I_H^{\varepsilon}(E_C : C)_{\mathcal{N}\left(\varphi_1^{AE_C}\otimes\varphi_2^{BF_C}\right)} - 2\log\left(\frac{1}{\varepsilon}\right),$$
$$R_2 \le I_H^{\varepsilon}(F_C : CE_C)_{\mathcal{N}\left(\varphi_1^{AE_C}\otimes\varphi_2^{BF_C}\right)} - 2\log\left(\frac{1}{\varepsilon}\right)$$

where

$$|\varphi_1\rangle^{AE_C} \equiv |\varphi\rangle^{E_A E_C},$$
$$|\varphi_2\rangle^{BE_C} \equiv |\varphi\rangle^{F_B F_C}$$

we have,

$$\Pr\left[\left(\widehat{M},\widehat{N}\right) \neq (m,n) \,|\, (M,N) = (m,n)\right] \le 28\sqrt{\varepsilon}.$$

We defer the proof of this claim to a later point. The proof directly follows from Claim C.8 which itself uses Lemma C.5 as an intermediate step. Throughout the analysis, we will now assume that R_1 and R_2 satisfy the conditions stated by the hypothesis of Claim C.3.

The next section is devoted to proving the above claim. We first focus on Charlie's decoding strategy for Alice and then on his decoding strategy for Bob.

C.3.1 Decoding Alice

Defining the POVM $\{\Omega_{1,m}\}$:

Consider

$$I_H^{\varepsilon}(E_C : C)_{\mathbb{I}^{E_C F_C} \otimes \mathcal{N}^{\mathcal{AB} \to \mathcal{C}}\left(\varphi_1^{AE_C} \otimes \varphi_2^{BF_C}\right)}$$

Let $\Pi_1^{E_C C}$ denote an optimal measurement for the above quantity. That is,

$$\begin{aligned} & 2^{-I_H^{\varepsilon}(E_C:C)_{\mathcal{N}\left(\varphi_1^{AE_C} \otimes \varphi_2^{BF_C}\right)}} \\ &= \operatorname{Tr}\left[\Pi_1^{E_C C}\ \left(\mathcal{N}\left(\varphi_1^{A} \otimes \varphi_2^{B}\right) \otimes \varphi_1^{E_C}\right)\right] \end{aligned}$$

and

$$\operatorname{Tr}\left[\Pi_1^{EcC}\left(\mathcal{N}(\varphi_1^{AE_C} \otimes \varphi_2^{B})\right)\right] \geq 1 - \varepsilon.$$

Let

$$\begin{aligned} \Lambda_{1,m} &= \mathbb{I}^{E_{C_1}} \otimes \cdots \otimes \mathbb{I}^{E_{C_{m-1}}} \otimes \\ &\quad \otimes \Pi^{CE_{C_m}} \otimes \mathbb{I}^{E_{C_{m+1}}} \otimes \cdots \otimes \mathbb{I}^{E_{C_{2^{R_1}}}}, \\ \Omega_{1,m} &= \left(\sum_i \Lambda_{1,i}\right)^{-\frac{1}{2}} \Lambda_{1,m} \left(\sum_i \Lambda_{1,i}\right)^{-\frac{1}{2}}. \end{aligned}$$

Claim C.4. *Define a channel $\mathcal{N}_0^{A \to C}(\sigma^A) := \mathcal{N}^{AB \to C}(\sigma^A \otimes \varphi_2^B)$. Then, $\Pi_1^{E_C C}$ defined above, is an optimal tester for $(I, \varepsilon, \varphi_1^{AE_C}, \mathcal{N}_0)$.*

Proof. The proof follows directly from definition A.2, equation (C.3.1) and (C.3.1) and noting that $\mathcal{N}$ does not act on F_C. □

Claim C.5.

$$\begin{aligned} \operatorname{Tr}\Bigg[&\left(\mathbb{I}^{E_{A_1} \ldots E_{A_M}} \otimes \Omega_{1,m}\right) \mathcal{N}^{AB \to C}\left(\varphi_1^{E_{C_m} A} \otimes \varphi_2^{B}\right) \\ &\bigotimes_{i \neq m} {\varphi_1}^{E_{A_i} E_{C_i}} \otimes \mathrm{JUNK}^{E_{A_m}}\Bigg] \\ &\geq 1 - 16\varepsilon. \end{aligned}$$

