A unified approach to source and message compression

Anurag Anshu* Rahul Jain† Naqueeb Ahmad Warsi‡

Abstract

In this work we consider source and message compression in various network communication scenarios and present a unified approach to arrive at communication bounds. We present our communication bounds in the one-shot setting which imply optimal bounds for these tasks in the asymptotic i.i.d setting. 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 [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 [2, 3, 4, 5]. The other technique that we use is *position based decoding* introduced in [6], which in turn uses *hypothesis testing* between distributions. These two techniques used together allow us to construct all our protocols.

1 Introduction

Source compression is a fundamental task in information theory first studied by Shannon in his landmark paper [7]. This task was later extended to various network settings for example by Slepian and Wolf [8], Wyner [9] and Wyner and Ziv [10]. These works considered the asymptotic, independent and identically distributed (i.i.d.) setting.

In this work we consider source and message compression in various network communication scenarios and present a unified approach to arrive at communication bounds. Message compression is a task when a random variable correlated with the source is sought to be sent with low communication. We start with a one-sender-one-receiver task. We then consider a two-senders-one-receiver task followed by a one-sender-two-receivers task. We combine these two to consider a two-senders-two-receivers task.

We present our communication bounds in the one-shot setting which imply optimal bounds for these tasks in the asymptotic i.i.d setting. 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. Answering these questions has applications in computational complexity as well.

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 [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 [2, 3, 4, 5]. The other technique that we use is *position based decoding* introduced in [6], which in turn uses *hypothesis testing* between distributions. These two techniques used together allow us to construct all our protocols.

*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

Our results

We start with the following one-sender-one-receiver task. For all our results in this section let $\varepsilon > 0$ be a sufficiently small constant which represents an error parameter[1].

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 $\|p_{XYM} - p_{XY\hat{M}}\| \leq O(\sqrt{\varepsilon})$, where $\|.\|$ is the ℓ_1 norm. They are allowed to use shared randomness between them which is independent of XYM at the beginning of the protocol.

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 an priori distribution and M can be viewed as the message Alice wants to send to Bob. It was studied in [2, 4] when the distribution of (X, Y) is product and in [5] for general (X, Y). Here, we present a new protocol for this task using the aforementioned techniques of convex split and position based decoding and show the following achievability result.

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

$$R \geq \min_{\substack{(\tilde{X},\tilde{Y},\tilde{M},T,E):\\ \|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{XYM}\| \leq \delta \\ \tilde{Y} - \tilde{X} - \tilde{M}E}} \left(\mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_T) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{\tilde{Y}\tilde{M}E} \| p_{\tilde{Y}} \times p_T) + O\left(\log \frac{1}{\varepsilon}\right) \right),$$

where E takes values over a set $\mathcal{E}$ and T takes values over set $\mathcal{E} \times \mathcal{M}$. There exists a shared randomness assisted protocol in which Alice communicates R bits to Bob and Bob outputs random variable $\hat{M}$ satisfying $\|p_{XYM} - p_{XY\hat{M}}\| \leq \delta + O(\sqrt{\varepsilon})$.

Please refer to Section 2 for the definitions of $\mathrm{D}_s^{\varepsilon}(\cdot)$ and $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(\cdot)$. The minimization above over $\tilde{X}, \tilde{Y}, \tilde{M}$ and E (which we refer to as *extension* of $\tilde{M}$) 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.

Theorem 2 (Converse for Task 1)**.** *Any communication protocol for Task 1 must satisfy:*

$$R \geq \min_{\substack{(\tilde{X},\tilde{Y},\tilde{M},U,E):\\ \|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{XYM}\| \leq 3\sqrt{\varepsilon} \\ \tilde{Y} - \tilde{X} - \tilde{M}E}} \left(\mathrm{D}_s^{3\sqrt{\varepsilon}}(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_U) - \mathrm{D}_{\mathrm{H}}^{3\sqrt{\varepsilon}}(p_{\tilde{Y}\tilde{M}E} \| p_{\tilde{Y}} \times p_U) - O\left(\log \frac{1}{\varepsilon}\right) \right),$$

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

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 $(\hat{M}, \hat{N})$ such that $\|p_{XYMN} - p_{XY\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$. Shared randomness is allowed between Alice and Charlie and between Bob and Charlie.

We show the following achievability result for this task.

Theorem 3 (Achievability for Task 2)**.** *Let R_A, R_B be natural numbers such that,*

$$R_A \geq \mathrm{D}_s^{\varepsilon}(p_{XM} \| p_X \times p_M) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MN} \| p_M \times p_N) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_B \geq \mathrm{D}_s^{\varepsilon}(p_{YN} \| p_Y \times p_N) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MN} \| p_M \times p_N) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_A + R_B \geq \mathrm{D}_s^{\varepsilon}(p_{XM} \| p_X \times p_M) + \mathrm{D}_s^{\varepsilon}(p_{YN} \| p_Y \times p_N) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MN} \| p_M \times p_N) + O\left(\log \frac{1}{\varepsilon}\right).$$

[1] We do not attempt to optimize constants appearing in this paper.

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 $\|p_{XYMN} - p_{XY\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$.

Remark: In our result above we can optimize over $(\tilde{X}, \tilde{Y}, \tilde{M}, \tilde{N}, E, T)$ as in Theorem 1. 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 $(\hat{M}, \hat{N})$ such that $\|p_{XYZMN} - p_{XYZ\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$. Shared randomness is allowed between Alice and Charlie and between Bob and Charlie.

We show the following achievability result for this task.

Theorem 4 (Achievability for Task 3)**.** *Let R_A, R_B be natural numbers such that,*

$$R_A \geq \mathrm{D}_s^{\varepsilon}(p_{XM}\|p_X \times p_M) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MNZ}\|p_M \times p_{NZ}) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_B \geq \mathrm{D}_s^{\varepsilon}(p_{YN}\|p_Y \times p_N) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MZN}\|p_{MZ} \times p_N) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_A + R_B \geq \mathrm{D}_s^{\varepsilon}(p_{XM}\|p_X \times p_M) + \mathrm{D}_s^{\varepsilon}(p_{YN}\|p_Y \times p_N) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MNZ}\|p_M \times p_N \times p_Z) + O\left(\log \frac{1}{\varepsilon}\right).$$

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 $\|p_{XYZMN} - p_{XYZ\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$.

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 $\|p_{XMN} - p_{X\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$. Shared randomness is allowed between Alice and Charlie and between Alice and Bob.

We show the following achievability result for this task.

Theorem 5 (Achievability for Task 4)**.** *Let R_B, R_C be natural numbers such that,*

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

$$R_C \geq \mathrm{D}_s^{\varepsilon}(p_{XN}\|p_X \times p_N) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_B + R_C \geq \mathrm{D}_s^{\varepsilon}(p_{XMN}\|p_X \times p_M \times p_N) + O\left(\log \frac{1}{\varepsilon}\right).$$

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 $\hat{M}$ and Charlie outputs $\hat{N}$ such that $\|p_{XMN} - p_{X\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$.

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 $\hat{M}$ and $\hat{N}$ respectively such that $\|p_{XYZMN} - p_{XYZ\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$. Shared randomness is allowed between Alice and Bob and between Alice and Charlie.

We show the following achievability result for this task.

Theorem 6 (Achievability for Task 5). *Let R_B, R_C be natural numbers such that,*

$$R_B \geq \mathrm{D}_s^{\varepsilon}(p_{XM}\|p_X \times p_M) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MY}\|p_M \times p_Y) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_C \geq \mathrm{D}_s^{\varepsilon}(p_{XN}\|p_X \times p_N) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{NZ}\|p_N \times p_Z) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_B + R_C \geq \mathrm{D}_s^{\varepsilon}(p_{XMN}\|p_X \times p_M \times p_N) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MY}\|p_M \times p_Y) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{NZ}\|p_N \times p_Z) + O\left(\log \frac{1}{\varepsilon}\right).$$

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 $\hat{M}$ and Charlie outputs $\hat{N}$ such that $\|p_{XYZMN} - p_{XYZ\hat{M}\hat{N}}\| \leq O(\sqrt{\varepsilon})$.

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, Dave, Bob 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 $(\hat{M}_{11}, \hat{M}_{21})$ and Charlie outputs $(\hat{M}_{12}, \hat{M}_{22})$ such that,

$$\|p_{X_1 M_{11} M_{12} X_2 M_{21} M_{22} Y_1 Y_2} - p_{X_1 \hat{M}_{11} \hat{M}_{12} X_2 \hat{M}_{21} \hat{M}_{22} Y_1 Y_2}\| \leq O(\sqrt{\varepsilon}).$$

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

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

Theorem 7 (Achievability for Task 6). *Let $R_1^{(1)}, R_2^{(1)}, R_1^{(2)}, R_2^{(2)}$ be natural numbers such that for $i, j \in \{1, 2\}$,*

$$R_j^{(i)} \geq \mathrm{D}_s^{\varepsilon}(p_{X_i M_{ij}}\|p_{X_i} \times p_{M_{ij}}) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{M_{ij} Y_j}\|p_{M_{ij}} \times p_{Y_j}) + O\left(\log \frac{1}{\varepsilon}\right),$$

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

$$R_j^{(i)} + R_l^{(k)} \geq \mathrm{D}_s^{\varepsilon}(p_{X_i M_{ij}}\|p_{X_i} \times p_{M_{ij}}) + \mathrm{D}_s^{\varepsilon}(p_{X_k M_{kl}}\|p_{X_k} \times p_{M_{kl}}) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{M_{ij} Y_j}\|p_{M_{ij}} \times p_{Y_j}) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{M_{kl} Y_l}\|p_{M_{kl}} \times \times p_{Y_l}) + O\left(\log \frac{1}{\varepsilon}\right),$$

