A unified approach to source and message compression

Anurag Anshu∗ Rahul Jain† Naqueeb Ahmad Warsi‡

Abstract

We study the problem of source and message compression in the one-shot setting for the point-to-point and multi-party scenario (with and without side information). We derive achievability results for these tasks in a unified manner, using the techniques of convex-split, which was introduced in [1] and position based decoding introduced in [2], which in turn uses hypothesis testing between distributions. These results are in terms of smooth max divergence and smooth hypothesis testing divergence. As a by-product of the tasks studied in this work, we obtain several known source compression results (originally studied in the asymptotic and i.i.d. setting) in the one-shot case.

One of our achievability results includes the problem of message compression with side information, originally studied in [5]. We show that both our result and the result in [5] are near optimal in the one-shot setting by proving a converse bound.

1 Introduction

Source compression is a fundamental task in information theory first studied by Shannon in his landmark paper [8]. This task was later extended to various network settings for example by Slepian and Wolf [9], Wyner [10] and Wyner and Ziv [11]. These works considered the asymptotic, independent and identically distributed (i.i.d.) setting. Compression protocols have been particularly relevant in communication complexity [12, 13], where Alice and Bob wish to compute a joint function of their inputs \(x, y\) (that are sampled from a joint distribution \(p_{XY}\)). Upon receiving her input, Alice sends a message \(M\) to Bob, who sends the next message to Alice conditioned on Alice’s message and Bob’s input. This process continues till both parties have computed the desired function up to some error. Observe that \(M\) and \(Y\) are independent conditioned on \(X\). An important question in communication complexity is to communicate \(M\) with small communication cost (see Task 1 below), which has been investigated by several authors [2, 3, 5, 6] and is connected to the fundamental problem of direct sum. We refer to this question as the task of message compression.

In this work we consider source and message compression in various network communication scenarios and present a unified approach to arrive at communication bounds. Starting from a one-sender-one-receiver task, we consider a two-senders-one-receiver task followed by a one-sender-two-receivers task. These tasks are summarized in Figure 1. We combine these two to consider a two-senders-two-receivers task. It can be observed that our approach extends to more complicated network scenarios as well. We particularly focus on message compression in network scenarios due to growing interest in the problems related to multi-party communication complexity [14, 15, 16, 17].

We present our communication bounds in the one-shot setting and sketch how these bounds behave in the asymptotic i.i.d setting. We leave the question of second order and asymptotic non-i.i.d. analysis of many of these results to future work (second order and asymptotic non-i.i.d. analysis of some of the results has already been achieved in known literature). One-shot information theory has been studied extensively in the recent years both in the classical and quantum models. Apart from being practically relevant (since there is no i.i.d. assumption) it often provides interesting new insights and conceptual advances into the working and design of communication protocols, as the complications and conveniences of the i.i.d assumption are not present. One-shot information theory has been particularly useful in communication complexity while dealing with the important and consequential direct sum, direct product and composition questions.

∗Center for Quantum Technologies, National University of Singapore, Singapore. a0109169@u.nus.edu
†Center for Quantum Technologies, National University of Singapore and MajuLab, UMI 3654, Singapore. rahul@comp.nus.edu.sg
‡Center for Quantum Technologies, National University of Singapore and School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore and IIITD, Delhi. warsi.naqueeb@gmail.com
As applications of our results we reproduce several known results in network communication theory both in the one-shot and i.i.d. settings, further exhibiting the power of our unified framework.

There are two main techniques that we use to arrive at our results. First is the convex-split technique, which was introduced in \cite{1} for a related problem in the quantum domain. Convex-split technique is closely related to the well known rejection sampling technique, used in various information theoretic tasks in several works \cite{2 3 4 5 6}. The other technique that we use is position based decoding introduced in \cite{7}, which in turn uses hypothesis testing between distributions. These two techniques used together allow us to construct all our protocols.

- **Convex-split technique**: Central to this technique is the convex-split lemma \cite{1}, which is a statement of the following form. Let \( p_{AB} \) be a probability distribution over the set \( \mathcal{A} \times \mathcal{B} \), \( p_{B'} \) be a probability distribution (possibly different from \( p_B \)) over the set \( \mathcal{B} \). Then

  \[
  \sum_{b_1, b_2, \ldots, b_{2^R}} p_{B'}(b_1)p_{B'}(b_2) \ldots p_{B'}(b_{2^R}) \sum_{j=1}^{2^R} p_{AB}(ab_j) \leq \varepsilon,
  \]

  if \( R \geq D_{\max}(p_{AB}\|p_A \times p_{B'}) + \log \frac{1}{\varepsilon} \). Above \( D(\|\cdot\cdot\cdot) \) is the relative entropy and \( D_{\max}(\cdot\cdot\cdot) \) is the max divergence. In this work, we shall use a corollary of above result, which is a statement of the form

  \[
  \sum_{b_1, b_2, \ldots, b_{2^R}} p_{B'}(b_1)p_{B'}(b_2) \ldots p_{B'}(b_{2^R}) \sum_{j=1}^{2^R} p_{AB}(ab_j) - p_A \leq O(\varepsilon),
  \]

  if \( R \geq D_{\max}^c(p_{AB}\|p_A \times p_{B'}) \) and \( D_{\max}^c(\cdot\cdot\cdot) \) is the information spectrum relative entropy. Convex-split lemma is reminiscent of \cite{18} Lemma 4.1, which was also independently obtained as the soft-covering lemma in \cite{19} (see also \cite{20, 22, 23} for applications), but there are two points of difference. First is that the former is in terms of relative entropy, whereas the latter is in terms of the variational distance. Second, convex-split lemma accommodates the random variable \( B' \) which is not related to the random variable \( B \), a feature that is not present in the soft-covering lemma. In fact, this feature is essential for our protocol, as we shall show various optimality results using the fact that we can construct protocols using an arbitrary random variable \( B' \).

- **Position-based decoding technique**: This technique uses hypothesis testing to locate the index \( j \) where correlation between random variables \( A \) and \( B \) is according to \( p_{AB} \) in the distribution

  \[
  p_{AB, B_2 \ldots B_{2^R}}(a, b_1, b_2, \ldots, b_{2^R}) = \frac{1}{2^R} \sum_j p_{AB}(a, b_j) \cdot p_{B'}(b_1) \cdot \ldots \cdot p_{B'}(b_{j-1}) \cdot p_{B'}(b_{j+1}) \ldots \cdot p_{B'}(b_{2^R}).
  \]

  The technique succeeds with error \( \varepsilon + \delta \) as long as \( R \leq D_{\max}^c(p_{AB}\|p_A \times p_{B'}) + \log \delta \).

**Our results**

We start with the following one-sender-one-receiver task. For all our results in this section let \( \varepsilon, \delta > 0 \) be sufficiently small constants which represent error parameters.

**Task 1: One-sender-one-receiver message compression with side information at the receiver**. There are two parties Alice and Bob. Alice possesses random variable \( X \), taking values over a finite set \( \mathcal{X} \) (all sets that we consider in this paper are finite) and a random variable \( M \), taking values over a set \( \mathcal{M} \). Bob possesses random variable \( Y \), taking values over a set \( \mathcal{Y} \) such that \( M \) and \( Y \) are independent given \( X \) represented by \( M = X - Y \). Alice sends a message to Bob and at the end Bob outputs random variable \( \hat{M} \) such that \( \frac{1}{2} \| p_{XY \hat{M}} - p_{XY \hat{M}} \|_1 \leq \varepsilon \), where \( \| \cdot \|_1 \) represents the \( \ell_1 \) norm. They are allowed to use shared randomness between them which is independent of \( XY \hat{M} \) at the beginning of the protocol.

As discussed earlier, this task is particularly relevant from the point of view of communication complexity, where \((X, Y)\) can be viewed as inputs given to Alice and Bob respectively from a priori distribution and \( M \) can be viewed as the message Alice wants to send to Bob. It was studied in \cite{2, 4} when the distribution of \((X, Y)\) is product and in \cite{8} for general \((X, Y)\). All these results are in the one-shot setting. It was also studied when the random variable \( Y \) is not present as a side information in the work \cite{23}, in the asymptotic and i.i.d. setting.
Here, we discuss two ways of analyzing Task 1. First is the protocol of Braverman and Rao from [5], who analyzed the expected communication cost of their protocol. We show that their protocol is nearly optimal in present setting (that is the worst case communication cost or the total number of bits communicated from Alice to Bob) in the following theorem. The proof directly follows from Theorem 10.

**Theorem 1** (Achievability and converse for Task 1 using Braverman and Rao’s protocol). Let $\varepsilon, \delta \in (0, 1)$. Let $R$ be a natural number such that,

$$R \geq D^*_p(p_{X|M|Y}\|p_Y(p_{X|Y} \times p_{M|Y})) + O(\log \frac{1}{\delta}),$$

where $p_Y(p_{X|Y} \times p_{M|Y})$ is the probability distribution defined as

$$p_Y(p_{X|Y} \times p_{M|Y})(x, m, y) = p_Y(y)p_{X|Y=y}(x)p_{M|Y=y}(m).$$

There exists a shared randomness assisted protocol in which Alice communicates $R$ bits to Bob and Bob outputs random variable $\hat{M}$ satisfying $\frac{1}{2}\|p_{X|M}\| - p_{X|Y\hat{M}}\| \leq \varepsilon + 4\sqrt{\delta}$. 
Further, any communication protocol for Task 1 must satisfy:

\[ R \geq D^\varepsilon_{s}(p_{X|Y}p_{Y}p_{M|Y}) - O(\varepsilon) \]

where \( R \) is the communication (in bits) between Alice and Bob.

A drawback of the protocol of Braverman and Rao is that it does not seem to generalize to the multi-party setting, where there is a restriction on the shared randomness across the communicating parties (for instance, in the protocol of Slepian and Wolf \[9\], the senders are not allowed to share randomness with each other). Thus, we construct a second way of analyzing Task 1. We construct a new protocol using the aforementioned techniques of convex split and position based decoding and show a new achievable result in Theorem 2.

**Theorem 2** (Achievability for Task 1). Let \( \varepsilon, \delta \in (0, 1) \). Let \( R \) be a natural number such that,

\[ R \geq \min_{E,T} \left( \sum_{i=1}^{T} \frac{D^\varepsilon_{s}(p_{X|Y}p_{Y}p_{M|Y})}{T} - O(\varepsilon) \right) \]

where \( E \) takes values over a set \( E \) and \( T \) takes values over set \( E \times M \). There exists a shared randomness assisted protocol in which Alice communicates \( R \) bits to Bob and Bob outputs random variable \( M \) satisfying \( \frac{1}{2}\|p_{X|Y}p_{M} - p_{X|Y} \| \leq \varepsilon + 4\sqrt{\delta} \).

Please refer to Section 2 for the definitions of \( D^\varepsilon_{s}(\cdot) \) and \( D^\varepsilon_{H}(\cdot) \). The minimization above over \( E \) (which we refer to as an extension of \( M \)) and \( T \) (which is used in shared randomness) may potentially decrease the amount of communication between Alice and Bob. In our converse result below, we show that this is indeed the case. This also establishes the near optimality of our protocol.

**Theorem 3** (Converse for Task 1). Fix a \( \delta \in (0, 1) \). Any communication protocol for Task 1 must satisfy:

\[ R \geq \min_{E,T} \left( \sum_{i=1}^{T} \frac{D^\varepsilon_{s}(p_{X|Y}p_{Y}p_{M|Y})}{T} - O(\varepsilon) \right) \]

where \( R \) is the communication (in bits) between Alice and Bob, \( E \) (taking values in \( E \)) is a specific extension (defined subsequently in the proof of this result) of \( M \) and \( U \) is uniformly distributed over \( M \times \mathcal{E} \).

We highlight two important aspects of above results.

- The condition \( M - X - Y \) is very crucially exploited in our protocol (and in the protocols given in aforementioned works for this task). In the case where \( M, X, Y \) do not form a Markov chain, one needs to optimize over a new random variable \( V \) that satisfies the conditions \( V - MX - Y, Y - V Y - M \). Owing to the lack of a better understanding of the random variable \( V \), we do not pursue this case further in present work.

- The achievability and converse results given above converge (in terms of rate of communication) to the conditional mutual information \( I(X : M | Y) \) in the asymptotic and i.i.d. setting. For this, we use the asymptotic i.i.d. analysis of information spectrum relative entropy given in \[24\] to conclude that the rate of communication is equal to \( D(p_{X|Y}p_{Y}p_{M|Y}) \) which evaluates to \( I(X : M | Y) \) by direct analysis.

- The quantum analogue of Task 1 is the problem of quantum state redistribution \[25\] \[26\]. To motivate this analogy, observe that the Task 1 captures one round of communication in a classical communication protocol and quantum state redistribution captures one round of communication in a quantum communication protocol \[27\]. Theorems 1, 2 and 3 give a near optimal one-shot result for Task 1. On the other hand, a similar result for quantum state redistribution is unknown despite several recent efforts \[28\] \[29\] \[1\] \[30\] and is one of the major open problems in quantum information theory.

Next we consider the following two-senders one-receiver task.

**Task 2:** Two-senders-one-receiver message compression. There are three parties Alice, Bob and Charlie. Alice holds a random variable pair \( (X, M) \) and Bob holds a random variable pair \( (Y, N) \) such that \( M - X - Y - N \). Alice wants to communicate \( M \) to Charlie and Bob wants to communicate \( N \) to Charlie. Alice and Bob send a message each to Charlie and at the end Charlie outputs \( M, N \) such that \( \frac{1}{2} \| p_{X|Y}p_{M|Y} - p_{X|Y}p_{M|Y} \| \leq O(\varepsilon + \sqrt{\delta}) \). Shared randomness is allowed between Alice and Charlie and between Bob and Charlie.

