On the Compression of Messages in the Multi-Party Setting

Anurag Anshu and Penghui Yao

Abstract—We consider the following communication task in the multi-party setting, which involves joint random variables $XYZMN$ with the property that $M$ is independent of $YZN$ conditioned on $X$, and $N$ is independent of $XZN$ conditioned on $Y$. Three parties Alice, Bob and Charlie, respectively, observe samples $x$, $y$ and $z$ from $XYZ$. Alice and Bob communicate messages to Charlie with the goal that Charlie can output a sample $(m, n)$ such that the distribution of $(x, y, z, m, n)$ is close to $XYZMN$. This task reflects the simultaneous message passing communication complexity. Furthermore, it is a generalization of some well studied problems in information theory, such as distributed source coding, source coding with a helper and one sender and one receiver message compression. It is also closely related to the lossy distributed source coding task. Our main result is an achievable communication region for this task in the one-shot setting, through which we obtain a nearly optimal characterization using auxiliary random variables of bounded size. We employ our achievability result to provide a nearly optimal one-shot communication region for the task of lossy distributed source coding, in terms of auxiliary random variables of bounded size. Finally, we show that interactions are necessary to achieve the optimal expected communication cost.

Index Terms—Message compression, communication complexity, information theory.

I. INTRODUCTION

SOURCE coding is a central task in information theory, where the task for a sender is to communicate a sample from a source. The constraint is that the error made by the receiver in the decoding process should be small and the goal is to communicate as less number of bits as possible. A tight characterization of this task was achieved by Shannon [1] in the asymptotic & i.i.d. setting, where the senders are assumed to have a large number of identical and independent samples from the source. Later, Slepian and Wolf [2] presented a tight characterization for a multi-party source coding in the same asymptotic & i.i.d. setting. The powerful techniques introduced by these authors were further generalized to asymptotic & non-i.i.d. setting [3].

In the recent years, there has been a growing interest in the study of various generalizations of source coding in non-asymptotic and non-i.i.d. settings. An important setting that has been actively investigated in the past few decades is the one-shot setting, where just one sample from the source is given to the senders. A notable example of source coding, that has been studied in both the asymptotic & i.i.d. and the one-shot settings, is that the sender observes a sample $x$ from a source and the receiver is supposed to output a sample from some random variable that depends on $x$. This task was investigated in the one-shot setting [4], [5] in the context of communication complexity (known as message compression), while in the asymptotic & i.i.d setting [6], [7] in the context of channel simulation. The task can be stated in more details as follows, where Alice is a sender and Charlie is a receiver.

Task A: Let $XM$ be joint random variables taking values over $X \times M$, where $X$ and $M$ are both finite sets. Alice and Charlie possess pre-shared randomness, which is independent of $X$. Alice observes a sample $x$ from $X$ and sends a message to Charlie, where the message depends on the input $x$ and the value observed from the pre-shared randomness. Charlie decodes the message he receives with the value of the pre-shared randomness and outputs a sample distributed according to a random variable $M'$, which satisfies $\frac{1}{2}\|XM - XM'\|_1 \leq \epsilon$.

Above, $\|\cdot\|_1$ is the $\ell_1$ distance and $\epsilon$ is an error parameter. It was shown in [5] that the expected communication cost of this task is equal to $I(X : M)$ (up to a additive factor) in the one-shot setting, generalizing the result of Huffman [8]. The work [5] also gave important applications to the direct sum results in two-party communication complexity. The work [9] considered an extension of Task A with side information about $X$ at Charlie.

Task B: Let $XMZ$ be joint random variables taking values over a set $X \times M \times Z$, such that $M = X - Z$. Alice and Charlie possess pre-shared randomness independent of $XZ$.
Alice observes a sample $x$ from $X$ and Charlie observes a sample $z$ from $Z$. Alice sends a message to Charlie, where the message only depends on the input $x$ and the pre-shared randomness. Charlie outputs a sample distributed according to a random variable $M'$ such that $\frac{1}{2}\|XMZ - XM'Z\|_1 \leq \epsilon$.

An important assumption in Task B is the Markov chain condition $M - X - Z$, which signifies the fact that $M$ is to be treated as a ‘message’ generated by Alice given $x$. Essentially the same condition arises when the side information $Z$ is available with receiver in the context of channel simulation, as the channel generates $M$ only depending on $X$. The authors in [9] present a protocol for Task B with the expected communication cost $\mathbb{E}(|X : M|Z)$ (the conditional mutual information) up to additive factors, in the one-shot setting. Recently, it has been shown that the protocol in [9] is nearly optimal also in terms of the worst communication cost [10].

In this work, we consider a generalization of Task B in the setting of two senders and one receiver. More precisely, we consider the following task, which was also studied in [10].

**Task C:** Let $XYZM|N$ be joint random variables taking values over a set $X \times Y \times Z \times M \times N$, and satisfying the Markov chain conditions $M - X - YZM$ and $Z - X - M - N$.

Alice and Charlie possess pre-shared randomness and Bob and Charlie independently possess another pre-shared randomness. Alice observes a sample $x$ from $X$, Bob observes a sample $y$ from $Y$ and Charlie observes a sample $z$ from $Z$. Alice and Bob, respectively, communicate a message to Charlie (which also depends on the pre-shared randomness). Charlie outputs a sample distributed according to random variables $M'|N'$ such that $\frac{1}{2}\|XYZM|N - XYZM'|N'|\|_1 \leq \epsilon$.

Task C is a generalization of the distributed source coding (DSC) studied by Slepian and Wolf [2] in the asymptotic & i.i.d setting and in the second order and one-shot settings in [10]–[13]. We show in Appendix C that Task C also generalizes the task of source coding with a helper (SCH), which has been studied in [10], [13]–[18]. A motivation for Task C is to achieve a message compression algorithm in the multi-party communication complexity. In the past two decades, many elegant message compression protocols in one-shot settings have been discovered in the context of communication complexity [4], [5], [9], [19] (some of which were discussed earlier). These protocols show how to achieve the communication cost of a task close to the information complexity [20], which measures the amount of information exchanged between the communicating parties. As a result, significant progress has been made towards the direct sum problems, one of the central open problems in communication complexity. However, the notion of information complexity in the model of multi-party communication complexity has not yet been established, partially due to the fact that the communication region of multi-party communication is more involved and less understood. Hence, giving a tight characterization of the communication cost of Task C is a first step towards developing a correct notion of information complexity in the multi-party communication complexity.

We begin with a rate region for Task C in the asymptotic & i.i.d. setting. By employing the time sharing technique [21, Section 4.4], it can be found that the following is an achievable rate region, where $R_1$ is the rate of the communication from Alice to Charlie and $R_2$ is the rate of the communication from Bob to Charlie.

$$\begin{align*}
R_1 &\geq I(X : M|NZ) \\
R_2 &\geq I(Y : N|MZ)
\end{align*}$$

$$R_1 + R_2 \geq I(XY : MN|Z).$$

(1)

Is it possible to show that this rate region is optimal? The answer is negative, which can be seen by considering the task of SCH, which is a special case of Task C as discussed earlier. In this task, Alice holds a random variable $X$ and Bob holds a random variable $Y$ correlated with $X$. Alice and Bob communicate messages to Charlie in a manner that Charlie is able to output $X$ with high probability. It is well known [21, Section 10.4] that the time sharing rate region for the SCH task (as obtained by setting $M = X$ and $N$ trivial in Eq. (1)) is not the optimal rate region. In fact, the known characterization of an optimal rate region requires auxiliary random variables as well. Thus, the rate region given in Eq. (1) is not an optimal characterization of Task C and an optimal characterization may require some auxiliary random variables. On the other hand, the utility of the achievable rate region in Eq. (1) is that it only involves the random variables occurring in the inputs of the task.

### A. Our Results

We obtain the following results in the paper.

- **We study Task C in the one-shot setting.** First, we show how to obtain a one-shot analogue of Eq. (1) in Theorem III.2, which is our main result. Observe that the time sharing method cannot be applied in the one-shot setting. Hence we require a new tool to obtain our result, which we achieve by an appropriate multi-partite generalization of the protocols constructed in [9], [10]. While this achievable communication region is not known to be optimal (which is not known even in the asymptotic & i.i.d. setting, as discussed earlier) it has the following applications.

1. **We obtain a nearly tight characterization of the one-shot lossy distributed source coding** (in the presence of side information at the receiver) in Theorem III.4. When side information at the receiver is absent, our communication region can be viewed as a one-shot analogue of the Berger-Tung bound [22], [23]. While our one-shot region is nearly optimal, it does not imply optimal characterization in the asymptotic & i.i.d. setting, due to the possibility of obtaining a multi-letter outer bound.

2. **We also obtain a nearly tight characterization of Task C with auxiliary random variables** (of bounded size that is comparable to the size of input random variables) in the one-shot setting, in Theorem III.8.