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

$$R_j^{(i)} + R_l^{(i)} + R_j^{(k)} \geq \mathrm{D}_s^{\varepsilon}(p_{X_i M_{ij} M_{il}}\|p_{X_i} \times p_{M_{ij}} \times p_{M_{il}}) + \mathrm{D}_s^{\varepsilon}(p_{X_k M_{kj}}\|p_{X_k} \times p_{M_{kj}}) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{M_{ij} M_{kj} Y_j}\|p_{M_{ij}} \times p_{M_{kj}} \times p_{Y_j}) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{M_{il} Y_l}\|p_{M_{il}} \times p_{Y_l}) + O\left(\log \frac{1}{\varepsilon}\right),$$

and,

$$R_1^{(1)} + R_2^{(1)} + R_1^{(2)} + R_2^{(2)} \geq \mathrm{D}_s^{\varepsilon}(p_{X_1 M_{11} M_{12}}\|p_{X_1} \times p_{M_{11}} \times p_{M_{12}}) + \mathrm{D}_s^{\varepsilon}(p_{X_2 M_{21} M_{22}}\|p_{X_2} \times p_{M_{21}} \times p_{M_{22}}) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{M_{11} M_{21} Y_1}\|p_{M_{11}} \times p_{M_{21}} \times p_{Y_1}) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{M_{12} M_{22} Y_2}\|p_{M_{12}} \times p_{M_{22}} \times p_{Y_2}) + O\left(\log \frac{1}{\varepsilon}\right).$$

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

$$\|p_{X_1 M_{11} M_{12} X_2 M_{21} M_{22} Y_1 Y_2} - p_{X_1 \hat{M}_{11} \hat{M}_{12} X_2 \hat{M}_{21} \hat{M}_{22} Y_1 Y_2}\| \leq O(\sqrt{\varepsilon}).$$

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 $d : \mathcal{X} \times \mathcal{Z} \to (0,\infty)$ 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 [10] and in the non-i.i.d. setting in [11]. We show the following achievability result which follows as a corollary of Theorem 1.

Corollary 1 (Achievability for Task 7)**.** *Let $\delta \geq 0$. Let R be a natural number such that,*

$$R \geq \min_{M,f} \left(\mathrm{D}_s^{\varepsilon}(p_{XM} \| p_X \times p_M) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{YM} \| p_Y \times p_M) + O\left(\log \frac{1}{\varepsilon}\right)\right), \tag{1}$$

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 1 in which Alice send R bits to Bob and at the end Bob is able to generate $\hat{M}$ such that $\|p_{XYM} - p_{XY\hat{M}}\| \leq O(\sqrt{\varepsilon})$. Bob then outputs $Z = f(Y, \hat{M})$. Consider,

$$\begin{aligned}\Pr\left\{d(X, f(Y,\hat{M})) \geq k\right\} &\leq \Pr\{d(X, f(Y,M)) \geq k\} + \|p_{XYM} - p_{XY\hat{M}}\| \\ &\leq \delta + O(\sqrt{\varepsilon}).\end{aligned}$$

This protocol uses shared randomness between Alice and Bob and $\Pr\left\{d(X, f(Y,\hat{M})) \geq k\right\} \leq \delta + O(\sqrt{\varepsilon})$ averaged over the shared randomness. Hence there exists a fixed shared string between Alice and Bob, conditioned on which $\Pr\left\{d(X, f(Y,\hat{M})) \geq k\right\} \leq \delta + O(\sqrt{\varepsilon})$. 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 [8] in the asymptotic setting. Its one-shot version was studied in [12, 13].

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 $\Pr\left\{(X,Y) \neq (\hat{X},\hat{Y})\right\} \leq O(\sqrt{\varepsilon})$. 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 3.

Corollary 2 (Achievability for Task 8)**.** *Let XX represent two copies of X and YY represent two copies of Y. Let R_A, R_B be natural numbers such that,*

$$\begin{aligned}R_A &\geq \mathrm{D}_s^{\varepsilon}(p_{XX} \| p_X \times p_X) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{XY} \| p_X \times p_Y) + O\left(\log \frac{1}{\varepsilon}\right), \\ R_B &\geq \mathrm{D}_s^{\varepsilon}(p_{YY} \| p_Y \times p_Y) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{XY} \| p_X \times p_Y) + O\left(\log \frac{1}{\varepsilon}\right), \\ R_A + R_B &\geq \mathrm{D}_s^{\varepsilon}(p_{XX} \| p_X \times p_X) + \mathrm{D}_s^{\varepsilon}(p_{YY} \| p_Y \times p_Y) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{XY} \| p_X \times p_Y) + O\left(\log \frac{1}{\varepsilon}\right).\end{aligned}$$

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 $\Pr\left\{(X,Y) \neq (\hat{X},\hat{Y})\right\} \leq O(\sqrt{\varepsilon})$.

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

Next we consider the following task which was first studied by Wyner [9] in the asymptotic and i.i.d. setting, subsequently in the *information-spectrum* setting by Miyakaye and Kanaya [14] and in the one-shot case in [12, 15].

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 $\Pr\left\{X \neq \hat{X}\right\} \leq O(\sqrt{\varepsilon})$.

We show the following achievability result for this task which follows as a corollary from Theorem 3.

Corollary 3 (Achievability for Task 9)**.** *Let XX represent two copies of X. Let R_A, R_B be natural numbers such that,*

$$R_A \geq \mathrm{D}_s^{\varepsilon}(p_{XX}\|p_X \times p_X) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{XN}\|p_X \times p_N) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_B \geq \mathrm{D}_s^{\varepsilon}(p_{YN}\|p_Y \times p_N) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{XN}\|p_X \times p_N) + O\left(\log \frac{1}{\varepsilon}\right),$$

$$R_A + R_B \geq \mathrm{D}_s^{\varepsilon}(p_{XX}\|p_X \times p_X) + \mathrm{D}_s^{\varepsilon}(p_{YN}\|p_Y \times p_N) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{XN}\|p_X \times p_N) + O\left(\log \frac{1}{\varepsilon}\right),$$

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 $\Pr\left\{X \neq \hat{X}\right\} \leq O(\sqrt{\varepsilon})$.

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

We note that all our results above imply the corresponding known results in the asymptotic settings [8, 9, 10, 14]. Our results also imply the results of [12, 13] in the one-shot setting (by changing shared randomness to be uniform in Task 2). The one-shot bounds in Corollary 3 asymptotically yield a rate region which a priori appears to be a superset of the rate region obtained by [9], however the two are the same due to the optimality of the latter.

Organization

In the next section we present a few information theoretic preliminaries. In Section 3 we present proofs of our results. 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 $X - Y - Z$, iff for each $x \in \mathcal{X}$, $Y|(X = x)$ and $Z|(X = x)$ are independent. We define various information theoretic quantities below.

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

- *ℓ_1 distance:* $\|p_X - p_{X'}\| := \sum_x |p_X(x) - p_{X'}(x)|$.
- *Relative entropy:* $\mathrm{D}(p_X \| p_{X'}) := \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{p_{X'}(x)}$.
- *Max divergence:* $\mathrm{D}_\infty(p_X \| p_{X'}) := \max_x \log \frac{p_X(x)}{p_{X'}(x)}$.
- *Smooth max divergence:* $\mathrm{D}^\varepsilon_\infty(p_X \| p_{X'}) := \min_{\|p_{X''} - p_X\| \leq \varepsilon} \mathrm{D}_\infty(p_{X''} \| p_{X'})$.
- *Max information spectrum divergence:* $\mathrm{D}^\varepsilon_s(p_X \| p_{X'}) := \min \left\{ a : \Pr_{x \leftarrow p_X} \left\{ \frac{p_X(x)}{p_{X'}(x)} > 2^a \right\} < \varepsilon \right\}$.
- *Smooth hypothesis testing divergence:* $\mathrm{D}^\varepsilon_{\mathrm{H}}(p_X \| p_{X'}) := \max \left\{ -\log(\Pr_{p_{X'}}\{\mathcal{A}\}) \mid \mathcal{A} \subseteq \mathcal{X}, \Pr_{p_X}\{\mathcal{A}\} \geq 1 - \varepsilon \right\}$.

We will use the following facts.

Fact 1 ([16], Proposition 13)**.** *Let X, X' be random variables over $\mathcal{X}$. It holds that,*

$$\mathrm{D}^\varepsilon_\infty(p_X \| p_{X'}) \geq \mathrm{D}^{2\varepsilon}_s(p_X \| p_{X'}) - 2 \log \frac{2}{\varepsilon}.$$

Fact 2 ([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 $\{P_i\}_i$. It holds that,*

$$\mathrm{D}(P \| Q) = \sum_i \lambda_i \left(\mathrm{D}(P_i \| Q) - \mathrm{D}(P_i \| P) \right).$$

Fact 3 (Monotonicity of relative entropy [17])**.** *Let (X, Y, Z) be jointly distributed random variables. It holds that,*

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

Fact 4 (Pinsker's inequality [17])**.** *Let P and Q be two distributions over the set $\mathcal{X}$. It holds that,*

$$\|P - Q\| \leq 2 \cdot \sqrt{\mathrm{D}(P \| Q)}.$$

Fact 5 (Monotonicity under maps [17])**.** *Let X be a random variable distributed over the set $\mathcal{X}$. Let $f : \mathcal{X} \to \mathcal{Z}$ 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 6 (Convex-split lemma [1])**.** *Let $\varepsilon \in (0, \frac{1}{4})$. 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,*

$$R \geq \mathrm{D}^\varepsilon_s(p_{XM} \| p_X \times p_W) + 4 \log \frac{1}{\varepsilon}.$$

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

$$\begin{aligned} &\Pr\{(X, M_1, \ldots, M_{2^R}) = (x, m_1, \ldots, m_{2^R}) \mid J = j\} \\ &= p_{XM}(x, m_j) \cdot p_W(m_1) \cdots p_W(m_{j-1}) \cdot p_W(m_{j+1}) \cdots p_W(m_{2^R}). \end{aligned}$$

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

$$\|p_{X M_1 \ldots M_{2^R}} - p_X \times p_{W_1} \times \ldots \times p_{W_{2^R}}\| \leq 6\sqrt{\varepsilon}.$$

We also need the following extension of this lemma whose quantum version was shown in [18]. The proof of the classical version that we consider follows on similar lines and is deferred to Appendix.