Proof. It follows from Claim C.4 and Fact C.1, that, $\operatorname{Tr}\left(\Omega_{1,m}\Theta_{1,m}\right) \geq 1 - 16\varepsilon$, where

$$\begin{aligned} \Theta_{1,m} &= \bigotimes_{i \neq m} \varphi_1^{E_{C_i}} \otimes \mathcal{N}_0^{A \to C}\left(\varphi_1^{AE_{C_m}}\right) \\ &= \bigotimes_{i \neq m} \varphi_1^{E_{C_i}} \otimes \mathcal{N}^{AB \to C}\left(\varphi_1^{AE_{C_m}} \otimes \varphi_2^{B}\right). \end{aligned}$$

The second inequality follows from the definition of $\mathcal{N}_0$. Now,

$$\begin{aligned}
&\operatorname{Tr}\Big[\left(\mathbb{I}^{E_{A_1}\cdots E_{A_M}}\otimes\Omega_{1,m}\right)\mathcal{N}^{AB\to C}\left(\varphi_1^{E_{C_m}A}\otimes\varphi_2^{B}\right)\\
&\bigotimes_{i\neq m}|\varphi_1\rangle\langle\varphi_1|^{E_{A_i}E_{C_i}}\otimes\mathrm{JUNK}^{E_{A_m}}\Big]\\
&=\operatorname{Tr}\left[\Omega_{1,m}\bigotimes_{i\neq m}\varphi_1^{E_{C_i}}\otimes\mathcal{N}^{AB\to C}\left(\varphi_1^{AE_{C_m}}\otimes\varphi_2^{B}\right)\right]\\
&=\operatorname{Tr}\left(\Omega_{1,m}\Theta_{1,m}\right)\\
&\geq 1-16\varepsilon.
\end{aligned}$$

□

Lemma C.6. *Let $\widehat{\Theta}_1$ be the (global) state after step 1(in Protocol 4) and*

$$\begin{aligned}
\Theta_{\text{IDEAL}}\coloneqq|m,n\rangle\langle m,n|^{MN}\otimes\mathcal{N}^{AB\to C}\left(\varphi_1^{E_{C_m}A}\otimes\varphi_2^{F_{C_n}B}\right)\\
\otimes|m\rangle\langle m|^{\widehat{M}}\bigotimes_{i\neq m}\varphi_1{}^{E_{A_i}E_{C_i}}\bigotimes_{j\neq n}\varphi_2{}^{F_{B_j}F_{C_j}}.
\end{aligned}$$

Then,

1. $\mathsf{Pr}\left(\widehat{M}=m|M=m\right)_{\widehat{\Theta}_1}\geq 1-16\varepsilon.$
2. $\|\widehat{\Theta}_1-\Theta_{\text{IDEAL}}\|_1\leq 12\sqrt{\varepsilon}.$

Proof. Suppose Alice wants to send message m and Bob wants to send n. The global joint state just after the encoding can be described as follows:

$$\begin{aligned}
|m,n\rangle\langle m,n|^{MN}\otimes\left(\varphi_1^{E_{C_m}A}\otimes\varphi_2^{F_{C_n}B}\right)\\
\bigotimes_{i\neq m}|\varphi_1\rangle\langle\varphi_1|^{E_{A_i}E_{C_i}}\bigotimes_{j\neq n}|\varphi_2\rangle\langle\varphi_2|^{F_{B_j}F_{C_j}}.
\end{aligned}$$

Recall that while decoding Alice's message, Charlie disregards any of the Bob's register (other than B which is taken as input to the channel). It follows from Claim C.5 that:

$$\begin{aligned}
&\operatorname{Tr}\Big[|m,n\rangle\langle m,n|^{MN}\bigotimes\\
&\quad\left(\Omega_{1,m}\circ\mathcal{N}^{AB\to C}\right)\left(\varphi_1^{E_{C_m}A}\otimes\varphi_2^{F_{C_n}B}\right)\bigotimes_{i\neq m}\\
&\quad|\varphi_1\rangle\langle\varphi_1|^{E_{A_i}E_{C_i}}\bigotimes_{j\neq n}|\varphi_2\rangle\langle\varphi_2|^{F_{B_j}F_{C_j}}\Big]\\
&\geq\ 1-16\varepsilon
\end{aligned}$$