3. **In Section III-B we recover the nearly optimal one-shot results on the DSC task and one-sender one-receiver message compression task** as obtained in [10].

- **We study the expected communication cost of Slepian-Wolf task, a special case of Task C, where $Y, N$ are...**
trivial and $M = X$. This two-party task (as Bob is not involved) was considered by Slepian and Wolf [2] and the communication rate in the asymptotic & i.i.d. setting was shown to be equal to $H(X|Z)$. We show in Theorem V.1 that any one-way protocol for this task incurs an expected communication cost of $\frac{1}{\epsilon} H(X|Z)$. On the other hand, the result of [9] implies that there is an interactive protocol achieving the expected communication cost of $H(X|Z) + \epsilon \left( \sqrt{H(X|Z)} + \log \frac{1}{\epsilon} \right)$, for some universal constant $c$. Thus, there is a stark contrast between the one-way protocols and interactive protocols in terms of the expected communication cost. This also exhibits how the side information $Z$ limits the performance of one-way protocols, since Huffman’s coding scheme [8] achieves the expected communication of $H(X)+1$, which is nearly optimal in the absence of $Z$.

In turn, this implies that one-way protocols cannot achieve the region given in Eq. (1) for Task C in expected communication, even when the side information $Z$ is trivial. On the other hand, we also observe that there is a simple interactive protocol that uses the protocol of [9] for Task B as a subroutine and achieves the region given in Eq. (1) (up to small additive factors).

C. Organisation

We discuss the notations used in our proofs in Section II (the facts and lemmas used are states in Appendix A). In Section III, we discuss our main result and its consequences. We also compare with the communication region given in [10], in this section. In Section IV, we construct our protocol for Task C which leads to the proof of the main result. We also provide an overview of earlier techniques that are relevant to us. In Section V, we discuss the expected communication cost of Task C.

II. Preliminaries

For a natural number $n$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$ and $[m : n]$ denote the set $\{m, m+1, \ldots, n\}$. Let $X$ be a random variable taking values in a finite set $X$ (all the sets considered in this paper are finite). We let $p_X$ represent the distribution of $X$, that is for each $x \in X$, $p_X(x) \overset{def}{=} \Pr[X = x]$. We say that $x$ is drawn from $X$, denoted by $x \sim X$, if $x$ is a sample from the distribution $p_X$. We use $x \sim_{XY} X$ to represent that $x$ is uniformly sampled from set $X$. The support of the random variable $X$ is the set $\{x : p_X(x) > 0\}$ and is denoted by $\text{supp}(X)$. Let random variables $XY$ take values in the set $X \times Y$. The random variable $X$ conditioned on $Y = y$ is denoted as $(X|Y = y)$. We say that $X$ and $Y$ are independent and denote the joint random variables by $X \times Y$, if for each $x \in X$ and $y \in Y$ it holds that $p_{XY}(x,y) = p_X(x)p_Y(y)$. We say that the joint random variables $XYZ$ form a Markov chain, represented as $X \rightarrow Y \rightarrow Z$, if for each $x \in X$, $(Y|X = x)$ and $(Z|X = x)$ are independent.

We will use the notion of an event in several ways.

• It may represent a subset $S \subseteq X$, in which case we will denote the associated event by $S$, with some abuse of notation. We will use $\Pr_X[S]$ to represent $\sum_{x \in S} p_X(x)$. If $X$ is correlated with another random variable $Y$, the probability of $S$ conditioned on $(Y = y)$ will be denoted by $\Pr_{X|Y = y}[S]$.

• It may represent a predicate over the set $X$, for instance, `$f(x) > 1$ and $f(x) < 5$' for some function $f : X \rightarrow [10]$. In such a case we represent the probability of the event as $\Pr_{x \in X}[f(x) > 1 \text{ and } f(x) < 5]$. If $X$ is correlated with $Y$, the probability of the event conditioned on $(Y = y)$ will be denoted by $\Pr_{x \in X|Y = y}[f(x) > 1 \text{ and } f(x) < 5]$.

• It may represent a predicate over the random variables, for instance `$I > J$' where $I, J$ take values over $[n]$. The probability of such an event will be represented as $\Pr[I > J]$. If $I, J$ are correlated with $X$, the probability of $I > J$ conditioned on $(K = k)$ will be denoted by $\Pr[I > J|K = k]$.

In almost all instances, we will find that the conditioned random variable is clear from the context. In such a case, we will drop the conditioned random variable from the probabilities, for instance, by abbreviating $p_{X|Y = y}$ as $p_{X|y}$.

The complement of an event $E$ is denoted by $\neg E$. The indicator of the event $E$ is denoted by $I(E)$. For a random variable $X$ over a set $X$ and a set $S \subseteq X$, the random variable $X'$ distributed as $p_{X'}(x) = \frac{\mathbb{1}_{x \in S} p_X(x)}{\Pr_X[S]}$ is called a restriction.
of $X$ to the set $S$. For a random variable $X$ taking values over $X$ and a function $f : X \to \mathcal{X}'$, we denote by $f(X)$ the random variable obtained by sampling $x$ according to $X$ and then applying $f$ to it. We use the same notation when $X$ is correlated with other random variables.

Given random variables $X$ and $X'$ taking values in $\mathcal{X}$, we define the $\ell_1$ distance between the two as
\[
\|X - X'\|_1 \overset{\text{def}}{=} \sum_{x \in \mathcal{X}} |p_X(x) - p_{X'}(x)|.
\]
The $\ell_1$ distance is a measure of how distinguishable $X$ is from $X'$. Note that we are slightly modifying the standard notation of $\|p_X - p_{X'}\|_1$. This will be convenient when several random variables are involved in the $\ell_1$ distance, for instance in Fact A.6. Another measure is the KL-divergence, also termed as the relative entropy.
\[
D(X \| X') \overset{\text{def}}{=} \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{p_{X'}(x)},
\]
where we assume $0 \log \frac{0}{0} = 0$. If $\text{supp}(X) \not\subseteq \text{supp}(X')$, then we choose the convention that $D(X \| X') \overset{\text{def}}{=} \infty$. We are interested in one-shot information theory, while the relative entropy possesses an operational meaning only in the asymptotic and i.i.d. setting. Thus, we will need the following two variations on the relative entropy. The max information spectrum divergence captures the largest ratio between $p_X$ and $p_{X'}$ (when viewed as functions from $\mathcal{X} \to [0,1]$), if we are willing to ignore very large ratios occurring with small probability according to $X$.
\[
\tilde{D}_s(X \| X') \overset{\text{def}}{=} \min \left\{ a : \Pr_{x \sim X} \left[ \frac{p_X(x)}{p_{X'}(x)} \geq 2^a \right] \leq \epsilon \right\},
\]
where $\epsilon \in [0,1)$. Note that the definition of max information spectrum divergence, adapted from [10], is slightly different from the ‘information spectrum relative entropy’ in [26]. This is because the above quantity fits better with the operational tasks we consider. The smooth hypothesis testing divergence is defined as follows (with a small variation with respect to [26], as we only consider deterministic tests; see Appendix B).
\[
\tilde{D}_s^{\text{sh}}(X \| X') \overset{\text{def}}{=} \max \left\{- \log \Pr_{X'}[S] : S \subseteq \mathcal{X}, \Pr_X[S] \geq 1 - \epsilon \right\}.
\]
The motivation behind this is that the set $S$ serves as a test that accepts $X$ with high probability and rejects $X'$ with high probability. This notion also has an alternate expression as an information spectrum quantity. This is made precise in Fact A.1.

### III. MAIN RESULT AND ITS CONSEQUENCES

We revisit Task C, restated here for convenience.

**Definition III.1:** A \((R_1, R_2, \epsilon)\) two-senders-one-receiver message compression with side information at the receiver: Given the joint random variables $XYZN$ satisfying that $M = X - (Y, Z)$ and $N = Y - (X, Z)$, there are three parties Alice, Bob and Charlie holding $X$, $Y$ and $Z$, respectively. Alice sends a message of $R_1$ bits to Charlie and Bob sends a message of $R_2$ bits to Charlie. Charlie outputs a sample $(m, n)$ distributed according to the random variables $M'N'$ such that $\frac{1}{2} ||XYZMN - XYZM'N'||_1 \leq \epsilon$. Sharing randomness is allowed between Alice and Charlie and independently between Bob and Charlie.

Our main result is the following achievability theorem for the above task.

**Theorem III.2:** Given $\epsilon \in (0, 1)$, $\delta \in (0, \frac{1}{9})$ such that $\frac{1}{\sqrt{\delta}}$ is an integer, let $R_1, R_2$ satisfy
\[
\Pr_{x,y,m,n,z \sim XYZMN} \left[ \frac{p_{M|X=m}(m)}{p_{M|M=n}(m)} \right] \leq \delta \cdot 2^{R_1}
\]
and
\[
\Pr_{x,y,m,n,z \sim XYZMN} \left[ \frac{p_{N|Y=n}(n)}{p_{N|M=n}(n)} \right] \leq \delta \cdot 2^{R_2}
\]
and
\[
\delta^4 \cdot 2^{R_1 + R_2 - \log \log \frac{1}{\epsilon}} \geq 1 - \epsilon.
\]

There exists a \((R_1 + 3 \log \frac{1}{\epsilon}, R_2 + 3 \log \frac{1}{\epsilon} + 10\delta)\) two-sender-one-receiver message compression with side information at the receiver (Task C).