Fact 7 (Bipartite convex-split lemma)**.** *Let $\varepsilon \in (0, \frac{1}{16})$. Let (X, M, N) (jointly distributed over $\mathcal{X} \times \mathcal{M} \times \mathcal{N}$), U (distributed over $\mathcal{M}$) and V (distributed over $\mathcal{N}$) be random variables. Let R_1, R_2 be natural numbers such that,*

$$R_1 \geq \mathrm{D}_s^{\varepsilon}(p_{XM} \| p_X \times p_U) + 8 \log \frac{1}{\varepsilon},$$
$$R_2 \geq \mathrm{D}_s^{\varepsilon}(p_{XN} \| p_X \times p_V) + 8 \log \frac{1}{\varepsilon},$$
$$R_1 + R_2 \geq \mathrm{D}_s^{\varepsilon}(p_{XMN} \| p_X \times p_U \times p_V) + 8 \log \frac{1}{\varepsilon}.$$

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 $(J, K, X, M_1, \ldots, M_{2^{R_1}}, N_1, \ldots, N_{2^{R_2}})$ be distributed as follows:

$$\begin{aligned}&\Pr\{(X, M_1, \ldots, M_{2^{R_1}}, N_1, \ldots, N_{2^{R_2}}) = (x, m_1, \ldots, m_{2^{R_1}}, n_1, \ldots, n_{2^{R_2}}) \mid J = j, K = k\} \\ &= p_{XMN}(x, m_j, n_k) \cdot p_U(m_1) \cdots p_U(m_{j-1}) \cdot p_U(m_{j+1}) \cdots p_U(m_{2^{R_1}}) \cdot \\ &\qquad p_V(n_1) \cdots p_V(n_{k-1}) \cdot p_V(n_{k+1}) \cdots p_V(n_{2^{R_2}}).\end{aligned}$$

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

$$\| p_{X M_1 \ldots M_{2^{R_1}} N_1 \ldots N_{2^{R_2}}} - p_X \times p_{U_1} \times \ldots \times p_{U_{2^{R_1}}} \times p_{V_1} \times \ldots \times p_{V_{2^{R_2}}} \| \leq 15\sqrt{\varepsilon}.$$

The other main tool that we use is the position based decoding from [6] 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 8 (Position based decoding [6])**.** *Let $\varepsilon \in (0, \frac{1}{4})$. Let (Y, M) (jointly distributed over $\mathcal{Y} \times \mathcal{M}$) and W (distributed over $\mathcal{M}$) be random variables. Let R be a natural number such that,*

$$R \leq \max\left\{ \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{YM} \| p_Y \times p_W) - \log \frac{1}{\varepsilon}, 0 \right\}.$$

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

$$\begin{aligned}&\Pr\{(Y, M_1, M_2, \ldots, M_{2^R}) = (y, m_1, \ldots, m_{2^R}) \mid J = j\} \\ &= p_{YM}(y, m_j) \cdot p_W(m_1) \cdots p_W(m_{j-1}) \cdot p_W(m_{j+1}) \cdots p_W(m_{2^R}).\end{aligned}$$

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

We will also need the following extension of this decoding strategy shown in [18] 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 9 (Bipartite position based decoding [18])**.** *Let $\varepsilon \in (0, \frac{1}{16})$. Let (M, N) (jointly distributed over $\mathcal{M} \times \mathcal{N}$). Let R_1, R_2 be natural numbers such that,*

$$R_1 + R_2 \leq \max\left\{ \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MN} \| p_M \times p_N) - \log \frac{1}{\varepsilon}, 0 \right\}.$$

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

$$\begin{aligned}&\Pr\{(M_1 \ldots M_{2^{R_1}} N_1 \ldots N_{2^{R_2}}) = (m_1, \ldots, m_{2^{R_1}}, n_1, \ldots, n_{2^{R_2}}) \mid J = j, K = k\} \\ &= p_{MN}(m_j, n_k) \cdot p_M(m_1) \cdots p_M(m_{j-1}) \cdot p_M(m_{j+1}) \cdots p_M(m_{2^{R_1}}) \cdot \\ &\qquad p_N(n_1) \cdots p_N(n_{k-1}) \cdot p_N(n_{k+1}) \cdots p_N(n_{2^{R_2}}).\end{aligned}$$

There is a procedure to produce random variables (J', K') from $(M_1, \ldots, M_{2^{R_1}}, N_1, \ldots, N_{2^{R_2}})$ such that $\Pr\{(J, K) \neq (J', K')\} \leq 2\varepsilon$.

3 Proofs of our results

In this section we present proofs of our results mentioned in the Introduction 1.

Proof of Theorem 1: Let $(\tilde{X}, \tilde{Y}, \tilde{M}, T, E)$ be such that $\|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{XYM}\| \leq \delta$ and $\tilde{Y} - \tilde{X} - (\tilde{M}, E)$. Let R, r be natural numbers such that,

$$r \leq \max\left\{ \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{\tilde{Y}\tilde{M}E} \| p_{\tilde{Y}} \times p_T) - \log\frac{1}{\varepsilon}, 0 \right\},$$
$$R + r \geq \mathrm{D}_{s}^{\varepsilon}(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_T) + 2\log\frac{1}{\varepsilon}.$$

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 $\mathcal{B}(j)$ denote the subset corresponding to j. Let us invoke convex-split lemma (Fact 6) with $X \leftarrow \tilde{X}, M \leftarrow (\tilde{M}, E), W \leftarrow T$ and $R \leftarrow R + r$ to obtain joint random variables $(J, \tilde{X}, \tilde{M}_1, \ldots, \tilde{M}_{2^{R+r}})$. Let us first consider a fictitious protocol $\mathcal{P}'$ as follows.

Fictitious protocol $\mathcal{P}'$: Alice possesses random variables $(\tilde{X}, \tilde{M})$, Bob possesses random variable $\tilde{Y}$ and they share $(\tilde{M}_1, \ldots, \tilde{M}_{2^{R+r}})$ as public randomness (from the joint random variables $(\tilde{X}, \tilde{M}_1, \ldots, \tilde{M}_{2^{R+r}})$ above).

Alice's operations: Alice generates J from $(\tilde{X}, \tilde{M}_1, \ldots, \tilde{M}_{2^{R+r}})$, using the conditional distribution of J given $(\tilde{X}, \tilde{M}_1, \ldots, \tilde{M}_{2^{R+r}})$, and communicates $\mathcal{B}(J)$ to Bob. This can be done using R bits of communication.

Bob's operations: Bob performs position based decoding as in Fact 8 using $\tilde{Y}$ and the subset $\mathcal{B}(J)$, by letting $Y \leftarrow \tilde{Y}, M \leftarrow (\tilde{M}, E), W \leftarrow T$ and $R \leftarrow r$, and determines J'. Let $(M', E') := \tilde{M}_{J'}$. Bob outputs M'.

From Fact 8 we have $\Pr\{J \neq J'\} \leq 2\varepsilon$ and hence,

$$\begin{aligned}\|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{\tilde{X}\tilde{Y}M'}\| &\leq \Pr\{J = J'\}\|p_{\tilde{X}\tilde{Y}\tilde{M} \mid J=J'} - p_{\tilde{X}\tilde{Y}M' \mid J=J'}\| \\ &\quad + \Pr\{J \neq J'\}\|p_{\tilde{X}\tilde{Y}\tilde{M} \mid J\neq J'} - p_{\tilde{X}\tilde{Y}M' \mid J\neq J'}\| \\ &\leq 0 + 4\varepsilon = 4\varepsilon. \end{aligned} \tag{2}$$

Now consider the another fictitious protocol $\mathcal{P}''$.

Fictitious protocol $\mathcal{P}''$: Alice possesses random variables $(\tilde{X}, \tilde{M})$ and Bob possesses random variable $\tilde{Y}$. Alice and Bob share 2^{R+r} i.i.d. copies of the random variable T, denoted $\{T_1, T_2, \ldots, T_{2^{R+r}}\}$. Alice and Bob proceed as in $\mathcal{P}'$. Therefore the only difference in $\mathcal{P}''$ and $\mathcal{P}'$ is shared randomness. Let M'' be the output of Bob in $\mathcal{P}''$.

Consider,

$$\begin{aligned}\|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{\tilde{X}\tilde{Y}M''}\| &\overset{a}{\leq} \|p_{\tilde{X}\tilde{M}_1\ldots\tilde{M}_{2^{R+r}}} - p_{\tilde{X}} \times p_{T_1} \times \ldots \times p_{T_{2^{R+r}}}\| + \|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{\tilde{X}\tilde{Y}M'}\| \\ &\overset{b}{\leq} 6\sqrt{\varepsilon} + \|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{\tilde{X}\tilde{Y}M'}\| \\ &\overset{c}{\leq} 10\sqrt{\varepsilon}, \end{aligned}$$

where (a) follows from triangle inequality and the property $\tilde{M} - \tilde{X} - \tilde{Y}$; (b) follows from convex-split lemma and (c) follows from Equation (2). Now consider the actual protocol $\mathcal{P}$.

Actual protocol $\mathcal{P}$: Alice possesses random variables (X, M) and Bob possesses random variable Y. Alice and Bob share 2^{R+r} i.i.d. copies of the random variable T, denoted $\{T_1, T_2, \ldots, T_{2^{R+r}}\}$. Alice and Bob proceed as in $\mathcal{P}''$. Therefore the only difference in $\mathcal{P}$ and $\mathcal{P}''$ is the inputs of Alice and Bob. Let $\hat{M}$ be the output of Bob in $\mathcal{P}$.

Consider,

$$\begin{aligned}\|p_{XYM} - p_{XY\hat{M}}\| &\overset{a}{\leq} \|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{XYM}\| + \|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{\tilde{X}\tilde{Y}M''}\| \\ &\leq \delta + 10\sqrt{\varepsilon}, \end{aligned}$$

where (a) follows from Fact 5 and triangle inequality. This shows the desired. □

Proof of Theorem 3: 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 [19, Appendix D] and the constraints in the statement of the Theorem),

$$\begin{aligned} R_A + r_A &\geq \mathrm{D}_s^{\varepsilon}(p_{XM} \| p_X \times p_M) + 2\log\frac{1}{\varepsilon}, \\ R_B + r_B &\geq \mathrm{D}_s^{\varepsilon}(p_{YN} \| p_Y \times p_N) + 2\log\frac{1}{\varepsilon}, \\ r_A + r_B &\leq \max\{\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MN} \| p_M \times p_N) - \log\frac{1}{\varepsilon}, 0\}. \end{aligned}$$

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 $\mathcal{B}(j)$ 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 $\mathcal{B}(k)$ denote the subset corresponding to k.