And hence,

$$\Pr\left(\widehat{M}=m\mid M=m\right)_{\widehat{\Theta}_1}\geq 1-16\varepsilon.$$

Then, the Gentle Measurement Lemma (Fact A.5) implies that the post measurement state $\widehat{\Theta}_1$ is close to the ideal state

$$\begin{aligned}
\Theta_{\text{IDEAL}}\coloneqq|m,n\rangle\langle m,n|^{MN}\otimes\mathcal{N}^{AB\to C}\left(\varphi_1^{E_{C_m}A}\otimes\varphi_2^{F_{C_n}B}\right)\\
\otimes|m\rangle\langle m|^{\widehat{M}}\bigotimes_{i\neq m}\varphi_1{}^{E_{A_i}E_{C_i}}\bigotimes_{j\neq n}\varphi_2{}^{F_{B_j}F_{C_j}}
\end{aligned}$$

in the 1-norm by $3\sqrt{16\varepsilon} = 12\sqrt{\varepsilon}$. This concludes the proof. $\square$

C.4 Decoding Bob

Defining the POVM $\{\Omega_{2,n}\}_n$:

Consider

$$I_H^{\varepsilon}(F_C : E_C C)_{\mathbb{I}^{E_C F_C} \otimes \mathcal{N}^{AB \to C}\left(\varphi_1^{AE_C} \otimes \varphi_2^{BF_C}\right)}$$

Let $\Pi_2^{F_C E_C C}$ denote an optimal measurement for the above quantity. That is,

$$\begin{aligned} &2^{-I_H^{\varepsilon}(F_C : E_C C)_{\mathcal{N}\left(\varphi_1^{AE_C} \otimes \varphi_2^{BF_C}\right)}} &(7)\\ &= \mathrm{Tr}\left[\Pi_2^{F_C E_C C}\left(\mathcal{N}\left(\varphi_1^{AE_C} \otimes \varphi_2^{B}\right) \otimes \varphi_2^{F_C}\right)\right] &(8)\end{aligned}$$

and

$$\mathrm{Tr}\left[\Pi_2^{F_C E_C C}\left(\mathcal{N}(\varphi_1^{AE_C} \otimes \varphi_2^{BF_C})\right)\right] \geq 1 - \varepsilon. \qquad (9)$$

Let

$$\begin{aligned}\Lambda_{2,n} = \mathbb{I}^{F_{C_1}} \otimes \cdots \otimes \mathbb{I}^{F_{C_{n-1}}} \otimes \Pi_2^{CE_C F_{C_n}} \\ \otimes \mathbb{I}^{F_{C_{n+1}}} \otimes \cdots \otimes \mathbb{I}^{F_{C_{2^{R_2}}}},\end{aligned}$$

and

$$\Omega_{2,n} = \left(\sum_j \Lambda_{2,j}\right)^{-\frac{1}{2}} \Lambda_{2,n} \left(\sum_i \Lambda_{2,j}\right)^{-\frac{1}{2}}.$$

Claim C.7. *Define a channel $\mathcal{N}_1^{B \to CE_{C_m}}(\sigma^B) := \mathcal{N}^{AB \to C}(\sigma^B \otimes \varphi_1^{AE_{C_m}})$. Then, $\Pi_2^{F_C E_C C}$ defined above, is an optimal tester for $(I, \varepsilon, \varphi_2^{BF_C}, \mathcal{N}_1)$.*

Proof. The proof follows directly from definition A.2, equation (7) and (9) and noting that $\mathcal{N}$ does not act on F_C. $\square$

Claim C.8. *Let $\widehat{\Theta}_2$ be the state at the end of protocol 4. Then it holds that*

$$\left\|\widehat{\Theta}_2^{\widehat{M}\widehat{N}} - |m\rangle\langle m|^{\widehat{M}} \otimes |n\rangle\langle n|^{\widehat{N}}\right\|_1 \leq 28\sqrt{\varepsilon}.$$

Proof. Let $\Theta_{2,\mathrm{IDEAL}}$ be the post measurement state obtained by applying the POVM $\{\Omega_{2,n}\}_n$ to the state Θ_{IDEAL}. By using Fact C.1 on the channel $\mathcal{N}_1$ and the ideal state Θ_{IDEAL}, we see that,

$$\Pr\left[\widehat{N} \neq n \mid N = n, M = m\right]_{\Theta_{2,\mathrm{IDEAL}}} \leq 16\varepsilon. \qquad (10)$$