We show the following achievability result for this task.
Theorem 4 (Achievability for Task 2). Let $S, T$ be random variables taking values over the same sets as $M, N$ respectively. Let $R_A, R_B$ be natural numbers such that,

$$R_A \geq D^s_X(p_{XM} || p_X \times p_S) - D^s_H(p_{MN} || p_S \times p_N) + O\left(\frac{\log 1}{\delta}\right),$$

$$R_B \geq D^s_Y(p_{YN} || p_Y \times p_T) - D^s_H(p_{MN} || p_M \times p_T) + O\left(\frac{\log 1}{\delta}\right),$$

$$R_A + R_B \geq D^s_X(p_{XM} || p_X \times p_S) + D^s_Y(p_{YN} || p_Y \times p_T) - D^s_H(p_{MN} || p_S \times p_T) + O\left(\frac{\log 1}{\delta}\right).$$

There exists a shared randomness assisted protocol with communication $R_A$ bits from Alice to Charlie and $R_B$ bits from Bob to Charlie, in which Charlie outputs random variable pair $(\hat{M}, \hat{N})$ such that $\frac{1}{2}\|p_{XYMN} - p_{XY\hat{M}\hat{N}}\| \leq 3\varepsilon + 9\sqrt{\delta}$.

Remark: In our result above we can optimize over extensions $E$ as in Theorem 2. However we skip explicit mention of this optimization for ease of exposition and for brevity, both in the statement above and in its proof. We do the same for all the results later in this section.

Next we consider the same task but with side information with Charlie.

Task 3: Two-senders-one-receiver message compression with side information at the receiver. There are three parties Alice, Bob and Charlie. Alice holds a random variable pair $(X, M)$, Bob holds a random variable pair $(Y, N)$ and Charlie holds a random variable $Z$ such that $M - X = (Y, Z)$ and $N - Y = (X, Z)$. Alice and Bob send a message each to Charlie and at the end Charlie outputs random variable pair $(\hat{M}, \hat{N})$ such that $\frac{1}{2}\|p_{XYZMN} - p_{XY\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$. Shared randomness is allowed between Alice and Charlie and between Bob and Charlie.

We show the following achievability result for this task.

Theorem 5 (Achievability for Task 3). Let $S, T$ be random variables taking values over the same sets as $M, N$ respectively. Let $R_A, R_B$ be natural numbers such that,

$$R_A \geq D^s_X(p_{XM} || p_X \times p_S) - D^s_H(p_{MNZ} || p_S \times p_NZ) + O\left(\frac{\log 1}{\varepsilon}\right),$$

$$R_B \geq D^s_Y(p_{YN} || p_Y \times p_T) - D^s_H(p_{MZ} || p_MZ \times p_T) + O\left(\frac{\log 1}{\varepsilon}\right),$$

$$R_A + R_B \geq D^s_X(p_{XM} || p_X \times p_S) + D^s_Y(p_{YN} || p_Y \times p_T) - D^s_H(p_{MNZ} || p_S \times p_T \times p_Z) + O\left(\frac{\log 1}{\varepsilon}\right).$$

There exists a shared randomness assisted protocol with communication $R_A$ bits from Alice to Charlie and $R_B$ bits from Bob to Charlie, in which Charlie outputs random variable pair $(\hat{M}, \hat{N})$ such that $\frac{1}{2}\|p_{XYZMN} - p_{XY\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$.

Next we consider the following one-sender-two-receivers task.

Task 4: One-sender-two-receivers message compression. There are three parties Alice, Bob and Charlie. Alice holds correlated random variables $(X, M, N)$. She sends a message to Bob and a message to Charlie. Bob and Charlie after receiving their respective messages, output random variables $\hat{M}$ and $\hat{N}$ respectively such that $\frac{1}{2}\|p_{XMN} - p_{X\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$. Shared randomness is allowed between Alice and Charlie and between Alice and Bob.

We show the following achievability result for this task.

Theorem 6 (Achievability for Task 4). Let $S, T$ be random variables taking values over the same sets as $M, N$ respectively. Let $R_B, R_C$ be natural numbers such that,

$$R_B \geq D^s_X(p_{XM} || p_X \times p_S) + O\left(\frac{\log 1}{\varepsilon}\right),$$

$$R_C \geq D^s_X(p_{SN} || p_X \times p_T) + O\left(\frac{\log 1}{\varepsilon}\right),$$

$$R_B + R_C \geq D^s_X(p_{XM} || p_X \times p_S \times p_T) + O\left(\frac{\log 1}{\varepsilon}\right).$$
There exists a shared randomness assisted protocol with communication \( R_B \) bits from Alice to Bob and \( R_C \) bits from Alice to Charlie, in which Bob outputs \( M \) and Charlie outputs \( N \) such that \( \frac{1}{2} \| p_{XMN} - p_{X_{\tilde{M}}N} \| \leq O(\sqrt{\varepsilon}) \).

Next we consider the same task but with side information at the receivers.

**Task 5: One-sender-two-receivers message compression with side information at receivers.** There are three parties Alice, Bob and Charlie. Alice holds random variables \((X, M, N)\), Bob holds random variable \( Y \) and Charlie holds random variable \( Z \) such that \((M, N) - X - (Y, Z)\). Alice sends a message to Bob and a message to Charlie. Bob and Charlie after receiving their respective messages, output random variables \( \tilde{M} \) and \( \tilde{N} \) respectively such that \( \frac{1}{2} \| p_{XYZMN} - p_{X\tilde{Y}Z\tilde{M}\tilde{N}} \| \leq O(\sqrt{\varepsilon}) \). Shared randomness is allowed between Alice and Bob and between Alice and Charlie.

We show the following achievability result for this task.

**Theorem 7** (Achievability for Task 5). Let \( S, T \) be random variables taking values over the same sets as \( M, N \) respectively. Let \( R_B, R_C \) be natural numbers such that,

\[
R_B \geq D_s^c(p_{XM} \| p_X \times p_S) - D_H^c(p_{MY} \| p_S \times p_Y) + O \left( \log \frac{1}{\varepsilon} \right),
\]

\[
R_C \geq D_s^c(p_{XN} \| p_X \times p_T) - D_H^c(p_{NZ} \| p_T \times p_Z) + O \left( \log \frac{1}{\varepsilon} \right),
\]

\[
R_B + R_C \geq D_s^c(p_{XMN} \| p_X \times p_S \times p_T) - D_H^c(p_{MY} \| p_S \times p_Y) - D_H^c(p_{NZ} \| p_T \times p_Z) + O \left( \log \frac{1}{\varepsilon} \right).
\]

There exists a shared randomness assisted protocol with communication \( R_B \) bits from Alice to Bob and \( R_C \) bits from Alice to Charlie, in which Bob outputs \( \tilde{M} \) and Charlie outputs \( \tilde{N} \) such that \( \frac{1}{2} \| p_{X\tilde{Y}Z\tilde{M}\tilde{N}} \| \leq O(\sqrt{\varepsilon}) \).

Finally we consider the following task with two senders and two receivers.

**Task 6: Two-senders-two-receivers message compression with side information at the receivers.** There are four parties Alice, Bob, Dave and Charlie. Alice holds random variables \((X_1, M_{11}, M_{12})\), Dave holds random variables \((X_2, M_{21}, M_{22})\), Bob holds random variable \( Y_1 \) and Charlie holds random variable \( Y_2 \) such that \((M_{11}, M_{12}) - X_1 - (Y_1, Y_2, X_2) \) and \((M_{21}, M_{22}) - X_2 - (Y_1, Y_2, X_1)\). Alice sends a message each to Bob and Charlie and Dave sends a message each to Bob and Charlie. At the end Bob outputs \( \tilde{M_{11}}, \tilde{M_{21}} \) and Charlie outputs \( \tilde{M_{12}}, \tilde{M_{22}} \) such that,

\[
\frac{1}{2} \| p_{X_1, M_{11}, M_{12}, X_2, M_{21}, M_{22}, Y_1, Y_2} - p_{X_1, \tilde{M}_{11}, \tilde{M}_{12}, X_2, \tilde{M}_{21}, \tilde{M}_{22}, Y_1, Y_2} \| \leq O(\sqrt{\varepsilon}).
\]

Shared randomness is allowed between pairs (Alice, Bob), (Alice, Charlie), (Dave, Bob) and (Dave, Charlie).

We obtain the following achievable result for this task using arguments similar to the arguments used in obtaining previous achievable results. We skip its proof for brevity.

**Theorem 8** (Achievability for Task 6). Let \( R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)} \) be natural numbers such that for \( i, j \in \{1, 2\} \),

\[
R_j^{(i)} \geq D_s^c(p_{X_i, M_{ij}} \| p_{X_i} \times p_{M_{ij}}) - D_H^c(p_{M_{ij}, Y_j} \| p_{M_{ij}} \times p_{Y_j}) + O \left( \log \frac{1}{\varepsilon} \right),
\]

for \( i, j, k \in \{1, 2\} \) such that \( i \neq k \) or \( j \neq l \),

\[
R_j^{(i)} + R_l^{(k)} \geq D_s^c(p_{X_i, M_{ij}}, p_{X_j} \times p_{M_{ij}}) + D_s^c(p_{X_k, M_{kl}}, p_{X_k} \times p_{M_{kl}})
\]

\[
- D_H^c(p_{M_{ij}, Y_j} \| p_{M_{ij}} \times p_{Y_j}) - D_H^c(p_{M_{kl}, Y_l} \| p_{M_{kl}} \times p_{Y_l}) + O \left( \log \frac{1}{\varepsilon} \right),
\]

for \( i, j, k \in \{1, 2\} \) such that \( i \neq k \) and \( j \neq l \),

\[
R_j^{(i)} + R_l^{(k)} \geq D_s^c(p_{X_i, M_{ij}, M_{kl}}, p_{X_i} \times p_{M_{ij}} \times p_{M_{kl}}) + D_s^c(p_{X_k, M_{kl}, M_{ij}}, p_{X_k} \times p_{M_{kl}})
\]

\[
- D_H^c(p_{M_{ij}, Y_j} \| p_{M_{ij}} \times p_{Y_j}) - D_H^c(p_{M_{kl}, Y_l} \| p_{M_{kl}} \times p_{Y_l}) + O \left( \log \frac{1}{\varepsilon} \right),
\]
and,
\[ R_1^{(1)} + R_2^{(1)} + R_1^{(2)} + R_2^{(2)} \geq D_{\Delta}^x(p_{X_1, M_{11}, M_{12}} || p_{X_1} \times p_{M_{11}} \times p_{M_{12}}) + D_{\Delta}^y(p_{X_2, M_{21}, M_{22}} || p_{X_2} \times p_{M_{21}} \times p_{M_{22}}) \]
\[ - D_{\Delta}^x(p_{M_{11}, M_{12}, Y_1} || p_{M_{11}} \times p_{M_{21}} \times p_{Y_1}) - D_{\Delta}^y(p_{M_{12}, M_{22}, Y_2} || p_{M_{12}} \times p_{M_{22}} \times p_{Y_2}) + O \left( \log \frac{1}{\varepsilon} \right). \]

There exists a shared randomness assisted protocol with communication \( R_1^{(1)} \) bits from Alice to Bob, \( R_2^{(1)} \) bits from Alice to Charlie, \( R_1^{(2)} \) bits from Dave to Bob and \( R_2^{(2)} \) bits from Dave to Charlie such that Bob outputs \( \hat{M}_{11}, \hat{M}_{21} \) and Charlie outputs \( \hat{M}_{12}, \hat{M}_{22} \) satisfying
\[ \frac{1}{2} \| p_{X_1, M_{11}, M_{12}, X_2, M_{21}, M_{22}, Y_1, Y_2} - p_{X_1, \hat{M}_{11}, \hat{M}_{12}, X_2, \hat{M}_{21}, \hat{M}_{22}, Y_1, Y_2} \| \leq O(\sqrt{\varepsilon}). \]

We state without giving further details, that the task above can be extended in a natural fashion to obtain an analogous task for multiple senders and multiple receivers and analogous communication bounds can be obtained using similar arguments.

### Applications of our results

Here we consider several tasks studied in previous works and show that our results imply the results shown in these works. Consider the following task.

**Task 7: Lossy source compression.** Let \( k \geq 0 \). There are two parties Alice and Bob. Alice holds a random variable \( X \) and Bob holds a random variable \( Y \). Alice sends a message to Bob and Bob outputs a random variable \( Z \) such that
\[ \Pr \{ d(X, Z) \geq k \} \leq O(\sqrt{\varepsilon}), \]
where \( d : \mathcal{X} \times \mathcal{Z} \rightarrow (0, \infty) \) is a distortion measure. There is no shared randomness allowed between Alice and Bob.

This problem was studied in the asymptotic i.i.d setting in [11] and in the non-i.i.d. setting in [31]. We show the following achievability result which follows as a corollary of Theorem [4]:

**Corollary 1** (Achievability for Task 7). Let \( \delta \geq 0 \). Let \( R \) be a natural number such that,
\[ R \geq \min_{M, f} \left( D_{\Delta}^x(p_{X, f(Y, M)} || p_X \times p_M) - D_{\Delta}^y(p_{Y, f(Y, M)} || p_Y \times p_M) + O \left( \log \frac{1}{\varepsilon} \right) \right), \]  
(1)
where \( M \) and \( f \) satisfy \( M = X - Y \) and \( \Pr \{ d(X, f(Y, M)) \geq k \} \leq \delta \). There exists a protocol with communication \( R \) bits from Alice to Bob, in which Bob outputs a random variable \( Z \) such that \( \Pr \{ d(X, Z) \geq k \} \leq \delta + O(\sqrt{\varepsilon}) \).