As mentioned in the introduction, Theorem III.2 can be viewed as a one-shot analogue of Equation 1. For this, observe that we can write the conditional mutual information quantities in Equation 1 as expectations over ratios of probabilities in Equation 3. More precisely,
\[
\begin{align*}
I(X : M | NZ) & = \mathbb{E}_{x,m,n,z \sim XZN} \log \frac{p_{M|X=x,n,z}(m)}{p_{M|M=x}(m)} \\
& = \mathbb{E}_{x,y,m,n,z \sim XYNZ} \log \frac{p_{M|X=x}(m)}{p_{M|M=x}(m)}, \\
I(Y : M | NZ) & = \mathbb{E}_{y,m,n,z \sim YNZ} \log \frac{p_{N|Y=y,m,z}(n)}{p_{N|M=m}(n)}, \\
& = \mathbb{E}_{x,y,m,n,z \sim XYNZ} \log \frac{p_{N|Y=y}(n)}{p_{N|M=m}(n)}
\end{align*}
\]
and
\[
\begin{align*}
I(XY : MN | Z) & = \mathbb{E}_{x,y,m,n,z \sim XYNZ} \log \frac{p_{MN|X=x,Y=y}(m,n)}{p_{M|M=x}(m)} \\
& = \mathbb{E}_{x,y,m,n,z \sim XYNZ} \log \frac{p_{MN|X=x,Y=y}(m,n)}{p_{M|M=x}(m)}. \\
\end{align*}
\]
Thus, theorem III.2 captures the ‘worst-case’ behaviour of the rate region in Equation 1, up to an error of $\epsilon$ (which is allowed by the protocol) and additive loss of $\approx \log \log \max \{M, |N|\}$ in communication (which could potentially be improved). The theorem can be used to study several other tasks as we show below, owing to the generality of Task C itself.

#### A. Lossy Distributed Source Coding

Lossy source coding is a well studied task in information theory [16], [27]–[31], where a sender observes a sample from a source and the receiver is allowed to output a distorted version of this sample. Our first application is the lossy distributed source coding [22], [23], [32], [33] (which is...
a distributed version of the lossy source coding), which is defined as follows (observe that we also include the size information with the receiver in our definition below).

**Definition III.3:** A $(k, \epsilon, \delta)$-lossy distributed source coding with side information. Given $k > 0$ and $\epsilon \in (0, 1)$ and joint random variables $XYZ$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, Alice, Bob and Charlie observe $X, Y$ and $Z$, respectively. Both Alice and Bob send messages to Charlie, who is required to output joint random variables $X', Y'$ such that $\Pr_{X'Y'|X'Y'}[d(X', Y'|X, Y') \geq k] \leq \epsilon$, where $d : \text{supp}(X) \times \text{supp}(Y) \times \text{supp}(X') \times \text{supp}(Y') \rightarrow (0, +\infty)$ is a distortion measure. There is no shared randomness among the parties.

The following theorem obtains a nearly tight one-shot bound for this task, that uses bounded-sized auxiliary random variables. It can be viewed as a one-shot analogue of the Berger-Tung bound [22], [23] (when the side information with the receiver is absent, see [21, Section 12.1]). Note that auxiliary random variables also arise in the characterization of the lossy source coding task (see [21, Section 3.6]).

**Theorem III.4:** Given $\epsilon, \delta, \delta' \in (0, 1)$ and joint random variables $XYZ$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. For every $(k, \epsilon)$-lossy distributed source coding protocol with $N$ bits of communication from Alice to Charlie and $R_2$ bits of communication from Bob to Charlie, there exist random variables $MN$ taking values over a set $\mathcal{M} \times \mathcal{N}$ and satisfying $M - X - YZN$, $N - Y - XMZ$. Let $\Pr_{XYZMN}[d(X, Y, f(M, N, Z)) \geq k] \leq \epsilon$, there exists a $(k, \epsilon + \delta + 8\delta')$-lossy distributed source protocol with communication $R_1$ from Alice to Charlie and communication $R_2$ from Bob to Charlie such that

$$\Pr_{X,Y,Z,M,N} \left[ \frac{p_M(x,z | m)}{p_{M|z}(m,z)} \leq \frac{\delta}{\delta'}, \frac{p_{N|y,z}(n,z)}{p_{M|z}(m,z)} \leq \frac{\delta}{\delta'}, \frac{p_{M|z}(m,z)}{p_{MN}(m,n)} \leq \frac{2R_1 + R_2}{\delta} \right] \geq 1 - 3\delta.$$ 

Furthermore, for any joint random variables $MN$ satisfying $M - X - YZN$, $N - Y - XMZ$ and function $f : \mathcal{M} \times \mathcal{N} \rightarrow \mathcal{X}' \times \mathcal{Y}'$ satisfying $\Pr_{XYZMN}[d(X, Y, f(M, N, Z)) \geq k] \leq \epsilon$, there exists a $(k, \epsilon + \delta + 8\delta')$-lossy distributed source protocol with communication $R_1$ from Alice to Charlie and communication $R_2$ from Bob to Charlie such that

$$\Pr_{X,Y,Z,M,N} \left[ \frac{p_M(x,z | m)}{p_{M|z}(m,z)} \leq \delta', \frac{p_{N|y,z}(n,z)}{p_{M|z}(m,z)} \leq \delta', \frac{p_{M|z}(m,z)}{p_{MN}(m,n)} \leq \frac{2R_1 + R_2}{\delta} \right] \geq 1 - \delta.$$

**Proof:** We divide the proof into two parts.

- **Converse.** Given a $(k, \epsilon)$-lossy distributed source coding protocol, we can fix the local randomness used by Alice and Bob for encoding and obtain a new protocol where $X$ and $Y$ are mapped deterministically to the messages sent. Observe that this does not change the error and does not increase the communication cost. We choose $M$ and $N$ to be the messages sent by Alice and Bob respectively. Without loss of generality, we can assume that $|M| \leq |X|$ and $|N| \leq |Y|$. It holds that $M - X - YZN$ and $N - Y - XMZ$. Let $f$ be the function applied by Charlie to obtain $X'Y'$. By the correctness of the protocol, it holds that $\Pr_{XYZMN}[d(X, Y, f(M, N, Z)) \geq k] \leq \epsilon$. Let $U_1$ and $U_2$ be random variables uniformly distributed over $M$ and $N$, respectively. Note that for any $(x, y, z, m, n),$

$$p_M(x,z | m) \leq 2R_1 p_{U_1}(m),$$
$$p_{N|y,z}(n) \leq 2R_2 p_{U_2}(n),$$
$$p_{MN}(m,n) \leq 2R_1 + R_2.$$ 

We further have

$$\Pr_{M,N,Z} \left[ \frac{p_{MN}(m,n)}{p_{MN}(m,z)} \leq \frac{\delta}{\delta'}, \frac{p_{U_1}(m)}{p_{U_2}(n)} \leq \delta \right] \leq \delta,$$

Then

$$\Pr_{X,Y,Z,M,N} \left[ \frac{p_{MN}(m,n)}{p_{MN}(m,z)} \leq \frac{2R_1 + R_2}{\delta} \right] \geq 1 - \delta.$$

The converse follows from the union bound.

- **Achievability.** Given joint random variables $XYZMN$ satisfying $M - X - YZN$ and $N - Y - XMZ$, the parties run the protocol present in Theorem III.2 with $\epsilon \leftarrow \delta, \delta \leftarrow \delta'$. As guaranteed by Theorem III.2, the protocol outputs random variables $M'N'$ satisfying that

$$\frac{1}{2} \|X'Y'ZMN - X'Y'ZM'N'\|_1 \leq \epsilon + \delta + 3\delta'.$$

Then Charlie applies $f$ to $(M', N', Z)$ to get $(X', Y')$. It holds that

$$\Pr_{X,Y,Z,M,N} \left[ d(x, y, f(m, n, z)) \geq k \right] \leq \Pr_{X',Y',Z,M,N} \left[ d(x, y, f(m, n, z)) \geq k \right] + \frac{1}{2} \|X'Y'ZMN - X'Y'ZM'N'\|_1 \leq \epsilon + \delta + 3\delta'.$$

This completes the proof. □
B. Recovering Achievable Communication for DSC Task and Task B

Another application is the following corollary, for the problem of DSC. While it is a special case of lossy distributed source coding, it is possible to obtain a simpler bound without introducing auxiliary random variables. We reproduce the nearly optimal one-shot bound given in [10], up to an additive factor of \( \log log \max \{|X|, |Y|\} \).