Now,

$$\begin{aligned}&\Pr\left[\widehat{N} \neq n \mid N = n, M = m\right]_{\widehat{\Theta}_2} \\ &\leq \Pr\left[\widehat{N} \neq n \mid N = n, M = m\right]_{\Theta_{2,\mathrm{IDEAL}}} + \|\Theta_{2,\mathrm{IDEAL}} - \widehat{\Theta}_2\|_1 \\ &\leq 16\varepsilon + \|\Theta_{\mathrm{IDEAL}} - \widehat{\Theta}_1\|_1 \\ &\leq 16\varepsilon + 12\sqrt{\varepsilon} \\ &\leq 28\sqrt{\varepsilon}.\end{aligned}$$

The second inequality follows from eq (10) and data processing. The third inequality follows from Lemma C.6. This concludes the proof. $\square$

C.5 A More General Situation

We note that the decoding procedure outlined in Fig. 4 also works in a more general case, which we describe below: Consider a pure state $|\varphi\rangle^{E_C F_C AB}$ and the purifications $|\varphi_1\rangle^{E_C A}$ and $|\varphi_2\rangle^{F_C B}$ of the states φ^{E_C} and φ^{F_C} respectively. Consider the following situation:

Table 2: General Decoding

1. Fix the index (m, n).
2. Let Alice share 2^{R_1} the states
$$\bigotimes_{i \neq m} |\varphi_1\rangle^{E_{C_i} E_{A_i}}$$
with Charlie, where as before, the systems E_{A_i} belong to Alice and E_{C_i} belong to Charlie. Note also that $E_{A_i} \equiv A$.
3. Similarly, let Bob share 2^{R_2} the states
$$\bigotimes_{j \neq n} |\varphi_1\rangle^{F_{C_j} F_{B_j}}$$
with Charlie, where as before, the systems F_{B_j} belong to Alice and F_{C_j} belong to Charlie. Note also that $F_{B_j} \equiv B$.
4. For $i = m$ and $j = n$, let Alice, Bob and Charlie share the tripartite state
$$\mathcal{N}^{AB \to C}\left(|\varphi\rangle\langle\varphi|^{E_{C_m} F_{C_n} AB}\right)$$
5. Then, to decode the indices m and n, Charlie runs the protocol outlined in Fig. 4, with a suitable setting of decoders $\{\Omega_{1,m}\}$ and $\{\Omega_{2,n}\}$.

Then, the following claim can be proved along similar lines to the proof of Claim C.3:

Claim C.9. *For any*
$$R_1 \leq I_H^{\varepsilon}(E_C : C)_{\mathcal{N}\left(\varphi^{ABE_C F_C}\right)} - 2\log\left(\frac{1}{\varepsilon}\right),$$
$$R_2 \leq I_H^{\varepsilon}(F_C : CE_C)_{\mathcal{N}\left(\varphi^{ABE_C F_C}\right)} - 2\log\left(\frac{1}{\varepsilon}\right)$$
there exist choices for the decoders $\{\Omega_{1,m}\}$ and $\{\Omega_{2,n}\}$ in Procedure Table 2 such that the following holds:
$$\Pr\left[\left(\widehat{M}, \widehat{N}\right) \neq (m, n) \,|\, (M, N) = (m, n)\right] \leq 28\sqrt{\varepsilon}.$$

D Useful Lemmas

Claim D.1. *If Alice measures the Q registers of of the state $|\varphi\rangle^{\otimes n}$, where $n = \frac{1}{\delta} \cdot 2^{\text{Imax}}$, she obtains a string with at least one 0 with probability at least $1 - e^{-1/\delta}$.*

Proof. The probability that Alice gets all 1's is
$$\left(1 - \frac{1}{2^{\text{Imax}}}\right)^n \leq e^{-n/2^{\text{Imax}}} = e^{-1/\delta}.$$
□

Claim D.2. *Consider a protocol $\mathcal{P} = \left(M, \mathcal{N}, \mathcal{E}, \mathcal{D}, |\varphi\rangle^{E_A E_B}\right)$ such that the encoder $\mathcal{E}$ can toss its own private coins and abort the protocol with probability $p < 1$. Suppose we are promised that, whenever the protocol does not abort, it creates the state ρ^A, averaged over all other systems, on the input to the channel. We are also promised that whenever* *the protocol does not abort, the decoder makes an error at most ε while decoding. Then, it holds that, if M is the number of messages that $\mathcal{P}$ can send through the channel, then*

$$\log M \leq I_H^{\varepsilon}\left(B : B'\right)_{\mathcal{N}(\tau^{AB'})}$$

where $|\tau\rangle^{AB'}$ is an arbitrary purification of ρ^A.