**Proof.** Let \( M \) and \( f \) be such that they achieve the minimum in Equation (1). Alice and Bob employ the protocol from Theorem [4] in which Alice send \( R \) bits to Bob and at the end Bob is able to generate \( \hat{M} \) such that \( \frac{1}{2} \| p_{X, M} - p_{X, \hat{M}} \| \leq O(\sqrt{\varepsilon}). \) Bob then outputs \( Z = f(Y, \hat{M}) \). Consider,
\[ \Pr \{ d(X, f(Y, \hat{M})) \geq k \} \leq \Pr \{ d(X, f(Y, M)) \geq k \} + \| p_{XY, M} - p_{XY, \hat{M}} \| \leq \delta + O(\sqrt{\varepsilon}). \]

This protocol uses shared randomness between Alice and Bob and \( \Pr \{ d(X, f(Y, \hat{M})) \geq k \} \leq \delta + O(\sqrt{\varepsilon}) \) averaged over the shared randomness. Hence there exists a fixed shared string between Alice and Bob, conditioned on which \( \Pr \{ d(X, f(Y, \hat{M})) \geq k \} \leq \delta + O(\sqrt{\varepsilon}) \) . Fixing this string finally gives us the desired protocol which does not use shared randomness.

Next we consider the following problem which was first studied by Slepian-Wolf [9] in the asymptotic setting. Its one-shot version was studied in [32, 33]. Its second order analysis was given in [34].

**Task 8: Two-senders-one-receiver source compression.** There are three parties Alice, Bob and Charlie. Alice possesses a random variable \( X \), Bob possesses a random variable \( Y \). Alice and Bob both send a message each to Charlie who at the end outputs random variables \( \hat{X}, \hat{Y} \) such that \( \Pr \{ (X, Y) \neq (\hat{X}, \hat{Y}) \} \leq O(\varepsilon + \sqrt{\delta}) \). There is no shared randomness allowed between any parties.

We show the following achievability result for this task which follows as a corollary of Theorem [4]:
**Proof.** We observe that are uniform. Moreover, over the shared randomness. Hence there exists a fixed shared string conditioned on which output of Charlie. We have,

\[ R_A \geq D_H^e(p_{XX} \| p_X \times p_S) - D_H^e(p_{XY} \| p_S \times p_Y) + O\left(\log \frac{1}{\delta}\right), \]

\[ R_B \geq D_H^e(p_{YY} \| p_Y \times p_T) - D_H^e(p_{XY} \| p_X \times p_T) + O\left(\log \frac{1}{\delta}\right), \]

\[ R_A + R_B \geq D_H^e(p_{XX} \| p_X \times p_S) + D_H^e(p_{YY} \| p_Y \times p_T) - D_H^e(p_{XY} \| p_S \times p_T) + O\left(\log \frac{1}{\delta}\right). \]

There exists a protocol with communication \( R_A \) bits from Alice to Charlie and \( R_B \) bits from Bob to Charlie, in which Charlie outputs random variable pair \((\hat{X}, \hat{Y})\) such that \( \Pr\left\{(X, Y) \neq (\hat{X}, \hat{Y})\right\} \leq 3\varepsilon + 9\sqrt{\delta} \).

**Proof.** Alice, Bob and Charlie use the protocol in Theorem, where we set \( M \leftarrow X \) and \( N \leftarrow Y \). Let \((\hat{X}, \hat{Y})\) be the output of Charlie. We have, \( \frac{1}{2}\|p_{XY} - p_{XY}\| \leq 3\varepsilon + 9\sqrt{\delta} \) which implies \( \Pr\left\{(X, Y) \neq (\hat{X}, \hat{Y})\right\} \leq 3\varepsilon + 9\sqrt{\delta} \). This protocol uses shared randomness between Alice and Bob and \( \Pr\left\{(X, Y) \neq (\hat{X}, \hat{Y})\right\} \leq 3\varepsilon + 9\sqrt{\delta} \) averaged over the shared randomness. Hence there exists a fixed shared string conditioned on which \( \Pr\left\{(X, Y) \neq (\hat{X}, \hat{Y})\right\} \leq 3\varepsilon + 9\sqrt{\delta} \). Fixing this string gives us the desired protocol which does not use shared randomness. □

**Comparison with previous work:** The bounds appearing in earlier works \([9, 32, 34, 33]\) employ either the conditional entropy \( H(X \mid Y) \) or its one-shot analogue

\[ H_0^e(X \mid Y) := \min\left\{ a : \Pr_{(x, y) \leftarrow p_{XY}} \left\{ p_X \mid Y = y(x) \geq 2^{-a} \right\} \geq 1 - \varepsilon \right\} \]

for achievability results in above task. To clarify the connection, we show the following lemma.

**Lemma 1.** Let \( X, T \) be distributed according to the uniform distribution. Then it holds that

\[ D_H^e(p_{XX} \| p_X \times p_S) - D_H^e(p_{XY} \mid p_S \times p_Y) \leq H_0^e(X \mid Y), \]

\[ D_H^e(p_{YY} \mid p_Y \times p_T) - D_H^e(p_{XY} \mid p_X \times p_T) \leq H_0^e(Y \mid X) \]

and

\[ D_H^e(p_{XX} \| p_X \times p_S) + D_H^e(p_{YY} \| p_Y \times p_T) - D_H^e(p_{XY} \| p_S \times p_T) \leq H_0^e(X, Y). \]

**Proof.** We observe that \( D_H^e(p_{XX} \| p_X \times p_S) = \log |\mathcal{X}| \) and \( D_H^e(p_{YY} \| p_Y \times p_T) = \log |\mathcal{Y}| \) for all \( \delta \in (0, 1) \), as \( p_S, p_T \) are uniform. Moreover,

\[ D_H^e(p_{XY} \mid p_S \times p_Y) = \max\left\{ a : \Pr_{(x, y) \leftarrow p_{XY}} \left\{ p_X \mid Y = y(x) \geq 2^{-a} \cdot p_S(x) = \frac{2^a}{|\mathcal{X}|} \right\} \geq 1 - \varepsilon \right\} \]

\[ = \max\left\{ a : \Pr_{(x, y) \leftarrow p_{XY}} \left\{ p_X \mid Y = y(x) \geq 2^a \cdot p_S(x) = 2^a \right\} \geq 1 - \varepsilon \right\} + \log |\mathcal{X}| \]

\[ = -\min\left\{ a : \Pr_{(x, y) \leftarrow p_{XY}} \left\{ p_X \mid Y = y(x) \geq 2^{-a} \right\} \geq 1 - \varepsilon \right\} + \log |\mathcal{X}| \]

\[ = -H_0^e(X \mid Y) + \log |\mathcal{X}|. \]

Thus,

\[ D_H^e(p_{XX} \| p_X \times p_S) - D_H^e(p_{XY} \mid p_S \times p_Y) = \log |\mathcal{X}| - D_H^e(p_{XY} \| p_S \times p_T) \leq H_0^e(X \mid Y), \]

proving the first inequality. Similar argument shows the other two inequalities. □
Thus, we find that our achievability result is at most the achievability result given in [32]. Due to one-shot near optimality of the result in [32], our bound is essentially equivalent to it, up to constants.

Next we consider the following task which was first studied by Wyner [10] in the asymptotic and i.i.d. setting, subsequently in the information-spectrum setting in [35, 40, 50], in the second order setting by [20, 36] and in the one-shot case in [32, 37].

**Task 9: Source compression with coded side information available at the decoder.** There are three parties Alice, Bob and Charlie. Alice possesses a random variable $X$, Bob possesses a random variable $Y$. Alice and Bob both send a message each to Charlie who at the end outputs a random variable $\hat{X}$ such that $\Pr\{X \neq \hat{X}\} \leq O(\varepsilon + \sqrt{\delta})$.

We show the following achievable result for this task which follows as a corollary from Theorem 4.

**Corollary 3** (Achievability for Task 9). Let $(X, X) \sim p_{XX}$, where $p_{XX}(x, x) = p_{X}(x)$. Let $S, T$ be random variables taking values over the sets $\mathcal{X}, \mathcal{Y}$ respectively. Let $R_A, R_B$ be natural numbers such that,

$$R_A \geq D_0^\delta(p_{XX} \mid p_X \times p_S) - D_H(p_{XX} \mid p_S \times p_N) + O\left(\frac{1}{\delta}\right),$$

$$R_B \geq D_0^\delta(p_{YN} \mid p_Y \times p_T) - D_H(p_{XX} \mid p_X \times p_T) + O\left(\frac{1}{\delta}\right),$$

$$R_A + R_B \geq D_0^\delta(p_{XX} \mid p_X \times p_S) + D_0^\delta(p_{YN} \mid p_Y \times p_T) - D_H(p_{YN} \mid p_S \times p_T) + O\left(\frac{1}{\delta}\right),$$

where $X - Y - N$. There exists a protocol with communication $R_A$ bits from Alice to Charlie and $R_B$ bits from Bob to Charlie, in which Charlie outputs random variable $\hat{X}$ such that $\Pr\{X \neq \hat{X}\} \leq 3\varepsilon + 9\sqrt{\delta}$.

**Proof.** Alice, Bob and Charlie use the protocol in Theorem 4 where we set $M \leftarrow X$ and $N \leftarrow N$. Let $(\hat{X}, \hat{N})$ be the output of Charlie. We have, $\frac{1}{2} \|p_{XYXN} - p_{XY\hat{X}\hat{N}}\| \leq 3\varepsilon + 9\sqrt{\delta}$ which implies $\Pr\{X \neq \hat{X}\} \leq 3\varepsilon + 9\sqrt{\delta}$. This protocol uses shared randomness between Alice and Bob and $\Pr\{X \neq \hat{X}\} \leq 3\varepsilon + 9\sqrt{\delta}$ averaged over the shared randomness. Hence there exists a fixed shared string conditioned on which $\Pr\{X \neq \hat{X}\} \leq 3\varepsilon + 9\sqrt{\delta}$. Fixing this string gives us the desired protocol which does not use shared randomness.

**Comparison with previous work:** It can be shown that the following is a subset of the achievable region given in Corollary 3 by setting $p_S$ to be the uniform distribution, $p_T = p_N$ and using arguments similar to those given in Lemma 1:

$$R_A \geq H_0^\delta(X \mid N) + O\left(\frac{1}{\delta}\right),$$

$$R_B \geq D_0^\delta(p_{YN} \mid p_Y \times p_N) - D_H(p_{XX} \mid p_X \times p_N) + O\left(\frac{1}{\delta}\right),$$

$$R_A + R_B \geq H_0^\delta(X \mid N) + D_0^\delta(p_{YN} \mid p_Y \times p_N) + O\left(\frac{1}{\delta}\right).$$

This rate region has two extreme points, one of which is

$$(R_A, R_B) = \left(H_0^\delta(X \mid N) + O\left(\frac{1}{\delta}\right), D_0^\delta(p_{YN} \mid p_Y \times p_N) + O\left(\frac{1}{\delta}\right)\right).$$

Thus, the following rate region is a subset of the rate region given in Corollary 3:

$$R_A \geq H_0^\delta(X \mid N) + O\left(\frac{1}{\delta}\right),$$

$$R_B \geq D_0^\delta(p_{YN} \mid p_Y \times p_N) + O\left(\frac{1}{\delta}\right).$$

This rate region is optimal [32, 37] when taken as a union over all $N$ satisfying $X - Y - N$. This yields the optimality of the rate region in Corollary 3 when taken as a union over all $N$ satisfying $X - Y - N$. 

9
Asymptotic and i.i.d. properties

As discussed earlier, our achievable communication for Task 1 is optimal in the asymptotic and i.i.d. setting. Using the asymptotic i.i.d. properties of the information spectrum relative entropy and hypothesis testing relative entropy from [24], we are able to establish the rate regions for all the remaining tasks. We discuss the rate regions for Task 3 (which subsumes Task 2), Task 5 (which subsumes Task 4), Task 7, Task 8 and Task 9.

- **Task 3**: The rate region is given as (where we use $R^*_A, R^*_B$ to represent the rates)

  \[
  R^*_A \geq I(X : M) - I(M : NZ), \\
  R^*_B \geq I(Y : N) - I(MZ : N), \\
  R^*_A + R^*_B \geq I(X : M) + I(Y : N) - I(Z : M : N),
  \]

  where $I(Z : M : N) = D(p_{ZMN} \| p_Z \times p_M \times p_N)$ is the tripartite mutual information.

- **Task 5**: The rate region is given as (where we use $R^*_A, R^*_B$ to represent the rates)

  \[
  R^*_A \geq I(X : M) - I(M : Y), \\
  R^*_B \geq I(X : N) - I(N : Z), \\
  R^*_A + R^*_B \geq I(X : M : N) - I(M : Y) - I(N : Z).
  \]

- **Task 7**: The achievable rate is

  \[
  R \geq \min_{M,f} (I(X : M | Y)),
  \]

  where $M$ and $f$ satisfy $M - X - Y$ and $\lim_{n \to \infty} \Pr \{d(X, f(Y, M)) \geq k\} = 0$.

- **Task 8**: The rate region is given as (where we use $R^*_A, R^*_B$ to represent the rates)

  \[
  R^*_A \geq H(X) - I(X : Y) = H(X \mid Y), \\
  R^*_B \geq H(Y) - I(X : Y) = H(Y \mid X), \\
  R^*_A + R^*_B \geq H(XY).
  \]

  This recovers the rate region obtained by Slepian and Wolf [9].