Let us invoke bipartite convex-split lemma (Lemma 7) with $X \leftarrow (X,Y), M \leftarrow M, N \leftarrow N, U \leftarrow M, V \leftarrow N, R_1 \leftarrow R_A + r_A$ 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 $\mathcal{B}(J)$ 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 $\mathcal{B}(K)$ to Charlie. This can be done using R_B bits of communication.

Charlie's operations: Charlie performs bipartite position based decoding as in Fact 9 inside the subset $\mathcal{B}(J) \times \mathcal{B}(K)$, by letting $M \leftarrow M, N \leftarrow N, R_A \leftarrow r_A$ 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 9 we have,

$$\|p_{XYMN} - p_{XYM'N'}\| \leq 2 \cdot \Pr\{(J,K) \neq (J',K')\} \leq 4\varepsilon. \tag{3}$$

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 M, denoted $\left\{M_1, M_2, \ldots, M_{2^{R_A+r_A}}\right\}$. Bob and Charlie share $2^{R_B+r_B}$ i.i.d. copies of the random variable N, denoted $\left\{N_1, N_2, \ldots, N_{2^{R_B+r_B}}\right\}$. Alice, Bob and Charlie proceed as in $\mathcal{P}'$. Therefore the only difference in $\mathcal{P}$ and $\mathcal{P}'$ is shared randomness. Let $(\hat{M}, \hat{N})$ represent Charlie's outputs in $\mathcal{P}$.

From convex-split lemma

$$\|p_{XYM_1 \ldots M_{2^{R_A+r_A}} N_1 \ldots N_{2^{R_B+r_B}}} - p_X \times p_Y \times p_{M_1} \times \ldots \times p_{M_{2^{R_A+r_A}}} \times p_{N_1} \times \ldots \times p_{N_{2^{R_B+r_B}}}\| \leq 12\sqrt{\varepsilon}.$$

From Fact 5, triangle inequality for ℓ_1 norm and Equation (3) we have,

$$\|p_{XYMN} - p_{XY\hat{M}\hat{N}}\| \leq 4\varepsilon + 12\sqrt{\varepsilon} \leq 16\sqrt{\varepsilon}.$$

This shows the desired. □

Proof of Theorem 4: The proof follows on similar lines as the proof of Theorem 3 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 [19, Appendix D] and the constraints in the statement of the Theorem),

$$\begin{aligned}
R_A + r_A &\geq \mathrm{D}_s^{\varepsilon}(p_{XM} \| p_X \times p_M) + O\left(\log \frac{1}{\varepsilon}\right), \\
R_B + r_B &\geq \mathrm{D}_s^{\varepsilon}(p_{YN} \| p_Y \times p_N) + O\left(\log \frac{1}{\varepsilon}\right), \\
r_A &\leq \max\left\{\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MNZ} \| p_M \times p_{NZ}) - O\left(\log \frac{1}{\varepsilon}\right), 0\right\}, \\
r_B &\leq \max\left\{\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MZN} \| p_{MZ} \times p_N) - O\left(\log \frac{1}{\varepsilon}\right), 0\right\}, \\
r_A + r_B &\leq \max\left\{\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MNZ} \| p_M \times p_N \times p_Z) - O\left(\log \frac{1}{\varepsilon}\right), 0\right\}.
\end{aligned}$$

Let $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3 \subseteq \mathcal{M} \times \mathcal{N} \times \mathcal{Z}$ be such that $\Pr_{p_{MNZ}}\{\mathcal{A}_i\} \geq 1 - \varepsilon$ for all $i \in \{1, 2, 3\}$ and

$$\begin{aligned}
\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MNZ} \| p_M \times p_{NZ}) &= -\log \Pr_{p_M \times p_{NZ}}\{\mathcal{A}_1\}; \\
\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MZN} \| p_{MZ} \times p_N) &= -\log \Pr_{p_{MZ} \times p_N}\{\mathcal{A}_2\}; \\
\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MNZ} \| p_M \times p_N \times p_Z) &- \log \Pr_{p_M \times p_N \times p_Z}\{\mathcal{A}_3\}.
\end{aligned}$$

Define $\mathcal{A} := \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$.

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 3.

Charlie's operations: Charlie on receiving $\mathcal{B}(J)$ and $\mathcal{B}(K)$ from Alice and Bob respectively, performs bipartite position based decoding (similar to Fact 9) involving Z and the random variables in the subset $\mathcal{B}(J) \times \mathcal{B}(K)$. He finds the first pair (J', K') (in lexicographic order) such that $(Z, M_{J'}, N_{K'}) \in \mathcal{A}$ and outputs $(\hat{M}, \hat{N}) := (M_{J'}, N_{K'})$.

Using arguments as in the proof of Fact 9 we get $\Pr\{(J, K) \neq (J', K')\} = O(\varepsilon)$. Now using Fact 5 and triangle inequality for ℓ_1 norm it can be argued that $\|p_{XYMN} - p_{XY\hat{M}\hat{N}}\| = O(\sqrt{\varepsilon})$. □

Proof of Theorem 5: Let us invoke bipartite convex-split lemma (Lemma 7) with $X \leftarrow X, M \leftarrow M, N \leftarrow N, U \leftarrow M, V \leftarrow N, R_1 \leftarrow R_B$ and $R_2 \leftarrow R_C$ to obtain joint random variables $(J, K, X, M_1, \ldots, M_{2^{R_B}}, N_1, \ldots, N_{2^{R_C}})$.

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

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

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

Bob's operations: Bob performs position based decoding as in Fact 8, by letting $Y \leftarrow Y, M \leftarrow M$ and $R \leftarrow R_B$ and determines J'. Bob outputs $M' := M_{J'}$.

Charlie's operations: Charlie performs position based decoding as in Fact 8, by letting $Y \leftarrow Z, M \leftarrow N$ and $R \leftarrow R_C$ and determines K'. Bob outputs $N' := N_{K'}$.

From Fact 8 we have,

$$\|p_{XMN} - p_{XM'N'}\| \leq 2(\Pr\{J \neq J'\} + \Pr\{K \neq K'\}) \leq 8\varepsilon.$$

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

Actual protocol $\mathcal{P}$: Alice and Bob share 2^{R_B} i.i.d. copies of the random variable M, denoted $\left\{M_1, M_2, \ldots, M_{2^{R_B}}\right\}$. Alice and Charlie share 2^{R_C} i.i.d. copies of the random variable N, denoted $\left\{N_1, N_2, \ldots, N_{2^{R_C}}\right\}$. Alice, Bob and

Charlie proceed as in $\mathcal{P}'$. Therefore the only difference in $\mathcal{P}$ and $\mathcal{P}'$ is shared randomness. Let $(\hat{M}, \hat{N})$ represent Bob and Charlie's outputs respectively in $\mathcal{P}$.

From bipartite convex-split lemma (Lemma 7),

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

From Fact 5, triangle inequality for ℓ_1 norm and Equation (4) we have,

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

□

Proof of Theorem 6: The proof follows on similar lines as the proof of Theorem 5 and we provide a proof sketch here. Let (R_B, R_C, r_B, r_C) be natural numbers such that,

$$\begin{aligned}
R_B + r_b &\geq \mathrm{D}_s^{\varepsilon}(p_{XM}\|p_X \times p_M) + O\left(\log\frac{1}{\varepsilon}\right),\\
r_b &\leq \max\left\{\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MY}\|p_M \times p_Y) - O\left(\log\frac{1}{\varepsilon}\right), 0\right\},\\
R_C + r_c &\geq \mathrm{D}_s^{\varepsilon}(p_{XN}\|p_X \times p_N) + O\left(\log\frac{1}{\varepsilon}\right),\\
r_c &\leq \max\left\{\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{NZ}\|p_N \times p_Z) - O\left(\log\frac{1}{\varepsilon}\right), 0\right\},\\
R_B + R_C + r_b + r_c &\geq \mathrm{D}_s^{\varepsilon}(p_{XMN}\|p_X \times p_M \times p_N) + O\left(\log\frac{1}{\varepsilon}\right).
\end{aligned}$$

Let $\mathcal{A}_1 \subseteq \mathcal{Y} \times \mathcal{M}$ and $\mathcal{A}_2 \subseteq \mathcal{Z} \times \mathcal{N}$ be such that $\Pr_{p_{YM}}\{\mathcal{A}_1\} \geq 1-\varepsilon$ and $\Pr_{p_{ZN}}\{\mathcal{A}_2\} \geq 1-\varepsilon$ and,

$$\begin{aligned}
\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MY}\|p_M \times p_Y) &= -\log \Pr_{p_M \times p_Y}\{\mathcal{A}_1\},\\
\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{NZ}\|p_N \times p_Z) &= -\log \Pr_{p_N \times p_Z}\{\mathcal{A}_2\}.
\end{aligned}$$

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 Alice and Bob. For $j \in [2^{R_B+r_B}]$, let $\mathcal{B}(j)$ denote the subset corresponding to j. Similarly let us divide $[2^{R_C+r_C}]$ into 2^{R_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 $\mathcal{B}(k)$ 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 M, denoted $\left\{M_1, M_2, \ldots, M_{2^{R_B+r_b}}\right\}$. Alice and Charlie share $2^{R_C+r_c}$ i.i.d. copies of the random variable N, denoted $\left\{N_1, N_2, \ldots, N_{2^{R_C+r_c}}\right\}$.

Alice's operations: Alice generates (J, K) as in protocol $\mathcal{P}$ in the proof of Theorem 5. She communicates $\mathcal{B}(J)$ to Bob (using R_B bits) and $\mathcal{B}(K)$ to Charlie (using R_C bits).