Corollary III.5: Let \( \epsilon \in (0,1) \) such that \( \frac{1}{\sqrt{N}} \) is an integer. Let \( R_1, R_2 \) satisfy

\[
\begin{align*}
\Pr_{x,y \sim \mathcal{X}} & \left[ \frac{1}{p_{X|Z}(x)} \leq \delta \cdot 2^{R_1} \right. \\
& \quad \left. \frac{1}{p_{Y|Z}(y)} \leq \delta \cdot 2^{R_2} \right. \\
& \quad \left. \text{and} \quad \frac{1}{p_{X,Y|Z}(x,y)} \leq \delta^4 \cdot 2^{R_1 + R_2 - \log log \max \{|X|, |Y|\}} \right] \\
& \geq 1 - \epsilon.
\end{align*}
\]

There exists a protocol satisfying the following:

- No players share public coins.
- Alice and Bob observe a sample from \( X \) and \( Y \) and then send \( R_1 + 3 \log \frac{1}{\delta} \) bits and \( R_2 + 3 \log \frac{1}{\delta} \) bits to Charlie, respectively;
- Charlie outputs the random variables \( X'Y' \) such that
  \[ \Pr\{XY \neq X'Y'\} \leq \epsilon + 8\delta. \]

Proof: Applying Theorem III.2 with \( Z \) trivial, \( M = X \) and \( N = Y \), we obtain a randomness-assisted protocol with communications \( R_1 \) and \( R_2 \) from Alice and Bob respectively. Charlie outputs random variables \( X', Y' \) such that

\[ \frac{1}{2} \|XYXY - XYX'Y'\|_1 \leq \epsilon + 8\delta. \]

Let \( S = \{(x, y, x', y'): x \in \mathcal{X}, y \in \mathcal{Y}\} \). Then

\[
1 - \Pr\{XY = X'Y'\} = \left| \Pr_{XYXY \sim S} - \Pr_{XYXY' \sim S} \right| \leq \frac{1}{2} \|XYXY - XYX'Y'\|_1 \leq \epsilon + 8\delta.
\]

This completes the proof by the standard derandomization argument to fix the shared randomness.

We also reproduce the main results in [9] and [10] for Task B, up to additive factor of \( \log log |N| \). This is obtained by setting \( Y \) and \( N \) to be trivial in Theorem III.2.

Corollary III.6: Let \( \epsilon \in (0,1) \) such that \( \frac{1}{\sqrt{N}} \) is an integer. Let \( R \) satisfy

\[
\Pr_{x \sim \mathcal{X}, z \sim \mathcal{Z}} \left[ \frac{1}{p_{M|z}(m)} \leq \delta \cdot 2^{R - \log \log \frac{1}{2\epsilon}} \right] \geq 1 - \epsilon.
\]

There exists a protocol satisfying the following:

- Alice and Charlie share public random coins;
- Alice sends \( R + 3 \log \frac{1}{\delta} \) bits to Charlie;
- Charlie outputs the random variable \( M' \) such that
  \[
  \frac{1}{2} \|XZM - XZM'\|_1 \leq \epsilon + 8\delta.
  \]

C. Nearly Optimal Characterization of Task C in Terms of the Auxiliary Random Variables

Theorem III.2 can be further used to obtain a near-optimal characterization of Task C itself. We first show how to reduce the amount of the shared randomness in any given \( (R_1, R_2, \epsilon) \) two-sender-one-receiver message compression with side information at the receiver, using an argument similar to Wyner [34] and Newman [35]. Since their arguments do not apply in the multi-partite setting (notice that the new randomness must be shared independently between Alice, Charlie and Bob, Charlie), we replace the Chernoff bound arguments in [34], [35] with an argument based on the bipartite convex-split lemma (Fact A.6).

Claim III.7: Fix \( \epsilon, \delta \in (0,1) \). For any \( (R_1, R_2, \epsilon) \) two-sender-one-receiver message compression with side information at the receiver, there exists another \( (R_1, R_2, \epsilon + 2\delta) \) two-sender-one-receiver message compression with side information at the receiver that uses at most \( \log \frac{24|N|}{\epsilon \delta} \) bits of the shared randomness between Alice and Charlie as well as Bob and Charlie.

Proof: Given a protocol where Alice sends \( S_1 \) bits to Charlie, Bob sends \( S_2 \) bits to Charlie and Charlie outputs \( M'N' \) such that \( \frac{1}{2} \|XYZM'N' - XZYM'N'|_1 \leq \epsilon \), let \( S_1 \) be the shared randomness between Alice and Charlie and \( S_2 \) be the shared randomness between Bob and Charlie. Let \( T_1 \) be the message generated by Alice conditioned on \( S_1, X \) and \( T_2 \) be the message generated by Bob conditioned on \( S_2, Y \). We apply Fact A.6 to the random variables \( XYZM'N'S_1S_2 \) with \( L_1, L_2 \) chosen such that

\[
\begin{align*}
\Pr_{x,y,z,m,n,s \sim \mathcal{X},\mathcal{Y},\mathcal{Z},\mathcal{M},\mathcal{N},\mathcal{S}} & \left[ \frac{p_{XYZM'N'S_1S_2}(x,y,z,m,n,s)}{p_{XYZM'N'(x,y,z,m,n,s)}} \leq \frac{\epsilon^2}{24} \cdot L_1 \right] \\
\text{and} & \quad \frac{p_{XYZM'N'(x,y,z,m,n,s)}}{p_{XYZM'N'(x,y,z,m,n,s)}} \leq \frac{\epsilon^2}{24} \cdot L_1 \frac{L_2}{L_1}
\end{align*}
\]

\[
\geq 1 - \delta
\]

to obtain the random variables \( XZYM'N'S_1 \ldots S_{2L_2} \) (with \( S_i = 1 \) and \( S_{2i-1} = S_{2i} = 2 \)) that satisfies

\[
\frac{1}{2} \|XYZM'N'S_1 \ldots S_{2L_2} - XZYM'N' \times S_1 \times \ldots \times S_{2L_2} \times S_1 \times \ldots \times S_{2L_2} \|_1 \leq 2\delta.
\]

This expression can be re-arranged to obtain

\[
\left\| \mathbb{E}_{S_1, \ldots, S_{2L_1}, S_1, \ldots, S_{2L_2} \sim S_1 \times \ldots \times S_1 \times S_2 \times \ldots \times S_2} \mathbb{E}_{j \sim [L_1], j \sim [L_2]} \left[ (XYZM'N'|S_1S_2 = s_1, s_2) \right] \right\|_1 \leq 2\delta.
\]

Thus, there exists \( s_1, \ldots, s_{2L_1}, s_1, \ldots, s_{2L_2} \) such that

\[
\left\| \mathbb{E}_{j \sim [L_1], j \sim [L_2]} \left[ (XYZM'N'|S_1S_2 = s_1, s_2) \right] \right\|_1 \leq 4\delta.
\]
The new protocol is as follows.

- Alice and Charlie share uniform randomness $U_1$ taking values in $[L_1]$. Bob and Charlie share uniform randomness $U_2$ taking values in $[L_2]$.
- Conditioned on the value $i \sim U_1$, Alice generates $(T_i | X Y Z, S_1 = s_1^i)$ and sends it to Charlie. Conditioned on the value $j \sim U_2$, Bob generates $(T_j | X Y Z, S_2 = s_2^j)$ and sends it to Charlie.
- Charlie, who also observes $(i, j) \sim U_1 \times U_2$, generates $M'N'$ conditioned on $(s_1^i, s_2^j)$.
- Let the output of Charlie, averaged over the shared randomness, be $M''N''$.

It holds that

$$XY Z M'' N'' = \mathbb{E}_{i \sim U_1, j \sim U_2} \left[ X Y Z (M'N' \mid S_1 S_2 = s_1^i, s_2^j) \right].$$

Thus, Eq. (12) guarantees that

$$\frac{1}{2} \| X Y Z M'' N'' - X Y Z M N \|_1 \leq \epsilon + 2\delta.$$

To bound the size of shared randomness, observe that Eq. (11) can be rephrased as follows, using the fact that $X Y Z$ is independent of $S_1 S_2$:

$$\Pr_{x, y, z, m, n, s_1, s_2} \left[ \frac{p_{M'' N'' | x, y, z}(m, n)}{p_{M'' N''}(m, n)} \leq \frac{\delta^2}{24} \cdot L_1 \right] \leq \delta,$$

and probabilities are at most 1, we have

$$\Pr_{x, y, z, m, n, s_1, s_2} \left[ \frac{p_{M'' N'' | x, y, z}(m, n)}{p_{M'' N''}(m, n)} \leq \frac{\delta^2}{24} \cdot L_1 \right] \geq 1 - \delta.$$

if we choose $L_1 = \frac{24 |M| |N|}{\delta^2}$ and $L_2 = \frac{24 |M| |N|}{\delta^2}$. This completes the proof.