- **Task 9**: The rate region is given as (where we use $R^*_A, R^*_B$ to represent the rates)

  \[
  R^*_A \geq H(X) - I(X : N) = H(X \mid N), \\
  R^*_B \geq I(Y : N) - I(X : N), \\
  R^*_A + R^*_B \geq H(X) - I(X : N) + I(Y : N) = H(X \mid N) + I(Y : N).
  \]

  A subset of this rate region is the one obtained in [10].

  \[
  R^*_A \geq H(X \mid N), \\
  R^*_B \geq I(Y : N).
  \]

  Both rate regions match when taken as a union over all $N$ (which satisfy $X - Y - N$), due to the optimality of the latter. However, for a given $N$, our achievability result also implies the result of Slepian and Wolf [9] (by setting $N = Y$).

**Organization**

In the next section we present a few information theoretic preliminaries. In Section 3 we present proofs of our results. In Section 4 we consider the question of near optimality of Task 1. In Appendix A we present some deferred proofs.
2 Preliminaries

In this section we set our notations, make the definitions and state the facts that we will need later for our proofs.

For a natural number \( n \), let \([n]\) denote the set \{1, 2, \ldots, n\}. Let random variable \( X \) take values in a finite set \( \mathcal{X} \) (all sets we consider in this paper are finite). We let \( p_X \) represent the distribution of \( X \), that is for each \( x \in \mathcal{X} \) we let \( p_X(x) := \Pr(X = x) \). Let random variable \( Y \) take values in the set \( \mathcal{Y} \). We say \( X \) and \( Y \) are independent iff for each \( x \in \mathcal{X}, y \in \mathcal{Y} \) \( p_{XY}(x,y) = p_X(x) \cdot p_Y(y) \) and denote \( p_X \times p_Y := p_{XY} \). We say random variables \( (X, Y, Z) \) form a Markov chain, represented as \( X \rightarrow Y \rightarrow Z \), if for each \( x \in \mathcal{X}, Y |(X = x) \) and \( Z |(X = x) \) are independent. For an event \( E \), its complement is denoted by \( \neg E \). We define various information theoretic quantities below.

Definition 1. Let \( \varepsilon > 0 \). Let random variables \( X \) and \( X' \) take values in \( \mathcal{X} \). Define,

- \( \ell_1 \) distance: \( \|p_X - p_{X'}\| := \sum_x |p_X(x) - p_{X'}(x)| \).
- Relative entropy: \( D(p_X \| p_{X'}) := \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{p_{X'}(x)} \).
- Max divergence: \( D_\infty(p_X \| p_{X'}) := \max_x p_X(x) \log \frac{p_X(x)}{p_{X'}(x)} \).
- Smooth max divergence: \( D_{\varepsilon}^\infty(p_X \| p_{X'}) := \min_{\|p_{X''} - p_X\| \leq \varepsilon} D_{\infty}(p_X \| p_{X''}) \).
- Max information spectrum divergence: \( D_{\varepsilon}^\lambda(p_X \| p_{X'}) := \min \{ \alpha : \Pr_{x \leftarrow p_X} \left\{ \frac{p_X(x)}{p_{X'}(x)} > 2^\alpha \right\} \leq \varepsilon \} \).
- Smooth hypothesis testing divergence: \( D_{\varepsilon}^h(p_X \| p_{X'}) := \max \left\{ -\log(\Pr_{p_{X'}}(A)) \mid A \subseteq \mathcal{X}, \Pr_{p_X}(A) \geq 1 - \varepsilon \right\} \).

We will use the following facts.

Fact 1 (1). Let \( P \) and \( Q \) be two distributions over the set \( \mathcal{X} \), where \( P = \sum_i \lambda_i P_i \) is a convex combination of distributions \( \{P_i\}_1 \). It holds that,

\[
D(P \| Q) = \sum_i \lambda_i (D(P_i \| Q) - D(P_i \| P)) .
\]

Fact 2 (Monotonicity of relative entropy [38]). Let \((X, Y, Z)\) be jointly distributed random variables. It holds that,

\[
D(p_{XYZ} \| p_X \times p_Y \times p_Z) \geq D(p_{XY} \| p_X \times p_Y) .
\]

Fact 3 (Pinsker’s inequality [38]). Let \( P \) and \( Q \) be two distributions over the set \( \mathcal{X} \). It holds that,

\[
\|P - Q\| \leq 2 \cdot \sqrt{D(P \| Q)} .
\]

Fact 4 (Monotonicity under maps [38]). Let \( X \) be a random variable distributed over the set \( \mathcal{X} \). Let \( f : \mathcal{X} \rightarrow \mathcal{Z} \) be a function. Let random variable \( Z \), distributed over \( \mathcal{Z} \) be defined as,

\[
\Pr\{Z = z\} := \frac{\Pr\{X \in f^{-1}(z)\}}{\sum_{z' : \Pr\{X \in f^{-1}(z')\}}} .
\]

Similarly define random variable \( Z' \) from random variable \( X' \). It holds that,

\[
\|p_X - p_{X'}\| \geq \|p_Z - p_{Z'}\| .
\]

Following convex-split lemma from [1] is a main tool that we use. [1] provided a proof for a quantum version of this lemma and the proof of the classical version that we consider follows on similar lines. We defer the proof to Appendix.

Fact 5 (Convex-split lemma [1]). Let \( \varepsilon \in (0, \frac{1}{2}) \). Let \((X, M)\) (jointly distributed over \( \mathcal{X} \times \mathcal{M}\)) and \( W \) (distributed over \( \mathcal{M}\) be random variables. Let \( R \) be a natural number such that,

\[
R \geq D_{\varepsilon}^\infty(p_{XM} \| p_X \times p_W) + 4 \log \frac{1}{\varepsilon} .
\]

Let \( J \) be uniformly distributed in \([2^R]\) and joint random variables \((J, X, M_1, \ldots, M_{2^n})\) be distributed as follows:

\[
\Pr\{\{X, M_1, \ldots, M_{2^n}\} = (x, m_1, \ldots, m_{2^n}) \mid J = j\} = p_{XM}(x, m_j) \cdot p_W(m_1) \cdots p_W(m_{j-1}) \cdot p_W(m_{j+1}) \cdots p_W(m_{2^n}) .
\]

Then (below for each \( j \in [2^R], p_{W_j} = p_W \)),

\[
\|p_{XM_1 \ldots M_{2^n}} - p_X \times p_{W_1} \times \ldots \times p_{W_{2^n}}\| \leq 6\sqrt{\varepsilon} .
\]
We also need the following extension of this lemma whose quantum version was shown in [29]. The proof of the classical version that we consider follows on similar lines and is deferred to Appendix.

**Fact 6** (Bipartite convex-split lemma). Let \( \varepsilon \in (0, \frac{1}{\sqrt{2}}) \). Let \((X, M, N)\) (jointly distributed over \(X \times M \times N\)), \(U\) (distributed over \(M\)) and \(V\) (distributed over \(N\)) be random variables. Let \(R_1, R_2\) be natural numbers such that,

\[
R_1 \geq D_\varepsilon^Y(p_X || p_X \times p_U) + 8 \log \frac{1}{\varepsilon},
\]

\[
R_2 \geq D_\varepsilon^X(p_X || p_V) + 8 \log \frac{1}{\varepsilon},
\]

\[
R_1 + R_2 \geq D_\varepsilon(p_{XMN} || p_X \times p_U \times p_V) + 8 \log \frac{1}{\varepsilon}.
\]

Let \(J\) be uniformly distributed in \([2^{R_1}]\), \(K\) be independent of \(J\) and be uniformly distributed in \([2^{R_2}]\) and joint random variables \((J, K, X, M_1, \ldots, M_{2^n}, N_1, \ldots, N_{2^n})\) be distributed as follows:

\[
\Pr\{(X, M_1, \ldots, M_{2^n}, N_1, \ldots, N_{2^n}) = (x, m_1, \ldots, m_{2^n}, n_1, \ldots, n_{2^n}) \mid J = j, K = k\}
\]

\[
= p_{XMN}(x, m_j, n_k) \cdot p_U(m_1) \cdots p_U(m_{j-1}) \cdot p_U(m_{j+1}) \cdots p_U(m_{2^n}).
\]

\[
p_V(n_1) \cdots p_V(n_{k-1}) \cdot p_V(n_{k+1}) \cdots p_V(n_{2^n}).
\]

Then (below for each \(j \in [2^{R_1}], p_{U_j} = p_U\) and for each \(k \in [2^{R_2}], p_{V_k} = p_V\)),

\[
\|p_{XM_1 \cdots M_{2^n}, N_1 \cdots N_{2^n}} - p_X \times p_{U_1} \times \cdots \times p_{U_{2^n}} \times p_{V_1} \times \cdots \times p_{V_{2^n}}\| \leq 15 \varepsilon.
\]

The other main tool that we use is the position based decoding from [7] where a quantum version was shown. The proof of the classical version that we consider follows on similar lines and is deferred to Appendix.

**Fact 7** (Position based decoding [7]). Let \(\varepsilon, \delta \in (0, 1)\). Let \((Y, M)\) (jointly distributed over \(Y \times M\)) and \(W\) (distributed over \(M\)) be random variables. Let \(R\) be a natural number such that,

\[
R \leq \max \left\{ D_\varepsilon^Y(p_{YM} || p_Y \times p_W) - \log \frac{1}{\delta}, 0 \right\}.
\]

Let joint random variables \((J, Y, M_1, M_2, \ldots, M_{2^n})\) be distributed as follows. Let \(J\) be uniformly distributed in \([2^R]\) and

\[
\Pr\{(Y, M_1, M_2, \ldots, M_{2^n}) = (y, m_1, \ldots, m_{2^n}) \mid J = j\}
\]

\[
= p_{YM}(y, m_j) \cdot p_W(m_1) \cdots p_W(m_{j-1}) \cdot p_W(m_{j+1}) \cdots p_W(m_{2^n}).
\]

There is a procedure to produce a random variable \(J'\) from \((Y, M_1, M_2, \ldots, M_{2^n})\) such that \(\Pr\{J \neq J'\} \leq \varepsilon + \delta\).

We will also need the following extension of this decoding strategy shown in [39] where a (more general) quantum version was shown. The proof of the classical version that we consider follows on similar lines and is deferred to Appendix.

**Fact 8** (Bipartite position based decoding [39]). Let \(\varepsilon, \delta \in (0, 1)\). Let \((M, N)\) (jointly distributed over \(M \times N\)), \(U\) (distributed over \(M\)) and \(V\) (distributed over \(N\)). Let \(R_1, R_2\) be natural numbers such that,

\[
R_1 \leq \max \left\{ D_\varepsilon^M(p_{MN} || p_U \times p_N) - \log \frac{1}{\delta}, 0 \right\}
\]

\[
R_2 \leq \max \left\{ D_\varepsilon^M(p_{MN} || p_M \times p_V) - \log \frac{1}{\delta}, 0 \right\}
\]

\[
R_1 + R_2 \leq \max \left\{ D_\varepsilon^M(p_{MN} || p_U \times p_V) - \log \frac{1}{\delta}, 0 \right\}.
\]

Let joint random variables \((J, K, M_1, \ldots, M_{2^n}, N_1, \ldots, N_{2^n})\) be distributed as follows. Let \(J\) be uniformly distributed in \([2^R]\). Let \(K\) be independent of \(J\) and be uniformly distributed in \([2^{R_2}]\). Let,

\[
\Pr\{(M_1 \ldots M_{2^n}, N_1 \ldots N_{2^n}) = (m_1, \ldots, m_{2^n}, n_1, \ldots, n_{2^n}) \mid J = j, K = k\}
\]

\[
= p_{MN}(m_j, n_k) \cdot p_U(m_1) \cdots p_U(m_{j-1}) \cdot p_U(m_{j+1}) \cdots p_U(m_{2^n}).
\]

\[
p_V(n_1) \cdots p_V(n_{k-1}) \cdot p_V(n_{k+1}) \cdots p_V(n_{2^n}).
\]

There is a procedure to produce random variables \((J', K')\) from \((M_1, \ldots, M_{2^n}, N_1, \ldots, N_{2^n})\) such that \(\Pr\{(J, K) \neq (J', K')\} \leq 3\varepsilon + 3\delta\).
3 Proofs of our results

In this section we present proofs of our results mentioned in the Introduction.

Proof of Theorem: Let $E$ be such that $Y - X - (M, E)$. Let $R, r$ be natural numbers such that,

$$r \leq \max \left\{ D^r_H(p_{Y M E} | p_Y \times p_T) - \log \frac{1}{\delta}, 0 \right\},$$

$$R + r \geq D^r_S(p_{X M E} | p_X \times p_T) + 2 \log \frac{1}{\delta}.$$ 

Let us divide $[2^{R+r}]$ into $2^R$ subsets, each of size $2^r$. This division is known to both Alice and Bob. For $j \in [2^{R+r}]$, let $B(j)$ denote the subset corresponding to $j$. Let us invoke convex-split lemma (Fact 5) with $X \leftarrow X, M \leftarrow (M, E), W \leftarrow T$ and $R \leftarrow R + r$ to obtain joint random variables $(J, X, M_1, \ldots, M_{2^{R+r}})$. Let us first consider a fictitious protocol $P'$ as follows.

Fictitious protocol $P'$: Alice possesses random variables $(X, M, E)$, Bob possesses random variable $Y$ and they share $(M_1, \ldots, M_{2^{R+r}})$ as public randomness (from the joint random variables $(X, M_1, \ldots, M_{2^{R+r}})$ above).

Alice’s operations: Alice generates $J$ from $(X, M_1, \ldots, M_{2^{R+r}})$, using the conditional distribution of $J$ given $(X, M_1, \ldots, M_{2^{R+r}})$, and communicates $B(J)$ to Bob. This can be done using $R$ bits of communication. A similar encoding strategy is used in the works [21, 22] (see also the references therein).