Bob's operations: Bob performs position based decoding as in Fact 8, by letting $Y \leftarrow Y, M \leftarrow M$ 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 8, by letting $Y \leftarrow Z, M \leftarrow N$ and $R \leftarrow R_C$ and determines K'. Charlie outputs $\hat{N} := N_{K'}$.

Using arguments as in the proof of Fact 8 we get $\Pr\{(J,K) \neq (J',K')\} = O(\sqrt{\varepsilon})$. Now using Fact 5 and triangle inequality for ℓ_1 norm it can be argued that $\|p_{XYZMN} - p_{XYZ\hat{M}\hat{N}}\| = O(\sqrt{\varepsilon})$.

□

4 Optimality of the protocol for Task 1

The aim of this section is to relate our achievability result 1 with the result of Braverman and Rao [5]. Towards this, we first discuss the result of Braverman and Rao [5]. We start with the lemma below, which is a rephrased version

of Theorem 2.1 in [5]. It may be noted that Braverman and Rao were considering expected communication cost, whereas we are considering the worst case communication cost. Thus, we have stated the lemma below accordingly.

Lemma 1 ([5], Theorem 2.1)**.** *Let Alice possess a random variable M and Bob possess a random variable N, where M and N take values over the same set. Suppose Alice and Bob know an upper bound c on the value of $\mathrm{D}_\infty(p_M\|p_N)$. Then for every $\varepsilon > 0$, there exists a randomness assisted protocol (with uniform shared randomness) in which Bob outputs the random variable $\hat{M}$ such that $\|p_{\hat{M}} - p_M\| \leq \varepsilon$, if the amount of communication from Alice to Bob is $c + \log(\frac{1}{\varepsilon})$.*

An immediate corollary to this is the following, which is a *smooth* version of Lemma 1.

Corollary 4. *Let Alice possess a random variable M and Bob possess a random variable N, where M and N take values over the same set. Fix an $\varepsilon \in (0, \frac{1}{2})$. Suppose Alice and Bob know an upper bound c on the value of $\mathrm{D}_s^\varepsilon(p_M\|p_N)$. Then there exists a randomness-assisted protocol (with uniform shared randomness) where Bob outputs a random variable $\hat{M}$ such that $\|p_{\hat{M}} - p_M\| \leq 5\varepsilon$, if the amount of communication from Alice to Bob is $c + \log(\frac{2}{\varepsilon})$.*

Proof. From the definition of $\mathrm{D}_s^\varepsilon(p_M\|p_N)$, it holds that $\Pr_{m\leftarrow p_M}\{\frac{p_M(m)}{p_N(m)} \geq 2^c\} \leq \varepsilon$. Let us define the following set

$$\mathrm{Good} := \left\{ m : \frac{p_M(m)}{p_N(m)} < 2^c \right\}.$$

Let M' be a random variable taking values over the set Good, formally defined as

$$p_{M'}(m) = \begin{cases} \frac{p_M(m)}{\Pr\{\mathrm{Good}\}} & \text{if } m \in \mathrm{Good}; \\ 0 & \text{otherwise.} \end{cases}$$

It holds that $\|p_{M'} - p_M\| \leq 2\varepsilon$ and $\Pr\{\mathrm{Good}\} \geq 1 - \varepsilon$. This implies

$$2^{\mathrm{D}_\infty(p_{M'}\|p_N)} = \frac{1}{\Pr\{\mathrm{Good}\}} \cdot \max_{m\in\mathrm{Good}} \frac{p_M(m)}{p_N(m)} \leq \frac{2^c}{1-\varepsilon}.$$

Suppose Alice possesses the random variable M' and Bob possesses the random variable N. Alice and Bob execute the protocol used in the proof of Lemma 1. Then Bob outputs the random variable M'' such that $\|p_{M''} - p_{M'}\| \leq \varepsilon$. The same protocol when now executed on the random variable M produces a random variable $\hat{M}$. It holds that

$$\begin{aligned} \|p_{\hat{M}} - p_M\| &\overset{a}{\leq} \|p_{\hat{M}} - p_{M''}\| + \|p_{M''} - p_{M'}\| + \|p_{M'} - p_M\| \\ &\overset{b}{\leq} \|p_M - p_{M'}\| + \|p_{M''} - p_{M'}\| + \|p_{M'} - p_M\| \\ &\leq 2\varepsilon + \varepsilon + 2\varepsilon = 5\varepsilon, \end{aligned}$$

where (a) follows from triangle inequality and (b) follows from Fact 5. The amount of communication from Alice to Bob is at most $c + \log(\frac{1}{\varepsilon}) + \log(\frac{1}{1-\varepsilon}) \leq c + \log(\frac{2}{\varepsilon})$. This proves the corollary. □

Using this, the following achievability result is obtained for Task 1.

Theorem 8 (Braverman and Rao protocol, [5])**.** *Let $\varepsilon \geq 0$. Let R be a natural number such that,*

$$R \geq \min_{(\tilde{X},\tilde{Y},\tilde{M})} \max_{y\in\mathrm{supp}(p_{\tilde{Y}})} \inf_{p_{N|\tilde{Y}=y}} \mathrm{D}_s^\varepsilon(p_{\tilde{X}\tilde{M}|\tilde{Y}=y}\|p_{\tilde{X}|\tilde{Y}=y} \times p_{N|\tilde{Y}=y}) + O\left(\log\frac{1}{\varepsilon}\right),$$

where $(\tilde{X}, \tilde{Y}, \tilde{M})$ satisfies $\tilde{M} - \tilde{X} - \tilde{Y}$ and $\|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{XYM}\| \leq \varepsilon$ and $(\tilde{Y}, N) \sim p_{\tilde{Y}N}$. There exists a shared randomness assisted protocol in which Alice communicates R bits to Bob and Bob outputs random variable $\hat{M}$ satisfying $\|p_{XYM} - p_{XY\hat{M}}\| \leq 8\varepsilon$.

Proof. Let $(\tilde{X}, \tilde{Y}, \tilde{M})$ and $(\tilde{Y}, N)$ be the random variables that achieve the optimization in the definition of R. Define

$$
\begin{aligned}
c &:= \max_{y\in\mathrm{supp}(p_{\tilde{Y}})} \mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}|\tilde{Y}=y}\|p_{\tilde{X}|\tilde{Y}=y}\times p_{N|\tilde{Y}=y}) \\
&= \max_{y\in\mathrm{supp}(p_{\tilde{Y}})} \min\left(a_y : \Pr_{(x,m)\leftarrow p_{\tilde{X}\tilde{M}|\tilde{Y}=y}}\left\{\frac{p_{\tilde{X}\tilde{M}|\tilde{Y}=y}(x,m)}{p_{\tilde{X}|\tilde{Y}=y}(x)\cdot p_{N|\tilde{Y}=y}(m)} > 2^{a_y}\right\}\leq\varepsilon\right) \\
&= \max_{y\in\mathrm{supp}(p_{\tilde{Y}})} \min\left(a_y : \Pr_{(x,m)\leftarrow p_{\tilde{X}\tilde{M}|\tilde{Y}=y}}\left\{\frac{p_{\tilde{M}|\tilde{X}=x,\tilde{Y}=y}(m)}{p_{N|\tilde{Y}=y}(m)} > 2^{a_y}\right\}\leq\varepsilon\right) \\
&= \max_{y\in\mathrm{supp}(p_{\tilde{Y}})} \min\left(a_y : \Pr_{(x,m)\leftarrow p_{\tilde{X}\tilde{M}|\tilde{Y}=y}}\left\{\frac{p_{\tilde{M}|\tilde{X}=x}(m)}{p_{N|\tilde{Y}=y}(m)} > 2^{a_y}\right\}\leq\varepsilon\right) \qquad (5)
\end{aligned}
$$

where the last equality follows because $\tilde{M}-\tilde{X}-\tilde{Y}$. From Equation (5), we conclude that for all $y\in\mathrm{supp}(p_{\tilde{Y}})$,

$$
\Pr_{x\leftarrow p_{\tilde{X}|\tilde{Y}=y}}\left\{\mathrm{D}_s^{\varepsilon}(p_{M|\tilde{X}=x}\|p_{N|\tilde{Y}=y}) > c\right\}\leq\varepsilon \qquad (6)
$$

Fictitious protocol $\mathcal{P}'$: Alice possesses random variables $(\tilde{X},\tilde{M})$ and Bob possess random variable $\tilde{Y}$. They also share uniform shared randomness as required for the protocol in Corollary 4. Alice, upon seeing a realization x of $\tilde{X}$ and Bob, upon seeing a realization y of $\tilde{Y}$, run the protocol as defined in Corollary 4 with $M\leftarrow\tilde{M}\mid(\tilde{X}=x)$, $N\leftarrow N\mid(\tilde{Y}=y)$ and c as defined in Equation 5. At the end of the protocol, let M' be the random variable output by Bob. From Equation (6) and Corollary 4, we conclude that for all $y\in\mathrm{supp}(p_{\tilde{Y}})$,

$$
\Pr_{x\leftarrow p_{\tilde{X}|\tilde{Y}=y}}\left\{\|p_{M'|\tilde{X}=x,\tilde{Y}=y}-p_{\tilde{M}|\tilde{X}=x}\|\geq 5\varepsilon\right\}\leq\varepsilon.
$$

This implies that $\|p_{\tilde{X}\tilde{Y}M'}-p_{\tilde{X}\tilde{Y}\tilde{M}}\|\leq 6\varepsilon$.

Actual protocol $\mathcal{P}$: Alice and Bob possess the random variable (X,Y,M). They run the protocol $\mathcal{P}'$ which outputs the random variable triplet $(X,Y,\hat{M})$. Since $\|p_{\tilde{X}\tilde{Y}\tilde{M}}-p_{XYM}\|\leq\varepsilon$, it holds by triangle inequality and Fact 5 that

$$
\|p_{XY\hat{M}}-p_{XYM}\|\leq\|p_{XY\hat{M}}-p_{\tilde{X}\tilde{Y}M'}\|+\|p_{\tilde{X}\tilde{Y}M'}-p_{\tilde{X}\tilde{Y}\tilde{M}}\|+\|p_{\tilde{X}\tilde{Y}\tilde{M}}-p_{XYM}\|\leq\varepsilon+6\varepsilon+\varepsilon.
$$

The communication cost is $R=c+O\left(\log\frac{1}{\varepsilon}\right)$. This completes the proof. □

We now compare our result (Theorem 1) with Theorem 8. To accomplish this, we first define a series of new quantities and relate them to each other. In what follows, we will use $\mathcal{P}$ to represent a protocol for the Task1 discussed in Section 1.