We have the following theorem for Task C (Definition III.1), in terms of auxiliary random variables of a bounded size.
the function that Charlie applies on $S_1, S_2, T_1, T_2, Z$ to obtain $M' N'$. Then

$$
p_{T_1 S_1 | x, z} (t_1, s_1) 
\leq p_{S_1 | x, z} (s_1) = 2^{R_1} p_{U_1} (t_1) p_{S_1} (s_1)
$$

and

$$
p_{S_2 | y, z} (s_2) = 2^{R_2} p_{U_2} (t_2) p_{S_2} (s_2)
$$

and

$$
p_{S_1 S_2 T_1 T_2 | x, y, z} (t_1, t_2, s_1, s_2)
= 2^{R_1 + R_2} p_{U_1} (t_1) p_{U_2} (t_2) p_{S_1 S_2} (s_1, s_2).
$$

The rest of the proof follows closely the converse proof given in Theorem III.4.

- **Achievability**: The achievability also follows along the lines similar to Theorem III.4. By a straightforward application of Theorem III.2, Alice and Bob communicate $R_1$ and $R_2$ bits respectively to Charlie such that Charlie is able to output $S_1' S_2' T_1' T_2'$ satisfying

$$
\frac{1}{2} \| XY Z S_1' S_2' T_1' T_2' - XY Z S_1 S_2 T_1 T_2 \|_1 \leq 9\delta_6.
$$

Charlie now applies the function $f$ to obtain the desired output. It holds that

$$
\frac{1}{2} \| XY f (Z S_1' S_2' T_1' T_2') - XY Z MN \|_1 \leq \epsilon + 9\delta_6.
$$

This completes the proof.

\[\square\]

D. Comparison With the Bound Obtained in [10]

In [10, Theorem 4], the following achievable communication region was obtained for the task in Definition III.1 (the authors also state a more general bound optimized over all possible extensions of $M N$). But it involves auxiliary random variables of unbounded size:

$$
R_1 \geq \hat{D}_s^d (XM \| X \times S) - \hat{D}_H^d (MNZ \| S \times NZ) + 4 \log \frac{3}{\delta},
$$

$$
R_2 \geq \hat{D}_s^d (YN \| Y \times T) - \hat{D}_H^d (MZN \| MZ \times T) + 4 \log \frac{3}{\delta},
$$

$$
R_1 + R_2 \geq \hat{D}_s^d (XM \| X \times S) + \hat{D}_s^d (YN \| Y \times T)
- \hat{D}_H^d (MZN \| S \times T \times Z) + 6 \log \frac{3}{\delta},
$$

(13)

giving the overall error of $\epsilon_1 + \epsilon_2 + \epsilon_3 + 13\delta$. Above, $S$ and $T$ are arbitrary random variables over $M$ and $N$, respectively. The following claim shows that this achievable communication region is contained in the achievable communication region stated in Theorem III.2, up to the additive factor of $\log \max \{|M|, |N|\}$.

**Theorem III.9**: For any $(R_1, R_2)$ satisfying Eqs (13), it holds that

$$
\text{Pr}_{x, y, z, m, n \sim XY Z M N}
\begin{bmatrix}
\frac{p_{M | x} (m)}{p_{M | z} (m)} \\
\frac{p_{N | y} (n)}{p_{N | z} (n)}
\end{bmatrix}
\leq \frac{2 \epsilon_1}{3 \epsilon_2}
$$

and

$$
\frac{p_{N | m} (n)}{p_{N | m} (n)} \leq \frac{2 \epsilon_2}{3 \epsilon_2}
$$

and

$$
\frac{p_{M | x} (m) p_{N | y} (n)}{p_{M N | m, n} (m, n)} \leq \frac{2 \epsilon_3}{3 \epsilon_2} R_1 + R_2
$$

\[\geq 1 - \epsilon_1 - \epsilon_2 - \epsilon_3 - 2\delta.
$$

**Proof**: Let $k_1 \overset{\text{def}}{=} \hat{D}_s^d (XM \| X \times S)$, $k_2 \overset{\text{def}}{=} \hat{D}_H^d (MNZ \| S \times NZ)$, $k_3 \overset{\text{def}}{=} \hat{D}_s^d (YN \| Y \times T)$, $k_4 \overset{\text{def}}{=} \hat{D}_H^d (MZN \| MZ \times T)$ and $k_5 \overset{\text{def}}{=} \hat{D}_H^d (MZN \| S \times T \times Z)$. Define the following sets, which naturally arise from the definitions of the above quantities:

$$
S_1 \overset{\text{def}}{=} \{(x, y, z, m, n) : \frac{p_{M | x} (m)}{p_{S} (m)} \leq 2^{k_1}\}
$$

$$
S_2 \overset{\text{def}}{=} \{(x, y, z, m, n) : \frac{p_{M | n} (m)}{p_{S} (m)} \geq 2^{k_2}\}
$$

$$
S_3 \overset{\text{def}}{=} \{(x, y, z, m, n) : \frac{p_{N | y} (n)}{p_{T} (n)} \leq 2^{k_3}\}
$$

$$
S_4 \overset{\text{def}}{=} \{(x, y, z, m, n) : \frac{p_{N | m} (m)}{p_{T} (n)} \geq 2^{k_4}\}
$$

$$
S_5 \overset{\text{def}}{=} \{(x, y, z, m, n) : \frac{p_{M N | m} (m, n)}{p_{S} (m) p_{T} (n)} \geq 2^{k_5}\}.
$$

From the definition of $\hat{D}_s^d (XM \| X \times S)$ and $\hat{D}_s^d (YN \| Y \times T)$, it holds that $\Pr_{X Y Z M N} [S_1] \geq 1 - \delta$ and $\Pr_{X Y Z M N} [S_3] \geq 1 - \delta$. Using Fact A.1, we further conclude that $\Pr_{X Y Z M N} [S_2] \geq 1 - \epsilon_1 - \delta$, $\Pr_{X Y Z M N} [S_4] \geq 1 - \epsilon_2 - \delta$ and $\Pr_{X Y Z M N} [S_5] \geq 1 - \epsilon_3 - \delta$. Define $S = S_1 \cap S_2 \cap S_3 \cap S_4 \cap S_5$. By union bound, we have $\Pr_{X Y Z M N} [S] \geq 1 - \epsilon_1 - \epsilon_2 - \epsilon_3 - 5\delta$. Moreover, for all $(x, y, z, m, n) \in S$, the following inequalities simultaneously hold:

$$
\frac{p_{M | x} (m)}{p_{S} (m)} \leq 2^{k_1}, \quad \frac{p_{M | n} (m)}{p_{S} (m)} \geq 2^{k_2}, \quad \frac{p_{N | y} (n)}{p_{T} (n)} \leq 2^{k_3}, \quad \frac{p_{N | m} (m)}{p_{T} (n)} \geq 2^{k_4}, \quad \frac{p_{M N | m} (m, n)}{p_{S} (m) p_{T} (n)} \geq 2^{k_5}.
$$

Combining them, we find that

$$
\frac{p_{M | x} (m)}{p_{M | n} (m)} \leq \frac{2^{k_1 - k_2}}{\delta}, \quad \frac{p_{N | y} (n)}{p_{N | m} (m)} \leq \frac{2^{k_3 - k_4}}{\delta}
$$

and

$$
\frac{p_{M | x} (m) p_{N | y} (n)}{p_{M N | m} (m, n)} \leq \frac{2^{k_1 + k_2 - k_5}}{\delta}.
$$

Substituting the values of $R_1, R_2$ the proof concludes.

Now we give an example where the communication region in Eq. (13) is strictly contained in the communication region in Theorem III.2. Consider the case where $X = Y = Z$ and $M = N$. Let $X = Y = Z$ and $(M | X = x) = (N | X = x)$ for all $x \in X$. Thus, for all $(x, y, z, m, n) \sim XY Z M N$, we have $x = y = z$. Furthermore, the condition $M - X - Z - Y - N$ equivalent to $M - X - N$. Using these observations, the following identities hold for all $(x, y, z, m, n) \sim XY Z M N$.

$$
\frac{p_{M | x} (m)}{p_{M | n} (m)} = \frac{p_{M | x} (m)}{p_{M | z} (m)} = \frac{p_{M | x} (m)}{p_{S} (m)} = 1
$$

$$
\frac{p_{N | y} (n)}{p_{N | m} (m)} = \frac{p_{N | y} (n)}{p_{N | z} (n)} = \frac{p_{N | y} (n)}{p_{N} (n)} = 1
$$

$$
\frac{p_{M | x} (m) p_{N | y} (n)}{p_{M N | m} (m, n)} = \frac{p_{M | x} (m) p_{N | y} (n)}{p_{M N} (m, n)} = \frac{p_{M | x} (m) p_{N | y} (n)}{p_{M | x} (m) p_{N | y} (n)} = 1.
$$

(14)
This immediately leads to the following corollary of Theorem III.2.