Bob’s operations: Bob performs position based decoding as in Fact 7 using $Y$ and the subset $B(J)$, by letting $Y \leftarrow Y, M \leftarrow (M, E), W \leftarrow T$ and $R \leftarrow r$, and determines $J'$. Let $(M', E') := M_{J'}$. Bob outputs $M'$.

From Fact 7 we have $\Pr\{J \neq J'\} \leq \epsilon + \delta$ and hence, From Fact 7 we have $\Pr\{J \neq J'\} \leq \epsilon + \delta$ and hence,

$$\|p_{X M_1 \ldots M_{2^{R+r}}} - p_X \times p_T^1 \times \ldots \times p_T^{2^{R+r}}\| \leq 6\sqrt{\delta}.$$ 

Thus,

$$\|p_{X Y M} - p_{X Y M'}\| \leq \|p_{X M_1 \ldots M_{2^{R+r}}} - p_X \times p_T^1 \times \ldots \times p_T^{2^{R+r}}\| + \|p_{X Y M} - p_{X Y M'}\|$$

$$\leq 6\sqrt{\delta} + 2\epsilon + 2\delta \leq 2\epsilon + 8\sqrt{\delta}.$$ 

where (a) follows from the property $M - X - Y$ and (b) follows from Equation 2. This shows the desired.

Proof of Theorem: Let $R_A, r_A, R_B, r_b$ be natural numbers such that (existence of these numbers is guaranteed by the Fourier-Motzkin elimination technique [40] Appendix D) and the constraints in the statement of the Theorem,

$$R_A + r_A \geq D^r_S(p_{X M} | p_X \times p_S) + 2 \log \frac{1}{\delta},$$

$$R_B + r_B \geq D^r_S(p_{Y N} | p_Y \times p_T) + 2 \log \frac{1}{\delta},$$

$$r_A \leq \max\{D^r_H(p_{M N} | p_S \times p_X) - \log \frac{1}{\delta}, 0\},$$

$$r_B \leq \max\{D^r_H(p_{M N} | p_M \times p_T) - \log \frac{1}{\delta}, 0\},$$

$$r_A + r_B \leq \max\{D^r_H(p_{M N} | p_S \times p_T) - \log \frac{1}{\delta}, 0\}.$$
Let us divide $[2^R_{A+r_A}]$ into $2^R_A$ subsets, each of size $2^{r_A}$. This division is known to both Alice and Charlie. For $j \in [2^R_{A+r_A}]$, let $B(j)$ denote the subset corresponding to $j$. Similarly let us divide $[2^R_{B+r_B}]$ into $2^R_B$ subsets, each of size $2^{r_B}$. This division is known to both Bob and Charlie. For $k \in [2^R_{B+r_B}]$, let $B(k)$ denote the subset corresponding to $k$.

Let us invoke bipartite convex-split lemma (Fact 6) with $X \leftarrow (X, Y)$, $M \leftarrow M, N \leftarrow N, U \leftarrow S, V \leftarrow T, R_1 \leftarrow R_A + r_A$ and $R_2 \leftarrow R_B + r_B$ to obtain joint random variables $(J, K, X, Y, M_1, \ldots, M_{2^R_{A+r_A}}, N_1, \ldots, N_{2^R_{B+r_B}})$.

Let us first consider a fictitious protocol $\mathcal{P}'$ as follows.

**Fictitious protocol $\mathcal{P}'$:** Let Alice and Charlie share $(M_1, \ldots, M_{2^R_{A+r_A}})$ as public randomness. Let Bob and Charlie share $(N_1, \ldots, N_{2^R_{B+r_B}})$ as public randomness.

**Alice’s operations:** Alice generates $J$ from $(X, M_1, \ldots, M_{2^R_{A+r_A}})$, using the conditional distribution of $J$ given $(X, M_1, \ldots, M_{2^R_{A+r_A}})$, and communicates $B(J)$ to Charlie. This can be done using $R_A$ bits of communication.

**Bob’s operations:** Bob generates $K$ from $(Y, N_1, \ldots, N_{2^R_{B+r_B}})$, using the conditional distribution of $K$ given $(Y, N_1, \ldots, N_{2^R_{B+r_B}})$, and communicates $B(K)$ to Charlie. This can be done using $R_B$ bits of communication.

**Charlie’s operations:** Charlie performs bipartite position based decoding as in Fact 8 inside the subset $B(J) \times B(K)$, by letting $M \leftarrow M, N \leftarrow N, U \leftarrow S, V \leftarrow T, R_A \leftarrow r_A$ and $R_B \leftarrow r_B$, and determines $(J', K')$. Charlie outputs $(M', N') := (M_J, N_K)$.

Note that Alice and Bob’s operation produce the right joint distribution $(J, K, X, Y, M_1, \ldots, M_{2^R_{A+r_A}}, N_1, \ldots, N_{2^R_{B+r_B}})$ since $M - X - Y - N$. Therefore from Fact 8 we have,

$$\frac{1}{2} \| p_{XYMN} - p_{XYM'N'} \| \leq \Pr \{ (J, K) \neq (J', K') \} \leq 3\varepsilon + 3\delta. \tag{3}$$

Now consider the actual protocol $\mathcal{P}$.

**Actual protocol $\mathcal{P}$:** Alice and Charlie share $2^R_{A+r_A}$ i.i.d. copies of the random variable $S$, denoted $\{ S_1, S_2, \ldots, S_{2^R_{A+r_A}} \}$. Bob and Charlie share $2^R_{B+r_B}$ i.i.d. copies of the random variable $T$, denoted $\{ T_1, T_2, \ldots, T_{2^R_{B+r_B}} \}$. Alice, Bob and Charlie proceed as in $\mathcal{P}'$. Therefore the only difference in $\mathcal{P}$ and $\mathcal{P}'$ is shared randomness. Let $(M, N)$ represent Charlie’s outputs in $\mathcal{P}$.

From convex-split lemma

$$\frac{1}{2} \| p_{XYMN} - p_{XYM'N'} \| \leq \Pr \{ (J, K) \neq (J', K') \} \leq 3\varepsilon + 3\delta. \tag{3}$$

From Fact 4 triangle inequality for $\ell_1$ norm and Equation 3 we have,

$$\frac{1}{2} \| p_{XYMN} - p_{XYM'N} \| \leq 3\varepsilon + 9\delta \leq 3\varepsilon + 9\delta.$$

This shows the desired.

**Proof of Theorem 5:** The proof follows on similar lines as the proof of Theorem 4 and we provide a proof sketch here. Let $\{ R_A, R_B, r_A, r_B \}$ be natural numbers such that (existence of these numbers is guaranteed by the Fourier-Motzkin elimination technique [40] Appendix D] and the constraints in the statement of the Theorem),

$$R_A + r_A \geq D^c_H(p_{XM} \| p_X \times p_S) + O \left( \log \frac{1}{\varepsilon} \right),$$

$$R_B + r_B \geq D^c_H(p_{YN} \| p_Y \times p_T) + O \left( \log \frac{1}{\varepsilon} \right),$$

$$r_A \leq \max \left\{ D^c_H(p_{MNZ} \| p_S \times p_NZ) - O \left( \log \frac{1}{\varepsilon} \right), 0 \right\},$$

$$r_B \leq \max \left\{ D^c_H(p_{MZN} \| p_MZ \times p_T) - O \left( \log \frac{1}{\varepsilon} \right), 0 \right\},$$

$$r_A + r_B \leq \max \left\{ D^c_H(p_{MNZ} \| p_S \times p_T \times p_Z) - O \left( \log \frac{1}{\varepsilon} \right), 0 \right\}. $$
Let $A_1,A_2,A_3 \subseteq \mathcal{M} \times \mathcal{N} \times \mathcal{Z}$ be such that $\Pr_{p_{MNZ}}\{A_i\} \geq 1 - \varepsilon$ for all $i \in \{1, 2, 3\}$ and

$$D_\Pr(p_{MNZ} \| p_{S \times P_{NZ}}) = -\log \Pr_{p_{S \times P_{NZ}}} \{A_1\};$$
$$D_\Pr(p_{MZN} \| p_{MZ} \times P_T) = -\log \Pr_{p_{MZ} \times P_T} \{A_2\};$$
$$D_\Pr(p_{MZN} \| p_{S \times P_T \times P_Z}) \geq -\log \Pr_{p_{S \times P_T \times P_Z}} \{A_3\}.$$ 

Define $A := A_1 \cap A_2 \cap A_3$.

**Protocol $\mathcal{P}$**: Shared randomness and Alice and Bob's operations remain same as in the actual protocol $\mathcal{P}$ of the proof of Theorem \[4\].

**Charlie's operations**: Charlie on receiving $B(J)$ and $B(K)$ from Alice and Bob respectively, performs bipartite position based decoding (similar to Fact \[8\]) involving $Z$ and the random variables in the subset $B(J) \times B(K)$. He finds the first pair $(J', K')$ (in lexicographic order) such that $(Z, M_{J'}, N_{K'}) \in A$ and outputs $(M, N):=(M_{J'}, N_{K'})$.

Using arguments as in the proof of Fact \[8\] we get $\Pr\{(J, K) \neq (J', K')\} = O(\varepsilon)$. Now using Fact \[4\] and triangle inequality for $\ell_1$ norm it can be argued that $\|p_{XYM} - p_{XYMN}\| = O(\sqrt{\varepsilon})$.

**Proof of Theorem \[6\]**: Let us invoke bipartite convex-split lemma (Fact \[6\]) with $X \leftarrow X, M \leftarrow M, N \leftarrow N, U \leftarrow S, V \leftarrow T, R_1 \leftarrow R_B$ and $R_2 \leftarrow R_C$ to obtain joint random variables $(J, K, X, M_1, \ldots, M_{2^{RB}}, N_1, \ldots, N_{2^{RC}})$. We first consider a fictitious protocol $\mathcal{P}'$ as follows.

**Fictitious protocol $\mathcal{P}'$**: Let Alice and Bob share $(M_1, \ldots, M_{2^{RB}})$ as public randomness. Let Alice and Charlie share $(N_1, \ldots, N_{2^{RC}})$ as public randomness.

**Alice's operations**: Alice generates $(J, K)$ from $(X, M_1, \ldots, M_{2^{RB}}, N_1, \ldots, N_{2^{RC}})$, using the conditional distribution of $(J, K)$ given $(X, M_1, \ldots, M_{2^{RB}}, N_1, \ldots, N_{2^{RC}})$. She communicates $J$ to Bob (using $R_B$ bits) and $K$ to Charlie (using $R_C$ bits).

**Bob's operations**: Bob outputs $M' := M_J$.

**Charlie's operations**: Charlie outputs $N' := N_K$.

It holds that $p_{XM'N'} = p_{XMN}$. Now consider the actual protocol $\mathcal{P}$.

**Actual protocol $\mathcal{P}$**: Alice and Bob share $2^{RB}$ i.i.d. copies of the random variable $S$, denoted $\{S_1, S_2, \ldots, S_{2^{RB}}\}$. Alice and Charlie share $2^{RC}$ i.i.d. copies of the random variable $T$, denoted $\{T_1, T_2, \ldots, T_{2^{RC}}\}$. Alice, Bob and Charlie proceed as in $\mathcal{P}'$. Therefore the only difference in $\mathcal{P}$ and $\mathcal{P}'$ is shared randomness. Let $(M, N)$ represent Bob and Charlie’s outputs respectively in $\mathcal{P}$.

From bipartite convex-split lemma (Fact \[6\]),

$$\|p_{XM_1\ldots M_{2^{RB}}N_1\ldots N_{2^{RC}}} - p_X \times p_{M_1} \times \ldots \times p_{M_{2^{RB}}} \times p_{N_1} \times \ldots \times p_{N_{2^{RC}}}\| \leq 15\sqrt{\varepsilon}. \tag{4}$$

From Fact \[4\] triangle inequality for $\ell_1$ norm and Equation \[4\] we have,

$$\|p_{XMN} - p_{XMN}\| \leq 8\varepsilon + 15\sqrt{\varepsilon} \leq 23\sqrt{\varepsilon}.$$ 

**Proof of Theorem \[7\]**: The proof follows on similar lines as the proof of Theorem \[6\] and we provide a proof sketch here. Let $(R_B, R_C, r_B, r_C)$ be natural numbers such that,

$$R_B + r_B \geq D_\Pr(p_{XM} \| p_X \times p_S) + O\left(\log \frac{1}{\varepsilon}\right),$$
$$r_B \leq \max \left\{D_\Pr(p_{MY} \| p_S \times p_Y) - O\left(\log \frac{1}{\varepsilon}\right), 0\right\},$$
$$R_C + r_C \geq D_\Pr(p_{XN} \| p_X \times p_T) + O\left(\log \frac{1}{\varepsilon}\right),$$
\[ r_c \leq \max \left\{ D_H(p_{NZ} \| p_T \times p_Z) - O \left( \log \frac{1}{\epsilon} \right), 0 \right\}, \]
\[ R_B + R_C + r_b + r_c \geq D_s(p_{XMN} \| p_X \times p_S \times p_T) + O \left( \log \frac{1}{\epsilon} \right). \]

Let \( A_1 \subseteq \mathcal{Y} \times \mathcal{M} \) and \( A_2 \subseteq \mathcal{Z} \times \mathcal{N} \) be such that \( \Pr_{p_Y,M} (A_1) \geq 1 - \epsilon \) and \( \Pr_{p_Z,N} (A_2) \geq 1 - \epsilon \) and,
\[ D_H(p_{MY} \| p_S \times p_Y) = -\log \Pr_{p_S \times p_Y} (A_1), \]
\[ D_H(p_{NZ} \| p_T \times p_Z) = -\log \Pr_{p_T \times p_Z} (A_2). \]

Let us divide \([2^{R_B + r_b}]\) into \(2^{2R_B} \) subsets, each of size \(2^{r_B} \). This division is known to both Alice and Bob. For \( j \in [2^{R_B + r_B}] \), let \( B(j) \) denote the subset corresponding to \( j \). Similarly let us divide \([2^{R_C + r_C}]\) into \(2^{2R_C} \) subsets, each of size \(2^{r_C} \). This division is known to both Alice and Charlie. For \( k \in [2^{R_C + r_C}] \), let \( B(k) \) denote the subset corresponding to \( k \).