- **Opt**$^{\varepsilon}$: Let $\mathcal{P}$ 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 $\hat{M}$ correlated with XY. Let $\mathcal{P}(X,Y):=(X,Y,\hat{M})$ represent the output of the protocol. We define $\mathrm{err}(\mathcal{P}):=\|p_{XY\hat{M}}-p_{XYM}\|$ as the error incurred by the protocol and $\mathrm{C}(\mathcal{P})$ as the communication cost of the protocol. Define

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

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

$$
\mathrm{Opt}_1^{\varepsilon}:=\min_{(X,Y,U,S,\hat{M},V)}\mathrm{D}_{\infty}(p_{XSV}\|p_{XS}\times p_U),
$$

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

- **BR**$^{\varepsilon}$: The amount of communication needed by the protocol of Braverman and Rao for Task 1 is denoted by $\mathrm{BR}^{\varepsilon}$ and formally defined below (see also Theorem 8). Let $(\tilde{X},\tilde{Y},\tilde{M})$ be a joint random variable such that $\tilde{Y}-\tilde{X}-\tilde{M}$ and $\|p_{\tilde{X}\tilde{Y}\tilde{M}}-p_{XYM}\|\leq\varepsilon$. Further, let $(\tilde{Y},N)\sim p_{\tilde{Y}N}$. Define

$$
\mathrm{BR}^{\varepsilon}:=\min_{(\tilde{X},\tilde{Y},\tilde{M})}\max_{y\in\mathrm{supp}(p_{\tilde{Y}})}\inf_{p_{N|\tilde{Y}=y}}\mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}|\tilde{Y}=y}\|p_{\tilde{X}|\tilde{Y}=y}\times p_{N|\tilde{Y}=y}).
$$

- **Ext$^\varepsilon$**: This is the quantity obtained in the result of Theorem 1 by setting T as uniform random variable U. Let $(\tilde{X}, \tilde{Y}, \tilde{M}, E)$ be a joint random variable such that $\tilde{Y} - \tilde{X} - (\tilde{M}, E)$ and $\|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{XYM}\| \leq \varepsilon$. Define

$$\mathrm{Ext}^\varepsilon := \min_{(\tilde{X},\tilde{Y},\tilde{M},E)} \left(\mathrm{D}_s^\varepsilon(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_U) - \mathrm{D}_{\mathrm{H}}^\varepsilon(p_{\tilde{Y}\tilde{M}E} \| p_{\tilde{Y}} \times p_U) \right).$$

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 2) along with the protocol of Braverman and Rao (Theorem 8).

Theorem 9. *Let $M - X - Y$. Then it holds that*

1. $\mathrm{Opt}^\varepsilon \geq \mathrm{Opt}_1^\varepsilon$.
2. $\mathrm{Opt}_1^\varepsilon \geq \mathrm{BR}^{3\sqrt{\varepsilon}} - \log(\frac{4}{\varepsilon})$.
3. $\mathrm{BR}^\varepsilon + O(\log(\frac{1}{\varepsilon})) \geq \mathrm{Opt}^{8\varepsilon}$.
4. $\mathrm{Ext}^\varepsilon + O(\log(\frac{1}{\varepsilon})) \geq \mathrm{Opt}^{O(\sqrt{\varepsilon})}$
5. $\mathrm{BR}^\varepsilon > \mathrm{Ext}^\varepsilon$.

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

1. In any one-way communication protocol $\mathcal{P}$ with a shared randomness S, Alice produces a message $V \in \mathcal{V}$ 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 $\hat{M}$ such that $\|p_{XY\hat{M}} - p_{XYM}\| \leq \varepsilon$ and $X - (Y, V, S) - \hat{M}$. The total number of bits communicated by Alice to Bob is $\mathrm{C}(\mathcal{P}) = \log|\mathcal{V}|$. The inequality now follows from the relation $\mathrm{D}_\infty(p_{XSV} \| p_{XS} \times p_U) \leq \log|\mathcal{V}|$ and the definition of $\mathrm{Opt}_1^\varepsilon$.
2. Let $\hat{M}$ be the output of the protocol $\mathcal{P}$ such that $\|p_{XY\hat{M}} - p_{XYM}\| \leq \varepsilon$. Define the following set

$$\text{Good} := \left\{ y : \|p_{X\hat{M}|Y=y} - p_{XM|Y=y}\| \leq \sqrt{\varepsilon} \right\}. \tag{7}$$

Using the fact that $\|p_{XY\hat{M}} - p_{XYM}\| \leq \varepsilon$ and Markov's inequality we have that $\Pr\{\text{Good}\} \geq 1 - \sqrt{\varepsilon}$.

A closeby Markov chain distribution: We now construct a random variable triplet (X_1, Y_1, M_1) distributed as follows:

$$p_{X_1Y_1M_1}(x, y, m) := \begin{cases} \frac{p_Y(y)}{\Pr\{\text{Good}\}} p_{X\hat{M}|Y=y}(x, m) & \text{if } y \in \text{Good}; \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Further, define the random variable triplet (X_2, Y_2, M_2) distributed as follows

$$p_{X_2Y_2M_2}(x, y, m) := p_{X_1Y_1}(x, y) p_{M|X=x}(m). \tag{9}$$

We observe that $M_2 - X_2 - Y_2$ holds. Moreover, we have the following:

$$\begin{aligned} \|p_{X_2Y_2M_2} - p_{XYM}\| &\leq \|p_{X_2Y_2M_2} - p_{X_1Y_1M_1}\| + \|p_{X_1Y_1M_1} - p_{XYM}\| \\ &\overset{a}{\leq} \|p_{X_2Y_2M_2} - p_{X_1Y_1M_1}\| + 2\sqrt{\varepsilon} \\ &\overset{b}{\leq} \sqrt{\varepsilon} + 2\sqrt{\varepsilon} = 3\sqrt{\varepsilon}, \end{aligned} \tag{10}$$

where (a) follows from $\|p_{X_1Y_1M_1} - p_{XYM}\| \leq 2\left(1 - \Pr\{\text{Good}\}\right) \leq 2\sqrt{\varepsilon}$; and (b) follows from the following set of inequalities:

$$\begin{aligned} \|p_{X_2Y_2M_2} - p_{X_1Y_1M_1}\| &\overset{a}{=} \sum_y p_{Y_1}(y) \|p_{X_2M_2|Y_2=y} - p_{X_1M_1|Y_1=y}\| \\ &\overset{b}{=} \sum_y p_{Y_1}(y) \|p_{XM|Y=y} - p_{X\hat{M}|Y_1=y}\| \\ &\overset{c}{\leq} \sqrt{\varepsilon}, \end{aligned}$$

where (a) follows from Definition (9); (b) follows from Definitions (8) and (9); (c) follows from the definition of the set Good (Definition (7)).

Lower bound: For the random variables (X, Y, V, S, U) as defined in $\mathrm{Opt}_1^{\varepsilon}$, we prove the following:

$$\begin{aligned}
\mathrm{D}_\infty(p_{XSV}\|p_X \times p_S \times p_U) &\overset{a}{=} \mathrm{D}_\infty(p_{YXSV}\|p_{YX} \times p_S \times p_U) \\
&= \max_{y\in\mathrm{supp}(p_Y)} \mathrm{D}_\infty(p_{XSV|Y=y}\|p_{X|Y=y} \times p_{S|Y=y} \times p_U) \\
&\overset{b}{\geq} \max_{y\in\mathrm{supp}(p_Y)} \min_{S'V'} \mathrm{D}_\infty(p_{XSV|Y=y}\|p_{X|Y=y} \times p_{S'V'}) \\
&\overset{c}{\geq} \max_{y\in\mathrm{supp}(p_Y)} \min_{p_{N|Y=y}} \mathrm{D}_\infty(p_{X\hat{M}|Y=y}\|p_{X|Y=y} \times p_{N|Y=y}) \\
&\overset{d}{\geq} \max_{y\in\mathrm{supp}(p_{Y_1})} \min_{p_{N|Y=y}} \mathrm{D}_\infty(p_{X\hat{M}|Y=y}\|p_{X|Y=y} \times p_{N|Y=y}) \\
&\overset{e}{\geq} \max_{y\in\mathrm{supp}(p_{Y_1})} \min_{p_{N|Y=y}} \mathrm{D}_\infty^{\sqrt{\varepsilon}}(p_{X_2M_2|Y=y}\|p_{X|Y=y} \times p_{N|Y=y}) \\
&\overset{f}{=} \max_{y\in\mathrm{supp}(p_{Y_2})} \min_{p_{N|Y=y}} \mathrm{D}_\infty^{\sqrt{\varepsilon}}(p_{X_2M_2|Y=y}\|p_{X_2|Y=y} \times p_{N|Y=y}) \\
&\overset{g}{\geq} \max_{y\in\mathrm{supp}(p_{Y_2})} \min_{p_{N|Y=y}} \mathrm{D}_s^{3\sqrt{\varepsilon}}(p_{X_2M_2|Y=y}\|p_{X_2|Y=y} \times p_{N|Y=y}) - \log\left(\frac{4}{\varepsilon}\right).
\end{aligned}$$

Above, (a) follows from the fact that $Y - X - (S, V)$; (b) follows by minimizing over all random variables (S', V'); (c) follows from Fact 5 ; (d) follows from the fact that $\mathrm{supp}(p_{Y_1}) \subseteq \mathrm{supp}(p_Y)$; (e) follows from the definition of smooth max divergence and the fact that for all $y \in \mathrm{supp}(p_{Y_1}) = \mathrm{Good}$, we have:

$$\begin{aligned}
\|p_{X_2M_2|Y=y} - p_{X\hat{M}|Y=y}\| &= \sum_x p_{X|Y=y}(x)\|p_{M_2|X=x} - p_{\hat{M}|X=x,Y=y}\| \\
&= \sum_x p_{X|Y=y}(x)\|p_{M|X=x} - p_{\hat{M}|X=x,Y=y}\| \\
&= \|p_{XM|Y=y} - p_{X\hat{M}|Y=y}\| \leq \sqrt{\varepsilon};
\end{aligned}$$

(f) follows from the fact that $\mathrm{supp}(p_{Y_2}) = \mathrm{Good}$ and for all $y \in \mathrm{Good}$, $p_{X_2|Y_2=y} = p_{X|Y=y}$; and (g) follows from Fact 1.