Proof. Define the event E to be the set of those coin tosses of the encoder $\mathcal{E}$ when the protocol $\mathcal{P}$ does not abort. Define

$$\mathcal{P}|_E$$

to be the execution of the protocol $\mathcal{P}$ conditioned on the coin tosses in E, i.e., the encoder samples its private coins from a distribution which is supported only on the set E. By the promise given in the statement of the claim, $\mathcal{P}|_E$ creates the state ρ^A on the input to the channel. Therefore it holds that

$$\mathcal{P}|_E \in \mathcal{S}^{\rho^A}(\mathcal{N}, \varepsilon).$$

This implies that the total number of bits that the protocol $\mathcal{P}|_E$ can send, with probability of error at most ε is at most $I_H^{\varepsilon}(B : B')_{\mathcal{N}(\tau^{AB'})}$, by Corollary 4.3. Note however, that the protocol $\mathcal{P}$ does not send any bits when the coin tosses of the encoder land outside of E. Therefore, the total number of bits that the protocol $\mathcal{P}$ can send is at most

$$\log M \leq I_H^{\varepsilon}(B : B')_{\mathcal{N}(\tau^{AB'})}.$$

This concludes the proof. □

Claim D.3. *If Alice measures the registers $Q_1 Q_2 \ldots Q_n$ in random order, where $n = \frac{1}{\delta} 2^{\text{Imax}}$, then, conditioned on getting at least one 0 outcome, it holds that*

$$\Pr\left[\text{ success at } i \mid \text{ success }\right] = \frac{1}{n}.$$

Proof. Fix a permutation σ of the set $[n]$. Suppose

$$\sigma(i) = j$$

i.e., the i-th index is measured at time $j \in [n]$. Then,

$$\Pr\left[\text{ success at time } j \mid \sigma\right] = \left(1 - \frac{1}{2^{\text{Imax}}}\right)^{j-1} \cdot \frac{1}{2^{\text{Imax}}}$$

Then,

$$\begin{aligned}
\Pr\left[\text{ success at } i\right] &= \sum_j \sum_{\sigma \mid \sigma(i)=j} \Pr\left[\text{ success at time } j \mid \sigma\right] \cdot \Pr\left[\sigma \text{ s.t. } \sigma(i) = j\right] \\
&= \sum_j \left(1 - \frac{1}{2^{\text{Imax}}}\right)^{j-1} \cdot \frac{1}{2^{\text{Imax}}} \cdot \frac{1}{n} \\
&= \left(1 - \left(1 - \frac{1}{2^{\text{Imax}}}\right)^n\right) \cdot \frac{1}{n}
\end{aligned}$$

Also, note that

$$\Pr\left[\text{ success }\right] = \left(1 - \left(1 - \frac{1}{2^{\text{Imax}}}\right)^n\right)$$

Therefore,

$$\Pr\left[\text{ success at } i \mid \text{ success }\right] = \frac{1}{n}.$$

This concludes the proof. □

Claim D.4. *Given the states ρ^A, σ^A and a maximally correlated state*

$$\varrho^{X_1 X_2} := \frac{1}{K} \sum_x |x\rangle\langle x|^{X_1} \otimes |x\rangle\langle x|^{X_2},$$

it holds that

$$D_H^{\varepsilon}(\rho^A \otimes \varrho^{X_1 X_2} \,||\, \sigma^A \otimes \pi^{X_1} \otimes \pi^{X_2}) \le D_H^{\sqrt{\varepsilon}}(\rho^A \,||\, \sigma^A) + \log K - \log(1-\sqrt{\varepsilon}).$$

Proof. Let $\Pi_{\text{OPT}}^{AX_1X_2}$ be an optimal tester for $d := D_H^{\varepsilon}(\rho^A \otimes \varrho^{X_1X_2} \,||\, \sigma^A \otimes \pi^{X_1} \otimes \pi^{X_2})$.