**Protocol \( \mathcal{P} \):** Alice and Bob share \(2^{R_B + r_B} \) i.i.d. copies of the random variable \( S \), denoted \( \{S_1, S_2, \ldots, S_{2^{R_B + r_B}}\} \). Alice and Charlie share \(2^{R_C + r_C} \) i.i.d. copies of the random variable \( T \), denoted \( \{T_1, T_2, \ldots, T_{2^{R_C + r_C}}\} \).

**Alice’s operations:** Alice generates \((J, K)\) as in protocol \( \mathcal{P} \) in the proof of Theorem 6. She communicates \( B(J) \) to Bob (using \( R_B \) bits) and \( B(K) \) to Charlie (using \( R_C \) bits).

**Bob’s operations:** Bob performs position based decoding as in Fact 7 by letting \( Y \leftarrow Y, M \leftarrow M, W \leftarrow S \) and \( R \leftarrow R_B \) and determines \( J' \). Bob outputs \( \hat{M} : = M_{J'} \).

**Charlie’s operations:** Charlie performs position based decoding as in Fact 7 by letting \( Y \leftarrow Y, M \leftarrow N, W \leftarrow T \) and \( R \leftarrow R_C \) and determines \( K' \). Charlie outputs \( \hat{N} : = N_{K'} \).

Using arguments as in the proof of Fact 7 we get \( \Pr\{ (J, K) \neq (J', K') \} = O(\sqrt{\epsilon}) \). Now using Fact 4 and triangle inequality for \( \ell_1 \) norm it can be argued that \( \| p_{XYZMN} - p_{XYZMN} \| = O(\sqrt{\epsilon}) \).

\[ \square \]

## 4 Optimality of the protocol for Task 1

The aim of this section is to relate our achievability result 2 with the result of Braverman and Rao 5. It may be noted that Braverman and Rao were considering expected communication cost, whereas we are considering the worst case communication cost for Task 1. Thus, we have restated the result below accordingly.

**Theorem 9** (Braverman and Rao protocol, 5). Let \( \epsilon, \delta \in (0, 1) \). Let \( R \) be a natural number such that,
\[ R \geq \inf_{P_{XY}N} D_s(\mathcal{P}_{XYMN} \| p_Y (p_X | Y \times p_N | Y)) + O \left( \log \frac{1}{\delta} \right), \]
where \((Y, N) \sim p_{Y,N} \). There exists a shared randomness assisted protocol in which Alice communicates \( R \) bits to Bob and Bob outputs random variable \( M \) satisfying \( \frac{1}{2}\| p_{XYM} - p_{XY\hat{M}} \| \leq \epsilon + 3\delta \).

**Proof.** Let \( N \) be the random variable that achieves the optimization above. Define
\[ c := D_s(\mathcal{P}_{XYMN} \| p_Y (p_X | Y \times p_N | Y)) \]
\[ = \min \left\{ a : \Pr_{(x,m,y) \leftarrow \mathcal{P}_{XYMN}} \left( \frac{p_{XMY}(x,m,y)}{p_Y(y)p_X|Y=y(x)} \cdot p_N|Y=y(m)} \right) > 2^a \right\} \leq \epsilon \]
\[ = \min \left\{ a : \Pr_{(x,m,y) \leftarrow \mathcal{P}_{XYMN}} \left( \frac{p_{M|X=x}(m)}{p_N|Y=y(m)} \right) > 2^a \right\} \leq \epsilon \] (5)
where the last equality follows because \( M = X - Y \). Further, define
\[ \epsilon_{x,y} := \Pr_{(m) \leftarrow p_{M|X=x}} \left( \frac{p_{M|X=x}(m)}{p_N|Y=y(m)} > 2^c \right). \]

16
It holds that \( \sum_{x,y} p_{XY}(x, y) \varepsilon_{x,y} \leq \varepsilon \).

Let \( K \) be the smallest integer such that \( K p_M | X=x(m), K p_N | Y=y \) are integers. This can be assumed to hold with arbitrarily small error. Further, define the set \( \mathcal{K} := \{1, \cdots, K\} \). Let \( U \) be a random variable taking values uniformly in \( \mathcal{M} \times K \). Define the following function:

\[
f_{NE|Y=y}(m, e) := \begin{cases} 
1 & \text{if } e < K \cdot 2^c \cdot p_{N|Y=y}(m), \\
0 & \text{otherwise},
\end{cases}
\]

the following probability distribution:

\[
p_{ME|X=x}(m, e) = \begin{cases} 
\frac{1}{K} & \text{if } e < K \cdot p_{M|X=x}(m), \\
0 & \text{otherwise},
\end{cases}
\]

and the following sub-normalized probability distribution:

\[
p_{M'E|X=x,Y=y}(m, e) = \begin{cases} 
\frac{1}{K} & \text{if } e < K \cdot \min(p_{M|X=x}(m), 2^c \cdot p_{N|Y=y}(m)), \\
0 & \text{otherwise}.
\end{cases}
\]

**The protocol:** Alice and Bob share \( \frac{|\mathcal{M}|}{2} \) i.i.d. copies of \( U \), denoted \( \{U_1, U_2, \ldots, U_{|\mathcal{M}|}\} \). They also share \( \frac{2^c}{\varepsilon} \) random hash functions \( \mathcal{H} : \{1, 2, \ldots, \frac{|\mathcal{M}|}{2}\} \rightarrow \{0, 1\} \). Upon observing \( x \leftarrow X \), Alice takes samples \( (m_j, e_i) \leftarrow U_i \) and accepts the first index \( l \) that satisfies \( p_{ME|X=x}(m_l, e_i) > 0 \). Upon observing \( y \leftarrow Y \), Bob takes samples \( (m_j, e_i) \leftarrow U_j \) and locates all the indices \( j \) that satisfy \( f_{NE|Y=y}(m_j, e_j) = 1 \). Let the set of indices located by Bob be \( J \). Alice aborts the protocol if she does not find any index \( j \). Bob aborts the protocol if \( |J| \geq \frac{2^c}{\varepsilon} \). Conditioned on not aborting, Alice samples all the hash functions, and sends to Bob the evaluation \( \{\mathcal{H}(i), \mathcal{H}_2(i), \ldots, \mathcal{H}_p(i)\} \). Bob evaluates all the hash functions for each index \( j \in J \), he aborts if there exist \( j, j' \in J \) such that \( \mathcal{H}_l(j) = \mathcal{H}_l(j') \) for all \( l \) or there exists no \( j \) such that \( \mathcal{H}_l(j) = \mathcal{H}_l(i) \) for all \( l \). Conditioned on not aborting, he has located the correct index \( i \), and considers \( (m_i, e_i) \). Conditioned on any abort, Bob considers a sample \( (m, e) \) the random variable \( U \). He outputs \( m \) from the sample he considers. Let the overall output of Bob be distributed according to \( \mathcal{M} \).

**Error analysis:** Let \( E_1 \) be the event that \( i \notin J \). Let \( E_2 \) be the event that Alice does not find any index \( i \) or \( |J| \geq \frac{2^c}{\varepsilon} \). Let \( E_3 \) be the event that there exist \( j, j' \in J \) such that \( H_l(j) = H_l(j') \) for all \( l \). From Equation \[ \Pr\{E_1 | X = x, Y = y\} \leq \varepsilon_{x,y} \]. Moreover, as argued in \[ \Pr\{E_2\} \leq 2\delta \] and \( \Pr\{-E_3\} \geq 1 - \delta \). Thus, \( \Pr\{-E_2 \cap -E_3\} \geq 1 - 3\delta \). Conditioned on the events \( -E_2 \cap -E_3 \), Bob has obtained a sample \( (m, e) \) distributed according to \( p_{M'E|X=x,Y=y} + \varepsilon_{x,y} \cdot U \), as he outputs a sample according to the sub-normalized distribution \( p_{M'E|X=x,Y=y} \) conditioned on event \( \neq E_1 \) and uniform otherwise. Now, we have:

\[
\frac{1}{2} \|p_{ME|X=x} - p_{M'E|X=x,Y=y} - \varepsilon_{x,y} \cdot U\| \leq \frac{1}{2} \|p_{ME|X=x} - p_{M'E|X=x,Y=y}\| + \frac{\varepsilon_{x,y}}{2} \\
\leq \sum_{(m,e):K \cdot p_{M|X=x}(m) \geq e \geq K \cdot \min(p_{M|X=x}(m), 2^c \cdot p_{N|Y=y}(m))} \frac{1}{2K} + \frac{\varepsilon_{x,y}}{2} \\
\leq \sum_{m:p_{M|X=x} > 2^c \cdot p_{N|Y=y}(m)} p_{M|X=x}(m) + \frac{\varepsilon_{x,y}}{2} \\
\leq \varepsilon_{x,y},
\]

where the last inequality follows from Equation \[ \Pr\{E_1 | X = x, Y = y\} \leq \varepsilon_{x,y} \]. Thus, we conclude that

\[
\frac{1}{2} \|p_{M|X=x} - p_{\hat{M}|X=x,Y=y}\| \leq \frac{1}{2} \|p_{ME|X=x} - p_{M'E|X=x,Y=y}\| \\
\leq \frac{\Pr\{-E_2 \cap -E_3\}}{2} \|p_{ME|X=x} - p_{M'E|X=x,Y=y} - \varepsilon_{x,y} \cdot U\| + \Pr\{E_2 \cup E_3\} \\
\leq (1 - 3\delta)\varepsilon_{x,y} + 3\delta \leq \varepsilon_{x,y} + 3\delta.
\]

This implies that

\[
\frac{1}{2} \|p_{XY} - p_{X\hat{M}Y}\| \leq \sum_{x,y} p_{XY}(x, y) \varepsilon_{x,y} + 3\delta \leq \varepsilon + 3\delta.
\]
This completes the proof.

We now compare our result (Theorem 10) with Theorem 9. To accomplish this, we first define a series of new quantities and relate them to each other. In what follows, we will use $P$ to represent a protocol for the Task discussed in Section 1.

- **Opt**: Let $P$ be any shared randomness assisted communication protocol in which Alice and Bob work on their respective inputs $(X, Y)$, and Bob outputs a random variable $M$ correlated with $XY$. Let $P(X, Y) := (X, Y, M)$ represent the output of the protocol. We define $\text{err}(P) := \frac{1}{2} ||p_{XYM} - p_{XY}M||$ as the error incurred by the protocol and $C(P)$ as the communication cost of the protocol. Define

$$\text{Opt}^\varepsilon := \min_{P:\text{err}(P) \leq \varepsilon} C(P).$$

- **Opt$_1$**: Let $S$ be the shared randomness in a protocol $P$. Note that $S$ is independent of $(X, Y)$. Let $V$ be a random variable such that $Y - (X, S) - V$, $X - (Y, V, S) - M$ and $\frac{1}{2} ||p_{XYM} - p_{XY}M|| \leq \varepsilon$, where $M$ is output by Bob (as discussed above). The random variable $V$ represents the message generated by Alice to Bob in $P$. Define

$$\text{Opt}_1^\varepsilon := \min_{(X, Y, S, M, V)} D_{\infty}(p_{XSV} \| p_X \times p_S \times p_U),$$

where $U$ is the uniformly distributed random variable taking values over same set as $V$.

- **BR**: The amount of communication needed by the protocol of Braverman and Rao for Task 1 is denoted by $BR^\varepsilon$ and formally defined below (see also Theorem 9). Let $(Y, N) \sim p_{YN}$. Define

$$\text{BR}^\varepsilon := \inf_{p_{YN}} D_s((p_{XY} \| p_Y \times p_{N|Y})).$$

- **Ext$_\delta$**: This is similar to the quantity obtained in the result of Theorem 2 by setting $T$ as uniform random variable $U$. Define

$$\text{Ext}_\delta^\varepsilon := D_s(p_{XME} \| p_X \times p_U) - D_H (p_{YM} \| p_Y \times p_U).$$

The following theorem relates all the quantities defined above to each other. This in turn allows us to prove the optimality of our protocol (see Theorem 3) along with the protocol of Braverman and Rao (Theorem 9).

**Theorem 10.** Let $M = X - Y$ and $\varepsilon, \delta \in (0, 1)$. Then it holds that

1. $\text{Opt}^\varepsilon \geq \text{Opt}_1^\varepsilon$.
2. $\text{Opt}_1^\varepsilon \geq \text{BR}^\varepsilon/(1 - \delta) - \log(\frac{1}{\delta})$.
3. $\text{BR}^\varepsilon + O(\log(\frac{1}{\delta})) \geq \text{Opt}^\varepsilon + 3\delta$.
4. $\text{Ext}_\delta^\varepsilon + O(\log(\frac{1}{\delta})) \geq \text{Opt}^\varepsilon + 4\delta$.
5. $\text{BR}^\varepsilon > \text{Ext}^0_\varepsilon$.
6. $\text{Ext}^0_\varepsilon \geq D_s(p_{XMEY} \| p_Y \times p_{M|Y}) \geq D_s(p_{XYE} \| p_Y \times p_{MY|Y})$.