Thus,

$$\max_{y\in\mathrm{supp}(p_{Y_2})} \min_{p_{N|Y=y}} \mathrm{D}_s^{3\sqrt{\varepsilon}}(p_{X_2M_2|Y=y}\|p_{X_2|Y=y} \times p_{N|Y=y}) \geq \mathrm{BR}^{3\sqrt{\varepsilon}},$$

where the inequality above follows because $M_2 - X_2 - Y_2$ and $\|p_{X_2Y_2M_2} - p_{XYM}\| \leq 3\sqrt{\varepsilon}$ (Equation (10)) and from the definition of $\mathrm{BR}^{3\sqrt{\varepsilon}}$. This proves the item.

3. This is a direct consequence of Theorem 8.
4. This is a direct consequence of Theorem 1 .
5. Let $(\tilde{X}, \tilde{Y}, \tilde{M})$ and $(\tilde{Y}, N)$ be as obtained from the definition of $\mathrm{BR}^{\varepsilon}$. From Theorem 10 below, it holds that there exists a random variable E such that $(\tilde{X}, \tilde{Y}, \tilde{M}, E)$ satisfies $\tilde{Y} - \tilde{X} - (\tilde{M}, E)$ and

$$\max_y \mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}|\tilde{Y}=y}\|p_{\tilde{X}|\tilde{Y}=y} \times p_{N|\tilde{Y}=y}) \geq \mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}E}\|p_{\tilde{X}} \times p_U) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{\tilde{Y}\tilde{M}E}\|p_{\tilde{Y}} \times p_U).$$

The item follows by observing that $\mathrm{Ext}^{\varepsilon}$ is obtained by minimizing right hand side over all $(\tilde{X}, \tilde{Y}, \tilde{M}, E)$ and U, such that $\tilde{Y} - \tilde{X} - (\tilde{M}, E)$ and $\|p_{\tilde{X}\tilde{Y}\tilde{M}} - p_{XYM}\| \leq \varepsilon$.

□

The following theorem shows that the information theoretic quantity obtained in Theorem 1 is upper bounded by the information theoretic quantity obtained in Theorem 8.

Theorem 10. *Let $(\tilde{X}, \tilde{Y}, \tilde{M})$ and $(\tilde{Y}, N)$ be the optimal random variables appearing in the definition of $\mathrm{BR}^{\varepsilon}$. Then there exists a random variable E such that $\tilde{Y} - \tilde{X} - (\tilde{M}, E)$ and*

$$\max_y \mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}|\tilde{Y}=y} \| p_{\tilde{X}|\tilde{Y}=y} \times p_{N|\tilde{Y}=y}) \geq \mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_U) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{\tilde{Y}\tilde{M}E} \| p_{\tilde{Y}} \times p_U),$$

where U is uniformly distributed over the set over which the random variable pair $(\tilde{M}, E)$ take values.

Proof. The proof is divided in the following steps.

Construction of appropriate extension: Let K be the smallest integer such that $K p_{\tilde{M}|\tilde{X}=x}(m)$ 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 $(\tilde{X}, \tilde{M})$ as follows: for every $(m, e, x) \in \mathcal{M} \times \mathcal{K} \times \mathcal{X}$,

$$p_{\tilde{X}\tilde{M}E}(x, m, e) := \begin{cases} \frac{p_{\tilde{X}}(x)}{K} & \text{if } e \leq K p_{\tilde{M}|\tilde{X}=x}(m), \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

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

$$\begin{aligned} \mathrm{D}_s(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_U) &\overset{a}{=} \max_{m,x,e} \log \frac{p_{\tilde{X}\tilde{M}E}(x, m, e)}{p_{\tilde{X}}(x) p_U(u)} \\ &\overset{b}{=} \log \frac{|\mathcal{M}|K}{K} \\ &= \log |\mathcal{M}|, \end{aligned} \tag{12}$$

where (a) follows from the definition of $\mathrm{D}_s(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_U)$; (b) follows from Equation (11) and the fact that U is uniform over the set $\mathcal{M} \times \mathcal{K}$.

Lower bounding hypothesis testing relative entropy: For brevity, let

$$D_{\infty}^* := \max_y \mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}|\tilde{Y}=y} \| p_{\tilde{X}|\tilde{Y}=y} \times p_{N|\tilde{Y}=y}).$$

Define the following set

$$\mathcal{A} := \left\{ (y, m, e) \in \mathcal{Y} \times \mathcal{M} \times \mathcal{K} : e \leq K 2^{D_{\infty}^*} p_{N|\tilde{Y}=y}(m) \right\}. \tag{13}$$

We will prove the following

$$\Pr_{p_{\tilde{Y}} \times p_U} \{\mathcal{A}\} = 2^{-(\log |\mathcal{M}| - D_{\infty}^*))}; \tag{14}$$

$$\Pr_{p_{\tilde{M}\tilde{Y}E}} \{\mathcal{A}\} \geq 1 - \varepsilon. \tag{15}$$

The theorem now follows from the definition of $\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{\tilde{Y}\tilde{M}E} \| p_{\tilde{Y}} \times p_U)$ and Equations (12),(14),(15) as follows:

$$\begin{aligned} \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{\tilde{Y}\tilde{M}E} \| p_{\tilde{Y}} \times p_U) &\geq \log |\mathcal{M}| - D_{\infty}^* \\ &= \mathrm{D}_s(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_U) - D_{\infty}^* \\ &\geq \mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_U) - D_{\infty}^*, \end{aligned}$$

which leads to

$$D_{\infty}^* \geq \mathrm{D}_s^{\varepsilon}(p_{\tilde{X}\tilde{M}E} \| p_{\tilde{X}} \times p_U) - \mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{\tilde{Y}\tilde{M}E} \| p_{\tilde{Y}} \times p_U).$$

Proof of Equation (14): Towards this notice the following

$$\begin{aligned}
\Pr_{p_{\tilde{Y}} \times p_U}\{\mathcal{A}\} &= \sum_{(y,m,e)\in\mathcal{A}} p_{\tilde{Y}}(y) p_U(m,e) \\
&= \sum_{y\in\mathcal{Y}} p_{\tilde{Y}}(y) \sum_{(m,e):(y,m,e)\in\mathcal{A}} \frac{1}{|\mathcal{M}|K} \\
&= \sum_{(y,m)\in\mathcal{Y}\times\mathcal{M}} p_{\tilde{Y}}(y) p_{N|\tilde{Y}=y}(m) \frac{K 2^{D^*_\infty}}{|\mathcal{M}|K} \\
&= \frac{2^{D^*_\infty}}{|\mathcal{M}|} \\
&= 2^{-(\log|\mathcal{M}| - D^*_\infty)}.
\end{aligned}$$

Proof of Equation (15). Towards this we have the following:

$$\begin{aligned}
\Pr_{p_{\tilde{Y}\tilde{M}E}}\{A\} &= \sum_x p_{\tilde{X}}(x) \sum_{(y,m,e)\in\mathcal{A}} p_{\tilde{Y}|\tilde{X}=x}(y) p_{\tilde{M}E|\tilde{X}=x}(m,e) \\
&\overset{a}{=} \sum_x p_{\tilde{X}}(x) \sum_y p_{\tilde{Y}|\tilde{X}=x}(y) \sum_m \sum_{\substack{e: e\le K p_{\tilde{M}|\tilde{X}=x}(m) \\ (y,m,e)\in\mathcal{A}}} \frac{1}{K} \\
&\overset{b}{\ge} \sum_{(x,y)} p_{\tilde{X}\tilde{Y}}(x,y) \sum_{m: p_{\tilde{M}|\tilde{X}=x}(m) \le 2^{D^*_\infty} p_{N|\tilde{Y}=y}(m)} p_{\tilde{M}|\tilde{X}=x}(m) \\
&\overset{c}{\ge} 1-\varepsilon,
\end{aligned}$$

where a follows from Definition (11), b follows because for every x

$$\left\{(y,m,e) : p_{\tilde{M}|\tilde{X}=x}(m) \le 2^{D^*_\infty} p_{N|\tilde{Y}=y}(m) \text{ and } e \le K p_{\tilde{M}|\tilde{X}=x}(m)\right\} \subseteq \mathcal{A},$$

and c follows from the definition of D^*_∞. This completes the proof. □

Acknowledgment

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 message compression with applications." arXiv:1410.3031, 2014.

[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 Transcations 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] 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.

[7] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, July 1948.

[8] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, Jul 1973.

[9] A. Wyner, "On source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 21, pp. 294–300, May 1975.

[10] 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.

[11] 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.

[12] 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.

[13] T. Uteymatsu and T. Matstuta, "Revisiting the Slepain-Wolf coding problem for general sources: A direct approach," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Honolulu, HI), June 2014.

[14] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Trans. Fundamentals*, vol. E78-A(9), pp. 1063–1070, Sept. 1995.

[15] T. Uteymatsu and T. Matstuta, "Source coding with side information at the decoder revisited," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Hong Kong), June 2015.

[16] 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.

[17] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley Series in Telecommunications, New York, NY, USA: John Wiley & Sons, 1991.

[18] A. Anshu, R. Jain, and N. A. Warsi, "A generalized quantum Slepian-Wolf." https://arxiv.org/abs/arXiv:1703.09961, 2017.

[19] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge University Press, 2012.