That is,

$$\operatorname{Tr}\left[\Pi_{\text{OPT}}^{AX_1X_2}\left(\rho^A \otimes \varrho^{X_1X_2}\right)\right] \ge 1-\varepsilon \tag{11}$$

$$\operatorname{Tr}\left[\Pi_{\text{OPT}}^{AX_1X_2}\left(\sigma^A \otimes \pi^{X_1} \otimes \pi^{X_2}\right)\right] \le 2^{-d} \tag{12}$$

Define

$$\Pi_{x_1,x_2}^A := \left(\mathbb{I}^A \otimes \langle x_1, x_2|^{X_1X_2}\right) \Pi_{\text{OPT}}^{AX_1X_2} \left(\mathbb{I}^A \otimes |x_1, x_2\rangle^{X_1X_2}\right)$$

It is then easy to see that

$$\begin{aligned}
1-\varepsilon &\le \operatorname{Tr}\left[\Pi_{\text{OPT}}^{AX_1X_2}\ \rho^A \otimes \varrho^{X_1X_2}\right] \\
&= \operatorname{Tr}\left[\Pi_{\text{OPT}}^{AX_1X_2}\ \rho^A \otimes \sum_x \frac{1}{K} |x,x\rangle\langle x,x|^{X_1X_2}\right] \\
&= \sum_x \frac{1}{K} \operatorname{Tr}\left[\Pi_{\text{OPT}}^{AX_1X_2}\ \rho^A \otimes |x,x\rangle\langle x,x|^{X_1X_2}\right] \\
&= \sum_x \frac{1}{K} \operatorname{Tr}\left[\left(\langle x,x|^{X_1X_2}\ \Pi_{\text{OPT}}^{AX_1X_2}\ |x,x\rangle^{X_1X_2}\right)^A \rho^A\right] \\
&= \sum_x \frac{1}{K} \operatorname{Tr}\left[\Pi_{x,x}^A \rho^A\right]
\end{aligned}$$

We define a set GOOD_X as follows:

$$\text{GOOD}_X = \{x \ :\ \operatorname{Tr}\left[\Pi_{x,x}^A\ \rho^A\right] \ge 1-\sqrt{\varepsilon}\}.$$

A standard Markov argument then gives that

$$|\text{GOOD}_X| \ge (1-\sqrt{\varepsilon})K.$$

Again, note that

$$\begin{aligned}
\operatorname{Tr}\left[\Pi_{\text{OPT}}^{AX_1X_2}\left(\sigma^A \otimes \pi^{X_1} \otimes \pi^{X_2}\right)\right] &= \sum_{x_1,x_2} \frac{1}{K^2} \operatorname{Tr}\left[\Pi_{\text{OPT}}^{AX_1X_2}\left(\sigma^A \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_1}\right)\right] \\
&= \sum_{x_1,x_2} \frac{1}{K^2} \operatorname{Tr}_A\left[\operatorname{Tr}_{X_1X_2} \Pi_{\text{OPT}}^{AX_1X_2}\left(\sigma^A \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2}\right)\right] \\
&= \sum_{x_1,x_2} \frac{1}{K^2} \operatorname{Tr}_A\left[\langle x_1x_2|^{X_1X_2}\ \Pi_{\text{OPT}}^{AX_1X_2}\ |x_1x_2\rangle^{X_1X_2} \sigma^A\right] \\
&= \sum_{x_1,x_2} \frac{1}{K^2} \operatorname{Tr}_A\left[\Pi_{x_1x_2}^A\ \sigma^A\right] \\
&\ge \sum_{x \in \text{GOOD}_X} \frac{1}{K^2} \operatorname{Tr}\left[\Pi_{x,x}^A\ \sigma^A\right] \\
&\ge \sum_{x \in \text{GOOD}_X} \frac{1}{K^2} 2^{-D_H^{\sqrt{\varepsilon}}(\rho^A \,||\, \sigma^A)} \\
&\ge \frac{1-\sqrt{\varepsilon}}{K} \cdot 2^{-D_H^{\sqrt{\varepsilon}}(\rho^A \,||\, \sigma^A)}
\end{aligned}$$

This concludes the proof. □