**Proof.** We will prove the inequalities in the order they appear in the Theorem.

1. In any one-way communication protocol $P$ with a shared randomness $S$, Alice produces a message $V \in V$ using $(X, S)$, and communicates this to Bob. Notice that for this choice of $V$ we have $Y - (X, S) - V$. Using the message $V$, shared randomness $S$ and his input $Y$, Bob outputs $M$ such that $\frac{1}{2} ||p_{XYM} - p_{XY}M|| \leq \varepsilon$ and $X - (Y, V, S) - M$. The total number of bits communicated by Alice to Bob is $C(P) = \log |V|$. The inequality now follows from the relation $D_{\infty}(p_{XSV} \| p_X \times p_S \times p_U) \leq \log |V|$ (as $p_{XS} = p_X \times p_S$) and the definition of $\text{Opt}_1^\varepsilon$. 
2. For the random variables $(X, Y, S, U)$ as defined in $\text{Opt}_1^\varepsilon$, we prove the following:

\[
\begin{align*}
D_\infty(p_{XSV||X,S,U}) & \triangleq D_\infty(p_{YXSV||YX,S,U}) \\
& = D_\infty(p_{YXSV||YX,S,U}) \\
& = \min_{S',V'} D_\infty(p_{YXSV||YX,S,U}) \\
& \geq \min_{S'} D_\infty(p_{YXSV||YX,S,U}) \\
& \geq \min_{P_{N|Y}} D_\infty(p_{YXSV||YX,S,U}) \\
& = \min_{P_{N|Y}} D_\infty(p_{YXSV||YX,S,U}) \\
& \varepsilon \geq \frac{1}{2} \|p_{XMY} - p_{XMY}|| \\
& \geq | \Pr_{p_{XMY}}(A) - \Pr_{p_{XMY}}(A)| \\
& \geq \Pr_{p_{XMY}}(A) - \Pr_{p_{XMY}}(A) \\
& \geq \Pr_{p_{XMY}}(A) - \delta \Pr_{p_{XMY}}(A) \\
& = (1 - \delta) \Pr_{p_{XMY}}(A)
\end{align*}
\]

Thus, we conclude that $\Pr_{p_{XMY}}(\neg A) \geq 1 - \frac{\varepsilon}{1 - \delta}$. This implies

\[
D_\infty(p_{XMY||YXSV}) \leq D_\infty(p_{XMY||YXSV}) + \log \frac{1}{\delta}.
\]

Combining with Equation (6), the item concludes.

3. This is a direct consequence of Theorem 9.

4. This is a direct consequence of Theorem 2.

5. Let $N$ be as obtained from the definition of $\text{BR}_\varepsilon^\varepsilon$. From Theorem 11 below, it holds that there exists a random variable $E$ such that $(X, Y, M, E)$ satisfies $Y - X - (M, E)$ and

\[
D_\varepsilon(p_{XMY||YXSV}) \geq D_\varepsilon(p_{XMY||YXSV}) - D_\varepsilon(p_{XME||YXSV}).
\]

This concludes the item.

6. Let $\alpha := D_\varepsilon(p_{XME||YXSV})$ and $\beta := D_\varepsilon(p_{XME||YXSV})$. Let $A \subseteq Y \times M \times \mathcal{E}$ be the set achieving the optimum for $\beta$. For every $(x, y, m, e) \in A$, we have

\[
p_{XMEY}(x, m, e, y) = p_{XMEY}(x, m, e, y) \leq 2^{-\alpha}p_{X|Y}(x)p_{Y}(y)p_{MU}(m, e) \leq 2^{\alpha - \beta}p_{X|Y}(x)p_{Y}(y)p_{MUE}(m, e)
\]

Moreover, $\Pr_{p_{XMEY}}(A) \geq 1 - \varepsilon$. Thus,

\[
D_\varepsilon(p_{XMEY||YXSV}) \leq \alpha - \beta.
\]

This completes the proof.

The following theorem shows that the information theoretic quantity obtained in Theorem 2 is upper bounded by the information theoretic quantity obtained in Theorem 9.

**Theorem 11.** Let $N$ be the optimal random variable appearing in the definition of $\text{BR}_\varepsilon^\varepsilon$. Then there exists a random variable $E$ such that $Y - X - (M, E)$ and

\[
D_\varepsilon(p_{XMY||YXSV}) \geq D_\varepsilon(p_{XME||YXSV}) - D_\varepsilon(p_{XME||YXSV}),
\]

where $U$ is uniformly distributed over the set over which the random variable pair $(M, E)$ take values.
Proof. The proof is divided in the following steps.

Construction of appropriate extension: Let $K$ be the smallest integer such that $K_p M_{X=x}(m)$ is an integer. This can be assumed to hold with arbitrarily small error. Further, let $E$ be a random variable taking values over the set $\mathcal{K} := \{1, \cdots, K\}$ and jointly distributed with $(X, M)$ as follows: for every $(m, e, x) \in \mathcal{M} \times \mathcal{K} \times \mathcal{X}$,

$$p_{XME}(x, m, e) := \begin{cases} \frac{p_x(x)}{K} & \text{if } e \leq K_p M_{X=x}(m), \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

It can be seen that the property $Y - X - (M, E)$ holds. Let $U$ be a uniform random variable distributed over the set $\mathcal{K} \times \mathcal{M}$. Now we can establish the following:

$$D_0^s(p_{XME} \parallel p_X \times p_U) \overset{a}{=} \max_{m,x,e} \log \frac{p_{XME}(x, m, e)}{p_X(x)p_U(u)}$$

$$= \log \frac{\mathcal{M} | K}{K} = \log |\mathcal{M}|, \quad (8)$$

where (a) follows from the definition of $D_0^s(p_{XME} \parallel p_X \times p_U)$; (b) follows from Equation (7) and the fact that $U$ is uniform over the set $\mathcal{K} \times \mathcal{M}$.

Lower bounding hypothesis testing relative entropy: For brevity, let

$$D^*_\infty := D_f^s(p_{YM} \parallel p_Y (p_X | Y \times p_N | Y)).$$

Define the following set

$$\mathcal{A} := \{(y, m, e) \in \mathcal{Y} \times \mathcal{M} \times \mathcal{K} : e \leq K 2^D^*_{\infty} p_N | Y = y(m)\} \quad (9)$$

We will prove the following

$$\Pr_{p_Y \times p_U} \{A\} = 2^{-(\log |\mathcal{M}| - D^*_\infty)}; \quad (10)$$

$$\Pr_{p_{YM} E} \{A\} \geq 1 - \varepsilon. \quad (11)$$

The theorem now follows from the definition of $D_f^H(p_{YM} E \parallel p_Y \times p_U)$ and Equations (8), (10), (11) as follows:

$$D_f^H(p_{YM} E \parallel p_Y \times p_U) \geq \log |\mathcal{M}| - D^*_\infty$$

$$= D_0^s(p_{XME} \parallel p_X \times p_U) - D^*_\infty$$

which leads to

$$D^*_\infty \geq D_0^s(p_{XME} \parallel p_X \times p_U) - D_f^H(p_{YM} E \parallel p_Y \times p_U).$$

Proof of Equation (10): Towards this notice the following

$$\Pr_{p_Y \times p_U} \{A\} = \sum_{(y, m, e) \in \mathcal{A}} p_Y(y)p_U(m, e)$$

$$= \sum_{y \in \mathcal{Y}} p_Y(y) \sum_{(m, e) : (y, m, e) \in \mathcal{A}} \frac{1}{|\mathcal{M}|K}$$

$$= \sum_{(y, m) \in \mathcal{Y} \times \mathcal{M}} p_Y(y)p_N | Y = y(m) \frac{K 2^D^*_\infty}{|\mathcal{M}|K}$$

$$= \frac{2^D^*_\infty}{|\mathcal{M}|}$$

$$= 2^{-(\log |\mathcal{M}| - D^*_\infty).}$$
Proof of Equation (11). Towards this we have the following:

\[ \Pr_{\text{PY ME}} \{ A \} = \sum_x p_X(x) \sum_{(y,m,e) \in A} p_{Y|X=x}(y) p_{ME|X=x}(m, e) \]

\[ \leq \sum_x p_X(x) \sum_y p_{Y|X=x}(y) \sum_m \frac{1}{K} \sum_{e \leq K_{p_{M|X=x}(m)}} \sum_{(y,m,e) \in A} p_{M|X=x}(m) \]

\[ \geq \sum_{(x,y)} p_{X|Y}(x, y) \sum_{m: p_{M|X=x}(m) \leq 2^{D^*_\infty} p_N|Y=y(m)} \sum_{e \leq K_{p_{M|X=x}(m)}} \]

\[ \leq 1 - \varepsilon, \]

where a follows from Definition (7), b follows because for every \( x \)

\[ \{ (y, m, e) : p_{M|X=x}(m) \leq 2^{D^*_\infty} p_N|Y=y(m) \text{ and } e \leq K_{p_{M|X=x}(m)} \} \subseteq A, \]

and c follows from the definition of \( D^*_\infty \). This completes the proof.

Acknowledgment

We thank Marco Tomamichel and Mario Berta for helpful discussions. This work is supported by the Singapore Ministry of Education and the National Research Foundation, also through the Tier 3 Grant Random numbers from quantum processes MOE2012-T3-1-009 and NRF RF Award NRF-NRFF2013-13.

References

[1] A. Anshu, V. K. Devabathini, and R. Jain, “Quantum communication using coherent rejection sampling.” To appear in Physical Review Letters, 2017.
[2] R. Jain, J. Radhakrishnan, and P. Sen, “A direct sum theorem in communication complexity via message compression,” in Proceedings of the 30th international conference on Automata, languages and programming, ICALP’03, (Berlin, Heidelberg), pp. 300–315, Springer-Verlag, 2003.
[3] R. Jain, J. Radhakrishnan, and P. Sen, “Prior entanglement, message compression and privacy in quantum communication,” in Proceedings of the 20th Annual IEEE Conference on Computational Complexity, (Washington, DC, USA), pp. 285–296, IEEE Computer Society, 2005.
[4] P. Harsha, R. Jain, D. Mc.Allester, and J. Radhakrishnan, “The communication complexity of correlation,” IEEE Transactions on Information Theory, vol. 56, pp. 438–449, 2010.
[5] M. Braverman and A. Rao, “Information equals amortized communication,” in Proceedings of the 52nd Symposium on Foundations of Computer Science, FOCS ‘11, (Washington, DC, USA), pp. 748–757, IEEE Computer Society, 2011.
[6] J. Radhakrishnan, P. Sen, and N. Warsi, “One-shot marton inner bound for classical-quantum broadcast channel,” IEEE Transactions on Information Theory, vol. 62, pp. 2836–2848, May 2016.
[7] A. Anshu, R. Jain, and N. A. Warsi, “One shot entanglement assisted classical and quantum communication over noisy quantum channels: A hypothesis testing and convex split approach.” https://arxiv.org/abs/1702.01940, 2017.
[8] C. E. Shannon, “A mathematical theory of communication,” The Bell System Technical Journal, vol. 27, pp. 379–423, July 1948.
[9] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” IEEE Transactions on Information Theory, vol. 19, pp. 471–480, Jul 1973.
[10] A. Wyner, “On source coding with side information at the decoder,” IEEE Transactions on Information Theory, vol. 21, pp. 294–300, May 1975.
[11] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, Jan 1976.

[12] A. C.-C. Yao, “Some complexity questions related to distributive computing (preliminary report),” in *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC ’79, (New York, NY, USA), pp. 209–213, ACM, 1979.

[13] E. Kushilevitz and N. Nisan, *Communication Complexity*. New York, NY, USA: Cambridge University Press, 1997.

[14] A. K. Chandra, M. L. Furst, and R. J. Lipton, “Multi-party protocols,” in *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, STOC ’83, (New York, NY, USA), pp. 94–99, ACM, 1983.

[15] A. A. Sherstov, “The multiparty communication complexity of set disjointness,” in *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC ’12, (New York, NY, USA), pp. 525–548, ACM, 2012.

[16] T. Lee and A. Shraibman, “Disjointness is hard in the multiparty number-on-the-forehead model,” *computational complexity*, vol. 18, pp. 309–336, Jun 2009.

[17] P. Beame and T. Huynh, “Multipart communication complexity and threshold circuit size of AC$^0$,” *SIAM Journal on Computing*, vol. 41, no. 3, pp. 484–518, 2012.

[18] R. Jain, Y. Shi, Z. Wei, and S. Zhang, “Efficient protocols for generating bipartite classical distributions and quantum states,” *IEEE Transactions on Information Theory*, vol. 59, pp. 5171–5178, Aug 2013.

[19] P. Cuff and E. C. Song, “The likelihood encoder for source coding,” in *2013 IEEE Information Theory Workshop (ITW)*, pp. 1–2, Sept 2013.

[20] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, “Nonasymptotic and second-order achievability bounds for coding with side-information,” *IEEE Transactions on Information Theory*, vol. 61, pp. 1574–1605, April 2015.

[21] E. C. Song, P. Cuff, and H. V. Poor, “The likelihood encoder with applications to lossy compression and secrecy,” in *2015 Information Theory and Applications Workshop (ITA)*, pp. 301–307, Feb 2015.

[22] Z. Goldfeld, P. Cuff, and H. H. Permuter, “Wiretap channels with random states non-causally available at the encoder,” in *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, pp. 1–5, Nov 2016.

[23] P. Cuff, “Communication requirements for generating correlated random variables,” in *2008 IEEE International Symposium on Information Theory*, pp. 1393–1397, July 2008.