**Corollary III.10:** Given \( \delta \in (0, \frac{1}{3}) \) such that \( \frac{1}{\delta} \) is an integer, suppose \( XYZMN \) is such that \( X = Y = Z \) and \( (M|X = x) = (N|X = x) \). Let \( R_1, R_2 \) satisfy
\[
R_1 \geq \log \frac{1}{\delta}, \quad R_2 \geq \log \frac{1}{\delta}
\]
and
\[
R_1 + R_2 \geq \log \log \frac{|N|}{\delta} + 4 \log \frac{1}{\delta}.
\]
There exists a \((R_1 + 3 \log \frac{1}{\delta}, R_2 + 3 \log \frac{1}{\delta}, 8\delta)\) two-sender-one-receiver message compression with side information at the receiver (Task C) for the random variables \( XYZMN \).

**Proof:** We substitute Eq. (14) in Eq. (3) and set \( \epsilon = 0 \). This shows that for any \((R_1, R_2)\) satisfying
\[
\Pr_{x,y_z,m,n} \left[ 1 \leq \delta \cdot 2^{R_1}, \text{ and } 1 \leq \delta^4 \cdot 2^{R_1 + R_2 - \log \log \frac{|N|}{\delta}} \right] = 1, \quad (15)
\]
there exists a \((R_1 + 3 \log \frac{1}{\delta}, R_2 + 3 \log \frac{1}{\delta}, 8\delta)\) two-sender-one-receiver message compression with side information at the receiver (Task C) for the random variables in the statement of the corollary. But the constraint on \((R_1, R_2)\) in Eq. (15) is equivalent to the constraint on \((R_1, R_2)\) in the statement of the corollary. This completes the proof.

The example is constructed in the following claim.

**Claim III.11:** Let \( X = M = [n] \). Define the random variables \( XYZMN \) as follows. \( X = Y = Z \), \( X \) is uniform in \([n]\), \((M|X = x) = (N|X = x)\) and
\[
p_{M|X=x}(m) = \begin{cases} \frac{\epsilon_1}{2|x|^3} & \text{if } m = x \\ \text{otherwise} \end{cases}
\]
Let \( \delta \in (0, \frac{1}{3}) \) and \( \epsilon_1, \epsilon_2 \in (0, \frac{1}{3}) \). It holds that for any random variables \( S, T \) taking values over \([n]\),
\[
\hat{D}_\delta^S(XM||X \times S) - \hat{D}_\delta^T(MNZ||S \times NZ) \geq \log(|X|) - 2
\]
and
\[
\hat{D}_\delta^S(YN||Y \times T) - \hat{D}_\delta^H(MZN||MZ \times T) \geq \log(|X|) - 2.
\]

The proof of this claim is given in Appendix E. Thus, there exist joint random variables \( XYZMN \) for which the communication region in Eq. (13) requires \( \log |X| \) bits of communication from Alice to Charlie and \( \log |X| \) bits from Bob to Charlie. This is in sharp contrast with the communication region in Corollary III.10, which gives a protocol where sum of the communication is \( \approx \log \log |N| = \log \log |X| \). Observe that the optimal communication for this random variable is zero (as Charlie can locally generate \( MN \)). But the example serves as a simple demonstration of the strict containment of the communication region.

**IV. PROOF OF THEOREM III.2 AND THE DETAILS OF TECHNIQUES USED**

This section is devoted to the proof of Theorem III.2. Before proceeding to the proof in Subsection IV-E, we discuss the key techniques that will be used to construct our protocol. For this, we consider the protocols in [9], [10] for the case of one sender and one receiver (Task B), and highlight the limitation in their extension to Task C.

**A. Methods of Convex-Split and Hypothesis Testing**

The protocol in [9] relies on the methods of rejection sampling [4], [5], [36]. The protocol in [10] uses convex-split [24] and hypothesis testing via position-based decoding [25]. Let us review the second and the third methods, which will be used in our proofs, and refer the reader to [4], [5], [9] for the description of rejection sampling.

The convex-split method is centered around Fact A.5, which shows that for \( R \) large enough, the random variables \( X_1M_2 \ldots M_{2^n} \) is close to \( X \times W \times W \times \ldots \times W \) (c.f. Fact A.5 for the definitions of these random variables). View \( X \) as the input given to Alice and let Alice and Charlie share \( 2^R \) copies of \( W \). Since the resulting random variables \( X \times W \times W \times \ldots \times W \) is close to \( X_1M_2 \ldots M_{2^n} \), the parties can assume that the shared randomness is in fact \( M_1, M_2, \ldots, M_{2^n} \) (with Alice additionally possessing \( X \)). Alice considers a sample \((x, m_1, m_2, \ldots, m_{2^n})\) from \( M_1M_2 \ldots M_{2^n} \). Recall that \( M \) and \( W \) take values over the same set, hence this is a valid step in the protocol. She generates the random variable \( J \) conditioned on this sample, distributed as
\[
\Pr[J = j|X_1 \ldots M_{2^n} = (x, m_1, \ldots, m_{2^n})] = \frac{\Pr[J = j|X_1 \ldots M_{2^n} = (x, m_1, \ldots, m_{2^n})]}{\Pr[X_1 \ldots M_{2^n} = (x, m_1, \ldots, m_{2^n})]}
\]
Here, the probabilities on the right hand side are given in Fact A.5. Now, conditioned on \( J = j \), Eq. (41) ensures that \( X \) and \( M_j \) are correlated according to the random variables \( X \times M \). Alice communicates the random variable \( J \) to Charlie, who uses it to output \( M_j \). In this way, Charlie is able to output \( M \) which is correlated with \( X \) according to the random variables \( X \times M \).

The position-based decoding method is based upon the smooth hypothesis testing divergence. Imagine a variation on the above protocol where Alice does not communicate the random variable \( J \) to Charlie, but sends the random variable \( J' \) defined as \( \lfloor J/2^b \rfloor \) for some integer \( b \). In such a case, Charlie is still unaware of the actual value of \( J \), as there are \( 2^b \) possible realizations of \( J \) conditioned on the \( J' \) he observes. But observe the following crucial property using Eq. (41); the random variable \( M_j \) is distributed according to \( M \) and all other random variables with Charlie are distributed according to \( W \). If Charlie performs hypothesis testing to distinguish between them, the probability that he confuses a \( W \) with \( M_j \) is at most \( 2^{-\hat{D}_H(M||W)} \). Through a simple application of the union bound, we infer that if \( (2^b - 1) \cdot 2^{-\hat{D}_H(M||W)} \leq \epsilon \), the probability that he confuses any of the \( 2^b \) copies of \( W \) with \( M_j \) is at most \( \epsilon \). This allows \( b \) to be chosen to be \( \hat{D}_H(M||W) - \log \frac{1}{\epsilon} \). Evidently, this helps in reducing the cost of communication, at the expense of small increase in error.

Equipped with these two notions, we now proceed to the discussion of previous protocols for Task B.
B. Revisiting Previous Protocol for Task B

We start with describing the set up in [9]. We will not touch upon the details of the encoding and decoding operations, as those will not be needed in our proofs. Let us assume that \( p_{M|x}(m) \) and \( p_{M|z}(m) \) are rational numbers for all \( x, m, z \), which can be achieved by perturbing by an arbitrarily small amount. Alice and Charlie share sufficiently large number of copies of a random variable \( U \) uniformly distributed over the set \( M \times [K] \), where \( K \) is a sufficiently large integer. The choice of \( K \) is made in a manner that \( K p_{M|x}(m) \) and \( K p_{M|z}(m) \) are integers for all \( x, m, z \). The motivation behind sharing \( U \) is that this is a natural choice that does not depend on \( x \) and \( z \), and hence can be ‘pre-agreed’ upon by both parties. The number of required copies of this uniformly distributed random variable suffices to be approximately \( |M|/\epsilon \), where \( \epsilon \) is the error parameter. The parties index each of these copies with a unique integer in \( |M|/\epsilon \). Recall that given \( x \), Alice has the knowledge of the distribution \( p_{M|x} \). Given \( z \), Charlie has the knowledge of the distribution \( p_{M|z} \).

\[
2^c \overset{\text{def}}{=} \max_{m,x,z:p_X(x,z)>0} \frac{p_{M|x}(m)}{p_{M|z}(m)}, \tag{16}
\]

as the largest possible ratio between these probabilities, for all \( x, z \) in the support of \( XZ \). It can be viewed as a one-shot analogue of the conditional mutual information, since the conditional mutual information can be written as

\[
I(X : M | Z) = \sum_{x,m,z} p_{X,M,Z}(x,m,z) \log \frac{p_{M|x}(m)}{p_{M|z}(m)}.
\]

In the last equality, we have used the condition \( M = X - Z \). The key property of this set-up is that for any element \((m, e) \in M \times [K] \) (in the domain of \( U \)) satisfying \( e \leq K p_{M|x}(m) \), it also holds that \( e \leq K \cdot 2^c \cdot p_{M|z}(m) \) (see Figure 1). This property is used in the encoding and decoding operations in [9], and the number of bits communicated is approximately \( c \).