A Deferred proofs

Proof of Fact 6: Let $c := \mathrm{D}_s^{\varepsilon}(p_{XM}\|p_X \times p_W)$. Define,

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

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

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

We note that,

$$\forall (x,m) : \frac{p_{X'M'}(x,m)}{p_X(x)p_W(m)} \leq \frac{2^c}{p} \quad \text{and} \quad p_{X'}(x) \leq \frac{p_X(x)}{p}. \tag{16}$$

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

$$\|p_{X'M'_1\ldots M'_{2^R}} - p_{XM_1\ldots M_{2^R}}\| = \|p_{XM} - p_{X'M'}\| \leq 4\varepsilon. \tag{17}$$

Consider,

$$\begin{aligned}
&\mathrm{D}\left(p_{X'M'_1\ldots M'_{2^R}}\|p_X\times p_{W_1}\times\ldots\times p_{W_{2^R}}\right)\\
&\overset{a}{=}\frac{1}{2^R}\sum_{j=1}^{2^R}\left(\mathrm{D}\left(p_{X'M'_j}\|p_X\times p_{W_j}\right)-\mathrm{D}\left(p_{X'M'_j}\times p_{W_1}\times\ldots\times p_{W_{j-1}}\times p_{W_{j+1}}\times\ldots\times p_{W_{2^R}}\|p_{X'M'_1\ldots M'_{2^R}}\right)\right)\\
&\overset{b}{\leq}\frac{1}{2^R}\sum_{j=1}^{2^R}\left(\mathrm{D}\left(p_{X'M'_j}\|p_X\times p_{W_j}\right)-\mathrm{D}\left(p_{X'M'_j}\|\frac{1}{2^R}p_{X'M'_j}+\left(1-\frac{1}{2^R}\right)p_{X'}\times p_{W_j}\right)\right)\\
&\overset{c}{\leq}\log\left(1+\frac{2^c}{2^R}\right)+\log\frac{1}{p}\\
&\overset{d}{\leq}4\varepsilon,
\end{aligned}$$

where (a) follows from Fact 2; (b) follows from Fact 3 and (c) follows from Equation (16) and (d) follows since $\log(1+x)\leq x$ for all real x and from choice of R. From Fact 4 we get,

$$\|p_{X'M'_1\ldots M'_{2^R}}-p_X\times p_{W_1}\times\ldots p_{W_{2^R}}\|\leq 2\sqrt{\varepsilon}.$$

This along with Equation (17) and the triangle inequality for ℓ_1 distance gives us the desired. □

Proof of Lemma 7: Define,

$$\begin{aligned}
c_1&:=\mathrm{D}_s^{\varepsilon}(p_{XMN}\|p_X\times p_U\times p_V),c_2:=\mathrm{D}_s^{\varepsilon}(p_{XM}\|p_X\times p_U),c_3:=\mathrm{D}_s^{\varepsilon}(p_{XN}\|p_X\times p_V),\\
\mathrm{Good}_1&:=\left\{(x,m,n):\frac{p_{XMN}(x,m,n)}{p_X(x)p_U(m)p_V(n)}\leq 2^{c_1}\right\},\\
\mathrm{Good}_2&:=\left\{(x,m,n):\frac{p_{XM}(x,m)}{p_X(x)p_U(m)}\leq 2^{c_2}\right\},\\
\mathrm{Good}_3&:=\left\{(x,m,n):\frac{p_{XN}(x,n)}{p_X(x)p_V(n)}\leq 2^{c_3}\right\},\\
\mathrm{Good}&:=\mathrm{Good}_1\cap\mathrm{Good}_2\cap\mathrm{Good}_3.
\end{aligned}$$

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

$$p_{X'M'N'}(x,m,n)=\begin{cases}\frac{p_{XMN}(x,m,n)}{p} & \text{if } (x,m,n)\in\mathrm{Good},\\ 0 & \text{otherwise.}\end{cases}$$

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

$$\frac{p_{X'M'N'}(x,m,n)}{p_X(x)p_U(m)p_V(n)}\leq\frac{2^{c_1}}{p}\quad,\quad\frac{p_{X'M'}(x,m)}{p_X(x)p_U(m)}\leq\frac{2^{c_2}}{p}\quad,\quad\frac{p_{X'N'}(x,n)}{p_X(x)p_V(n)}\leq\frac{2^{c_3}}{p}\quad,\quad p_{X'}(x)\leq\frac{p_X(x)}{p}.\tag{18}$$

Let us construct joint random variables $\left(J',K',X',M'_1,\ldots,M'_{2^{R_1}},N'_1,\ldots,N'_{2^{R_2}}\right)$ 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,

$$\|p_{X'M'_1\ldots M'_{2^{R_1}}N'_1\ldots N'_{2^{R_2}}}-p_{XM_1\ldots M_{2^{R_1}}N_1\ldots N_{2^{R_2}}}\|=\|p_{XMN}-p_{X'M'N'}\|\leq 12\varepsilon.\tag{19}$$

For notational convenience lets define,

$$\begin{aligned}
\forall j\in[2^{R_1}]:\quad p_{U_{-j}}&:=p_{U_1}\times\ldots\times p_{U_{j-1}}\times p_{U_{j+1}}\times\ldots\times p_{U_{2^{R_1}}},\\
\forall k\in[2^{R_2}]:\quad p_{V_{-k}}&:=p_{V_1}\times\ldots\times p_{V_{k-1}}\times p_{V_{k+1}}\times\ldots\times p_{V_{2^{R_2}}},\\
q_{X'M'N'}&:=\frac{1}{2^{R_1+R_2}}p_{X'M'_jN'_k}+\frac{1}{2^{R_1}}\left(1-\frac{1}{2^{R_2}}\right)p_{X'M'_j}\times p_{V_k}\\
&\quad+\frac{1}{2^{R_2}}\left(1-\frac{1}{2^{R_1}}\right)p_{X'N'_k}\times p_{U_j}+\left(1-\frac{2^{R_1}+2^{R_2}-1}{2^{R_1+R_2}}\right)p_{X'}\times p_{U_j}\times p_{V_k}.
\end{aligned}$$

Consider,

$$\begin{aligned}
&\mathrm{D}\left(p_{X'M'_1\ldots M'_{2^{R_1}}N'_1\ldots N'_{2^{R_2}}}\|p_X\times p_{U_1}\times\ldots\times p_{U_{2^{R_1}}}\times p_{V_1}\times\ldots\times p_{V_{2^{R_2}}}\right)\\
&\overset{a}{=}\frac{1}{2^{R_1+R_2}}\sum_{j,k}\left(\mathrm{D}\left(p_{X'M'_jN'_k}\|p_X\times p_{U_j}\times p_{V_k}\right)-\mathrm{D}\left(p_{X'M'_jN'_k}\times p_{U_{-j}}\times p_{V_{-k}}\|p_{X'M'_1\ldots M'_{2^{R_1}}N'_1\ldots N'_{2^{R_2}}}\right)\right)\\
&\overset{b}{\leq}\frac{1}{2^{R_1+R_2}}\sum_{j,l}\left(\mathrm{D}\left(p_{X'M'_jN'_k}\|p_X\times p_{U_j}\times p_{V_k}\right)-\mathrm{D}\left(p_{X'M'_jN'_k}\|q_{X'M'N'}\right)\right)\\
&\overset{c}{\leq}\log\left(1+\frac{2^{c_1}}{2^{R_1+R_2}}+\frac{2^{c_2}}{2^{R_1}}+\frac{2^{c_3}}{2^{R_2}}\right)+\log\frac{1}{p}\\
&\overset{d}{\leq}9\varepsilon,
\end{aligned}$$

where (a) follows from Fact 2; (b) follows from Fact 3; (c) follows from Equation (18) and (d) follows since $\log(1+x)\leq x$ for all real x and from choice of parameters. From Fact 4 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 (19) and the triangle inequality for ℓ_1 distance gives us the desired. □

Proof of Fact 8: Let $\mathcal{A}\subseteq\mathcal{Y}\times\mathcal{M}$ be such that $\Pr_{p_{YM}}\{\mathcal{A}\}\geq 1-\varepsilon$, and

$$c:=\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{YM}\|p_Y\times p_W)=-\log\Pr_{p_Y\times p_W}\{\mathcal{A}\}.$$

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,

$$\begin{aligned}
\Pr\{J'\neq j\}&\leq\Pr\{(Y,M_j)\notin\mathcal{A}\}+\Pr\{(Y,M_{j'})\in\mathcal{A}\text{ for some }j'\neq j\}\\
&\leq\varepsilon+2^R\cdot 2^{-c}\leq 2\varepsilon.
\end{aligned}$$

Therefore,

$$\Pr\{J\neq J'\}=\sum_{j\in[2^R]}\Pr\{J=j\}\cdot\Pr\{J'\neq j\mid J=j\}\leq 2\varepsilon.$$

□

Proof of Fact 9: Let $\mathcal{A}\subseteq\mathcal{M}\times\mathcal{N}$ be such that $\Pr_{p_{MN}}\{\mathcal{A}\}\geq 1-\varepsilon$, and

$$c:=\mathrm{D}_{\mathrm{H}}^{\varepsilon}(p_{MN}\|p_M\times p_N)=-\log\Pr_{p_M\times p_N}\{\mathcal{A}\}.$$

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,

$$\begin{aligned}
\Pr\{(J',K')\neq(j,k)\}&\leq\Pr\{(M_j,N_k)\notin\mathcal{A}\}+\Pr\left\{(M_{\tilde{j}},N_{\tilde{k}})\in\mathcal{A}\text{ for some }(\tilde{j},\tilde{k})\neq(j,k)\right\}\\
&\leq\varepsilon+2^{R_1+R_2}\cdot 2^{-c}\leq 2\varepsilon.
\end{aligned}$$

Therefore,

$$\Pr\{(J,K)\neq(J',K')\}=\sum_{(j,k)\in[2^{R_1}]\times[2^{R_2}]}\Pr\{(J,K)=(j,k)\}\cdot\Pr\{(J',K')\neq(j,k)\mid(J,K)=(j,k)\}\leq 2\varepsilon.$$

□