[24] M. Tomamichel and M. Hayashi, “A hierarchy of information quantities for finite block length analysis of quantum tasks,” *IEEE Transactions on Information Theory*, vol. 59, pp. 7693–7710, Nov 2013.

[25] I. Devetak and J. Yard, “Exact cost of redistributing multipartite quantum states,” *Phys. Rev. Lett.*, vol. 100, 2008.

[26] J. T. Yard and I. Devetak, “Optimal quantum source coding with quantum side information at the encoder and decoder,” *IEEE Transactions on Information Theory*, vol. 55, pp. 5339–5351, 2009.

[27] D. Touchette, “Direct sum theorem for bounded round quantum communication complexity,” 2014. http://arxiv.org/abs/1409.4391.

[28] M. Berta, M. Christandl, and D. Touchette, “Smooth entropy bounds on one-shot quantum state redistribution,” *IEEE Transactions on Information Theory*, vol. 62, pp. 1425–1439, March 2016.

[29] N. Datta, M.-H. Hsieh, and J. Oppenheim, “Optimality of one-shot state redistribution.” In preparation, 2014.

[30] A. Anshu, R. Jain, and N. A. Warsi, “Smooth min-max relative entropy based bounds for one-shot classical and quantum state redistribution.” https://arxiv.org/abs/1702.02396, 2017.

[31] K. Iwata and J. Muramatsu, “An information-spectrum approach to rate-distortion function with side information,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E85-A, no. 6, pp. 1387–1395, 2002.

[32] N. A. Warsi, “Simple one-shot bounds for various source coding problems using smooth Rényi quantities,” *Problems of Information Transmission*, vol. 52, no. 1, pp. 39–65, 2016.

[33] T. Utematsu and T. Matstuta, “Revisiting the Slepian-Wolf coding problem for general sources: A direct approach,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Honolulu, HI), June 2014.
A Deferred proofs

Proof of Fact 5: Let \( c := D'_r(p_{XM} \parallel p_{X} \times p_W) \). Define,

\[
\text{Good} := \left\{(x, m) : \frac{p_{XM}(x, m)}{p_X(x)p_W(m)} \leq 2^c \right\}.
\]

This implies (from definition of \( c \)) that \( p := \Pr\{(X, M) \in \text{Good}\} \geq 1 - \varepsilon \). Let us define joint random variables \((X', M')\) as follows:

\[
p_{X'M'}(x, m) = \begin{cases} 
\frac{p_{XM}(x, m)}{p} & \text{if } (x, m) \in \text{Good}, \\
0 & \text{otherwise}.
\end{cases}
\]

We note that,

\[
\forall (x, m) : \frac{p_{X'M'}(x, m)}{p_{X}(x)p_{W}(m)} \leq \frac{2^c}{p} \quad \text{and} \quad p_{X'}(x) \leq \frac{p_{X}(x)}{p}.
\]

Let us construct joint random variables \((J', X', M'_1, \ldots, M'_{2R})\) from \((X', M')\) in a similar fashion as we constructed joint random variables \((J, X, M_1, \ldots, M_{2R})\) from \((X, M)\). We note that,

\[
\|p_{X'M'_1 \ldots M'_{2R}} - p_{XM_1 \ldots M_{2R}}\| = \|p_{XM} - p_{X'M'}\| \leq 4\varepsilon. \tag{13}
\]

Consider,

\[
D \left(p_{X'M'_1 \ldots M'_{2R}} \parallel p_{X} \times p_{W_1} \times \cdots \times p_{W_{2R}}\right)
\]

\[
\leq \frac{a}{2R} \sum_{j=1}^{2R} \left( D \left(p_{X'M'_j} \parallel p_{X} \times p_{W_j}\right) - D \left(p_{X'M'_j} \times p_{W_1} \times \cdots \times p_{W_{j-1}} \times p_{W_{j+1}} \times \cdots \times p_{W_{2R}} \parallel p_{X'M'_1 \ldots M'_{2R}}\right) \right)
\]

\[
\leq \frac{b}{2R} \sum_{j=1}^{2R} \left( D \left(p_{X'M'_j} \parallel p_{X} \times p_{W_j}\right) - D \left(p_{X'M'_j} \parallel \frac{1}{2R} p_{X'M'_j} + \left(1 - \frac{1}{2R}\right) p_{X'} \times p_{W_j}\right) \right)
\]

\[
\leq \log \left(1 + \frac{2^c}{2R}\right) + \log \frac{1}{p}
\]

\[
\leq 4\varepsilon,
\]

where (a) follows from Fact 1, (b) follows from Fact 2 and (c) follows from Equation 12 and (d) follows since \(\log(1 + x) \leq x\) for all real \(x\) and from choice of \(R\). From Fact 3 we get,

\[
\|p_{X'M'_1 \ldots M'_{2R}} - p_{X} \times p_{W_1} \times \cdots \times p_{W_{2R}}\| \leq 2\sqrt{\varepsilon}.
\]

23
This along with Equation (13) and the triangle inequality for $\ell_1$ distance gives us the desired.

**Proof of Fact 5**: Define,

$$ c_1 := D^x_s(p_{X|MN}||p_X \times p_{UV}), c_2 := D^y_s(p_{X|M}||p_X), c_3 := D^z_s(p_{XN}||p_X \times p_Y), $$

Good_1 := \{ (x, m, n) : \frac{p_{X|MN}(x, m, n)}{p_X(x) p_U(m) p_Y(n)} \leq 2^{c_1} \},

Good_2 := \{ (x, m, n) : \frac{p_{X|M}(x, m)}{p_X(x) p_U(m)} \leq 2^{c_2} \},

Good_3 := \{ (x, m, n) : \frac{p_{X|N}(x, n)}{p_X(x) p_V(n)} \leq 2^{c_3} \},

Good := Good_1 \cap Good_2 \cap Good_3.

This implies (from definitions of $c_1, c_2, c_3$) that $p := \Pr \{ (X, M, N) \in Good \} \geq 1 - 3\varepsilon$. Let us define joint random variables $(X', M', N')$ as follows:

$$ p_{X'M'N'}(x, m, n) = \begin{cases} \frac{p_{X|MN}(x, m, n)}{p_X(x) p_U(m) p_Y(n)} & \text{if } (x, m, n) \in \text{Good}, \\ 0 & \text{otherwise}. \end{cases} $$

We note that $\forall (x, m, n) :$

$$ \frac{p_{X'M'}(x, m)}{p_X(x) p_U(m)} \leq 2^{c_1} p, \quad \frac{p_{X|N'}(x, n)}{p_X(x) p_V(n)} \leq 2^{c_2} p, \quad \frac{p_{X'}(x)}{p_X(x)} \leq 2^{c_3} p. \quad (14) $$

Let us construct joint random variables $(J', K', X', M'_1, \ldots, M'_{2^R_1}, N'_1, \ldots, N'_{2^R_2})$ from $(X', M', N')$ in the same way as we constructed $(J, K, X, M_1, \ldots, M_{2^R_1}, N_1, \ldots, N_{2^R_2})$ from $(X, M, N)$. We note that,

$$ \|p_{X'M'_1 \ldots M'_{2^R_1} N'_1 \ldots N'_{2^R_2}} - p_{X'M_1 \ldots M_{2^R_1} N_1 \ldots N_{2^R_2}}\| = \|p_{X|MN} - p_{X'|M'|N'}\| \leq 12\varepsilon. \quad (15) $$

For notational convenience let us define,

$$ \forall j \in [2^R_1] : \quad p_{U_{-j}} := p_{U_1} \times \ldots \times p_{U_{j-1}} \times p_{U_{j+1}} \times \ldots \times p_{U_{2^R_1}}, $$

$$ \forall k \in [2^R_2] : \quad p_{V_{-k}} := p_{V_1} \times \ldots \times p_{V_{k-1}} \times p_{V_{k+1}} \times \ldots \times p_{V_{2^R_2}}, $$

$$ q_{X'|M'|N'} := \frac{1}{2^{R_1 + R_2}} p_{X'|M'|N'_k} + \frac{1}{2^{R_1}} \left(1 - \frac{1}{2^{R_2}} \right) p_{X'|M'|N_k} \times p_{V_k} \times p_{U_j} \times p_{V_k} + \left(1 - \frac{2^{R_1} + 2^{R_2} - 1}{2^{R_1 + R_2}} \right) p_{X'} \times p_{U_j} \times p_{V_k}. $$

Consider,

$$ D \left( p_{X'M'_1 \ldots M'_{2^R_1} N'_1 \ldots N'_{2^R_2}} || p_X \times p_U \times \ldots \times p_{U_{2^R_1}} \times p_{V_1} \times \ldots \times p_{V_{2^R_2}} \right) $$

$$ = a \frac{1}{2^{R_1 + R_2}} \sum_{j,k} D \left( p_{X'|M'|N'_k} || p_X \times p_{U_j} \times p_{V_k} \right) - D \left( p_{X'|M'|N'_k} \times p_{U_{-j}} \times p_{V_{-k}} || p_X \times p_{U_{2^R_1}} \times p_{V_{2^R_2}} \right) $$

$$ \leq b \frac{1}{2^{R_1 + R_2}} \sum_{j,l} D \left( p_{X'|M'|N'_l} || p_X \times p_{U_j} \times p_{V_k} \right) - D \left( p_{X'|M'|N'_l} || q_{X'|M'|N'} \right) $$

$$ \leq \log \left( 1 + \frac{2^{c_1}}{2^{R_1 + R_2}} + \frac{2^{c_2}}{2^{R_1}} + \frac{2^{c_3}}{2^{R_2}} \right) + \log \frac{1}{p} $$

$$ \leq d \leq 9\varepsilon, $$

where (a) follows from Fact 1, (b) follows from Fact 2, (c) follows from Equation (14) and (d) follows since $\log(1 + x) \leq x$ for all real $x$ and from choice of parameters. From Fact 3 this implies

$$ \|p_{X'M'_1 \ldots M'_{2^R_1} N'_1 \ldots N'_{2^R_2}} - p_{X'} \times p_{U_1} \times \ldots \times p_{U_{2^R_1}} \times p_{V_1} \times \ldots \times p_{V_{2^R_2}}\| \leq 3\sqrt{\varepsilon}. $$
This along with Equation \((15)\) and the triangle inequality for \(\ell_1\) distance gives us the desired.

**Proof of Fact 7:** Let \(\mathcal{A} \subseteq \mathcal{Y} \times \mathcal{M}\) be such that \(\Pr_{p_{YM}} \{\mathcal{A}\} \geq 1 - \varepsilon\), and
\[
c := D_H(p_{YM} \| p_Y \times p_M) = -\log \Pr_{p_Y \times p_M} \{\mathcal{A}\}.
\]

Define \(J'\) to be the first index in \([2^{|R|}]\) such that \((Y, M_{J'}) \in \mathcal{A}\). For the arguments below, let us condition on the event \(J = j\) for some fixed \(j \in [2^{|R|}]\). Consider,
\[
\Pr\{J' \neq j\} \leq \Pr\{(Y, M_j) / \in \mathcal{A}\} + \Pr\{(Y, M_{j'}) \in \mathcal{A}\text{ for some }j' \neq j\}
\]
\[
\leq \varepsilon + 2^{|R|} \cdot 2^{-c} \leq \varepsilon + \delta.
\]
Therefore,
\[
\Pr\{J \neq J'\} = \sum_{j \in [2^{|R|}]} \Pr\{J = j\} \cdot \Pr\{J' \neq j \mid J = j\} \leq \varepsilon + \delta.
\]

**Proof of Fact 8:** For \(i \in \{1, 2, 3\}\), let \(\mathcal{A}_i \subseteq \mathcal{M} \times \mathcal{N}\) be such that \(\Pr_{p_{MN}} \{\mathcal{A}_i\} \geq 1 - \varepsilon\), and
\[
c_1 := D_H(p_{MN} \| p_U \times p_N) = -\log \Pr_{p_U \times p_N} \{\mathcal{A}_1\}
\]
\[
c_2 := D_H(p_{MN} \| p_M \times p_V) = -\log \Pr_{p_M \times p_V} \{\mathcal{A}_2\}
\]
\[
c_3 := D_H(p_{MN} \| p_U \times p_V) = -\log \Pr_{p_U \times p_V} \{\mathcal{A}_3\}.
\]

Let \(\mathcal{A} := \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3\). Define \((J', K')\) to be the first pair of indices (in lexicographic order) in \([2^{|R_1|}] \times [2^{|R_2|}]\) such that \((M_{J'}, N_{K'}) \in \mathcal{A}\). For the arguments below, let us condition on the event \((J, K) = (j, k)\) for some fixed \((j, k) \in [2^{|R_1|}] \times [2^{|R_2|}]\). Consider,
\[
\Pr\{(J', K') \neq (j, k)\} \leq \Pr\{(M_j, N_k) \notin \mathcal{A}\} + \Pr\{(M_{j'}, N_{k'}) \in \mathcal{A}\text{ for some } (j', k') \neq (j, k)\}
\]
\[
\leq 3\varepsilon + 2^{|R_1|} \cdot 2^{-c_1} + 2^{|R_2|} \cdot 2^{-c_2} + 2^{|R_1| + |R_2|} \cdot 2^{-c_3} \leq 3\varepsilon + 3\delta.
\]
Therefore,
\[
\Pr\{(J, K) \neq (J', K')\} = \sum_{(j, k) \in [2^{|R_1|}] \times [2^{|R_2|}]} \Pr\{(J, K) = (j, k)\} \cdot \Pr\{(J', K') \neq (j, k) \mid (J, K) = (j, k)\} \leq 3\varepsilon + 3\delta.
\]