In [10], it was shown that the protocol of [9] can be viewed in the context of convex-split and hypothesis testing. Alice and Charlie divide \( |M|/\epsilon \) copies of shared randomness in approximately \( 2^c \) blocks (see Figure 2). Alice uses the convex-split method to find the index of the first block properly correlated with \( X \). Namely the first element of the shared randomness \((m, e)\) that satisfies \( e \leq K p_{M|x}(m) \). Denote the random variable associated to \( e \) by \( E \), which extends \( X \) to \( XME \). Alice sends the index of the block to Charlie using approximately \( c \) bits of communication. Within this block lies the correct index that Alice wants Charlie to pick up. Charlie needs to distinguish this index with other indices, where any sample \((m, e)\) has the property that \( e \) is uniform in \([K]\). Charlie uses hypothesis testing in this step. Let \( A \) be the set of all \((m, e)\) satisfying \( p_{M|E}(m,e) > 0 \). In Figure 1, this is the set of all points lying below all the blue curves. The key observation is that this set ensures the following relation:

\[
\Pr_{ME}[A] = 1, \quad \Pr_U[A] \leq \frac{2^c}{|M|}. \tag{17}
\]

The first equality is clear. To prove the second, observe from Figure 1 that the set \( A \) is contained below the dashed red curve, which is the set of all points \((m, e)\) satisfying \( e < K \cdot 2^c \cdot p_{M|z}(m) \). Thus,

\[
|A| \leq \sum_{m} K \cdot 2^c \cdot p_{M|z}(m) \leq 2^c K.
\]

Since \( \Pr_U[A] = \frac{|A|}{|M|} \), Eq. (17) follows. Thus, the test that Charlie uses for the hypothesis testing accepts the random variables \( ME \) at the correct index with probability 1 and accepts the uniform distribution over \( M \times [K] \) with probability at most \( \frac{2^c}{|M|} \).
C. Difficulty in Using Previous Method for Task C

One can try the same approach for Task C, by ‘extending’ the random variables \( M \) and \( N \) to new random variables \( ME \) and \( NF \), where \( e \) is uniform in \([K \cdot p_{M|z}(m)]\) and \( f \) is uniform in \([K \cdot p_{N|y}(n)]\) for a given \( x, y, m, n \). As before, we assume that \( K \) is chosen in a manner such that \( K \cdot p_{M|z}(m) \) and \( K \cdot p_{N|y}(n) \) are integers. Alice and Bob both perform the convex-split steps and Charlie attempts to perform the hypothesis testing step. The key challenge is to construct the correct hypothesis test. Below, we analyze the test given in [10] by considering the full support of the random variables \( MENF \). For the ease of argument, we assume that \( Z \) is trivial.

Recall the quantity \( I(\{XY : MN\}) \) from (1), which was the rate of total communication from Charlie to Alice and Bob. Expanding this quantity, and using \( M - X - Y - N \), we find that

\[
I(\{XY : MN\}) = \sum_{x,y,m,n} p_{XYMN}(x,y,m,n) \log \frac{p_{MN|xy}(m,n) p_{MN}(m,n)}{p_{M}(m)}.
\]

We need a one-shot version of above expression for the protocol we are trying to construct, which is as follows (c.f. Eq. (16)):

\[
2^{e^c} \geq \max_{x,y,m,n:p_{XYMN}(x,y,m,n) > 0} \frac{p_{MN}(m,n)}{p_{M}(m)}.
\]

As discussed earlier, the insight from the protocol outlined in [10] is that the total communication from Alice and Bob to Charlie is \( c^2 \) if the following holds (c.f. Eq. (17)): there exists a test through which Charlie accepts the uniform distribution over \( M \times [K] \times N \times [K] \) with probability at most \( \frac{2^{e^c}}{p_{MN}(m,n)} \).

Recall that \( e \leq K \cdot p_{M|z}(m) \) and \( f \leq K \cdot p_{N|y}(n) \). The definition of \( c^2 \) gives the following guarantee:

\[
e^c \leq K^2 \cdot 2^{c^2} p_{MN}(m,n). \tag{18}
\]

If \( 2^{c^2} p_{MN}(m,n) \geq 1 \) then all pairs \( (e,f) \) satisfy the condition in Eq. (18). Let the set of all such \( (m,n) \) be ‘Bad’ and let the rest be ‘Good’. For \( (m,n) \in \text{Bad} \), the number of pairs \( (e,f) \) that satisfy the condition in Eq. (18) is \( K^2 \). Suppose the hypothesis test \( A \), that aims to distinguish \( MENF \) from the uniform distribution over \( M \times [K] \times N \times [K] \), is taken to be the support of \( MENF \). Then for all \( (m,n) \in \text{Bad} \), the test \( A \) accepts the uniform distribution over \( [K] \times [K] \) with probability 1. This can be a severe limitation on \( A \) if the probability of the set ‘Bad’ is large. Below, we argue that this indeed is the case, by constructing an example where the probability of ‘Bad’ is close to 1. Consider the following inequality, using the fact that for all \( (m,n) \in \text{Good} \), \( 2^{c^2} p_{MN}(m,n) < 1 \) implies \( p_{MN}(m,n) \log \frac{1}{2^{c^2} p_{MN}(m,n)} > 0 \).

\[
0 \leq \sum_{(m,n) \in \text{Good}} p_{MN}(m,n) \log \frac{1}{2^{c^2} p_{MN}(m,n)} \leq H(MN) - c^2 \cdot \text{Pr}_{MN}[\text{Good}],
\]

This inequality implies

\[
\text{Pr}_{MN}[\text{Good}] \leq \frac{H(MN)}{c^2}.
\]

We have the following claim.

Claim IV.1: Fix an \( \epsilon \in (0,1) \) and let \( \alpha \in (0,1) \) be such that \( \alpha(1-\alpha)^2 = 2\epsilon \). There exists random variables \( XYMN \) over \( M \times X \times Y \times N \) with \( M = X = Y = N \) such that \( H(MN) \leq (1-\alpha^3) \log |X| + 3 \) and \( c^2 \geq \log |X| - 3 \). Thus,

\[
\frac{H(MN)}{c^2} \leq 7\sqrt{\epsilon}.
\]

The proof of this claim is given in Appendix D. This implies that the probability \( \text{Pr}_{MN}[\text{Bad}] \) is at least \( 1 - 7\sqrt{\epsilon} \), leading to a large error. Hence, constructing a hypothesis test using the full support of the random variables \( MENF \) does not give the desired result.

Note that if \( N,Y \) are trivial (Task B), then

\[
2^{c^2} p_{M}(m) \leq 1 \text{ for all } m, \text{ which renders the set } \text{Bad} \text{ empty. This shows that the arguments in this subsection do not apply to Task B.}
\]

D. Resolving the Difficulty: A New Hypothesis Testing Method

We will resolve this problem by arguing that one can remove some \( (e,f) \) that occur with low probability, such that the resulting support size of \( (EF|MN = m,n) \) is not too large for every \( m,n \). To lay the groundwork, let us revisit the protocol in [10] for Task B and first obtain a new hypothesis test for this case. We recall the joint random variables \( XMEZ \), where \( XMZ \) satisfy \( M - X - Z \) and \( E \) is uniform over \([K \cdot p_{M|z}(m)]\) conditioned on \( m \) and \( x \). The random variable \( E \), conditioned on \( m, z \), satisfies

\[
p_{E|m,z}(e) = \frac{1}{2^{c^2} K \cdot p_{M|z}(m)} \sum_{x,m,z} p_{X|(m,z)}(x) \geq \frac{1}{2^{c^2} K \cdot p_{M|z}(m)} \sum_{x,m,z} p_{X|(m,z)}(x),
\]

where we have used the definition of \( c \). Now we define a random variable \( G \) correlated with \( X \). Given \( X = x \), \( G \) jointly correlates with \( MZ \) in a manner that it takes the value \( K \cdot p_{M|z}(m) \) with probability \( p_{X|m,z}(x) \) (note that by perturbing the conditional distribution in a negligible manner and choosing large enough \( K \), we can assume that the value \( K \cdot p_{M|z}(m) \) is unique for every \( m, x \)). Using this, the above inequality simplifies to

\[
\frac{p_{E|m,z}(e)}{2^{c^2} K \cdot p_{M|z}(m)} \geq \frac{1}{2^{c^2} K \cdot p_{M|z}(m)} \sum_{g \in \mathcal{G}} p_{G|m,z}(g) = \frac{\text{Pr}_{G \sim \mathcal{G}|m,z}(g \geq e)}{2^{c^2} K \cdot p_{M|z}(m)}. \tag{19}
\]

We now show the following result, which gives a test \( A_{m,z} \) to distinguish the random variable \( (E|MZ = m, z) \) from the uniformly distributed random variable over \([K] \).

Lemma IV.2: Fix \( m, z \) and a \( \delta \in (0,1) \). There exists a set \( A_{m,z} \) such that

\[
\text{Pr}_{E|m,z} [A_{m,z}] \geq 1 - \delta
\]
and
\[ P_{U}[A_{m,z}] = \frac{|A_{m,z}|}{K} \leq \frac{2^c \cdot P_{M|z}(m)}{\delta}, \]
where \( U \) is distributed uniformly in \([K]\).

Proof: Let \( A_{m,z} \) be the set of all \( e \) for which
\[ P_{g \sim G}[g \geq e] \geq \delta. \]

Lemma IV.3 proved towards the end of this subsection, shows that
\[ P_{E|m,z}[A_{m,z}] \geq 1 - \delta, \]
finishing first part of the proof. For the second part, observe from Eq. (19) that for all \( e \in A_{m,z}, \)
\[ p_{E|m,z}(e) \geq \frac{\delta}{2^c \cdot K \cdot P_{M|z}(m)}. \]
Thus,
\[ |A_{m,z}| \leq \frac{2^c \cdot K \cdot P_{M|z}(m)}{\delta}. \]

Since
\[ P_{U}[A_{m,z}] = \frac{|A_{m,z}|}{K}, \]
the proof concludes.

Lemma IV.2 now reproduces Eq. (17), up to a small multiplicative factor of \( \frac{2}{3} \). For this, we will construct a set \( A_{z} \in \mathcal{M} \times [K] \) for all \( z \), such that
\[ P_{M'|z}[A_{z}] \geq 1 - \delta, \quad P_{U'}[A_{z}] \leq \frac{2^c}{|M|} \delta, \]
where \( U' \) is a random variable uniformly distributed over \( \mathcal{M} \times [K] \). The definition of the desired set is
\[ A_{z} \overset{\text{def}}{=} \{(m, e) : e \in A_{m,z}\}. \]
To verify Eq. (20), use Lemma IV.2 to obtain
\[ P_{M'|z}[A_{z}] = \sum_{m} P_{M|z}(m) P_{E|m,z}[A_{m,z}] \geq 1 - \delta \]
and
\[ P_{U'}[A_{z}] = \frac{1}{|M|} \sum_{m} P_{U'}[A_{m,z}] \leq \frac{2^c}{|M|} \delta. \]
The key difference between Equations 20 and 17 is that the former will generalize to the two-party setting, as shown in Theorem IV.4. Before proceeding, let us finish the proof of the following lemma, which was used in Lemma IV.2 and will also be used in Theorem IV.4.

Lemma IV.3: Let \( EG \) be joint random variables taking values over \([K] \times [K] \), such that \( p_{E|g=g}(e) = 0 \) for \( e > g \). Then for any \( \delta \in (0,1) \), it holds that
\[ P_{e \sim E}[g \sim G][g \geq e] \leq \delta. \]

Proof: Let \( g^* \) be the smallest integer such that
\[ P_{g \sim G}[g \geq g^*] \leq \delta. \]
Then for any \( e \geq \delta \)
\[ P_{e \sim E}[g \sim G][g \geq e] \leq \delta. \]

E. Achievable Communication Region for Task C

Here we provide the proof of Theorem III.2. First, we define some new random variables, which will also be used in Theorem IV.4. We assume that \( p_{M|y}(m) \) and \( p_{N|y}(n) \) are distinct rationales for all \( (x, y, m, n) \), which is possible by perturbing the distributions and introducing an arbitrary small error. Let \( K \) be a sufficient large integer such that \( K_{P_{M|y}(m)} \) and \( K_{P_{N|y}(n)} \) are integers for all \( (x, y, m, n) \). We further require that \( \delta \) is also an integer for all \( p \in \left[ \frac{2}{\log 3} \right] \log_{\max}\{[M],[N]\} \),
which can be ensured by noting that \( \sqrt{\pi} \) is a rational (this constraint on \( K \) will be used in Lemma IV.6). We define random variables \( EF \) over \([K] \times [K] \) such that \( E \) is generated conditioned on \( MX \) and \( F \) is generated conditioned on \( NY \) as follows:
\[ p_{E|m,x}(e) = \frac{1}{K_{P_{M|y}(m)}} \]
and
\[ p_{F|n,y}(f) = \frac{1}{K_{P_{N|y}(n)}}. \]

Define \( L, L' \) to be the uniformly distributed and independent random variables over \([K] \). Let
\[ \bar{S} : \text{uniformly distributed over } \mathcal{M}; \]
\[ \bar{T} : \text{uniformly distributed over } N. \]

Finally, define \( S \equiv \bar{S} \times L \) and \( T \equiv \bar{T} \times L' \).

Proof of Theorem III.2: The proof proceeds in the following steps.

Random variables used in the convex-split method: Let
\[ r_{1} \overset{\text{def}}{=} \log \left[ \frac{\mathcal{M}}{2^{r_{1}}} \right]; \quad r_{2} \overset{\text{def}}{=} \log \left[ \frac{\mathcal{N}}{2^{r_{2}}} \right]; \]
\[ R_{3} \overset{\text{def}}{=} \left[ R_{1} + 2 \log 3/\delta \right] \]
and
\[ R_{4} \overset{\text{def}}{=} \left[ R_{2} + 2 \log 3/\delta \right]. \]

Let \( J_{1} \) be a random variable uniformly distributed over \([2^{R_{2}+r_{1}}]\) and the joint random variables \( J_{1}X_{1}S_{1}^{1}S_{2}^{2} \ldots S_{2^{R_{2}+r_{1}}} \) be defined to be:
\[ \Pr\left(X, S_{1}^{1}S_{2}^{2} \ldots S_{2^{R_{2}+r_{1}}} = x, (m_{1}, e_{1}) \ldots (m_{2^{R_{2}+r_{1}}}, e_{2^{R_{2}+r_{1}}}) \right) = \operatorname{Pr}_{J_{1} = j} \]
\[ \times \Pr_{X_{1} \sim X}(x, (m_{1}, e_{1}) \ldots (m_{R_{1}}, e_{R_{1}})) \ldots \Pr_{S_{2^{R_{2}+r_{1}}} \sim S_{2^{R_{2}+r_{1}}}}(m_{2^{R_{2}+r_{1}}}, e_{2^{R_{2}+r_{1}}}) \ldots \]
Similarly, let \( J_2 \) be a random variable uniformly distributed over \([2^{R_4 + r_2}]\) and the joint random variables \( J_2 Y T'_1 T'_2 \ldots T'_{2^R_4 + r_2} \) be defined to be:

\[
\Pr \left[ Y, T'_1 \ldots T'_{2^R_4 + r_2} = (y, (n_1, f_1) \ldots (n_{2^R_4 + r_2}, f_{2^R_4 + r_2}) \right]_{J_2 = j} = \frac{p_{Y|\{X \text{ and } E\}}(y, n_2, f_2) \cdot \cdots \cdot p_T(n_{j-1}, f_{j-1}) \cdot p_T(n_{j+1}, f_{j+1}) \cdots p_T(n_{2^R_4 + r_2}, f_{2^R_4 + r_2})}{p_{J_2 = j}}.
\]

Define the joint random variables \( J_1 J_2 XY Z S'_1 S'_2 \ldots S'_{2^R_4 + r_1} T'_1 T'_2 \ldots T'_{2^R_4 + r_2} \) as

\[
\Pr \left[ J_1 J_2 XY Z S'_1 S'_2 \ldots S'_{2^R_4 + r_1} T'_1 T'_2 \ldots T'_{2^R_4 + r_2} = (j_1, j_2, x, y, z, (m_1, e_1), \ldots, (m_{2^R_4 + r_1}, e_{2^R_4 + r_1}), (n_1, f_1) \ldots (n_{2^R_4 + r_2}, f_{2^R_4 + r_2}) \right] = p_{X|\{Y \text{ and } Z\}}(x|y) p_{X|\{E\}}(x|e) p_{Y|\{X \text{ and } E\}}(y|x, e) \cdots p_T(n_{j-1}, f_{j-1}) \cdots p_T(n_{j+1}, f_{j+1}) \cdots p_T(n_{2^R_4 + r_2}, f_{2^R_4 + r_2}).
\]

Now, observe that

\[
D^*_s(XME|X) = \min \left( \sum_{xyz} p_{X|\{Y \text{ and } Z\}}(x|y) p_{X|\{E\}}(x|e) p_{Y|\{X \text{ and } E\}}(y|x, e) \cdots p_T(n_{j-1}, f_{j-1}) \cdots p_T(n_{j+1}, f_{j+1}) \cdots p_T(n_{2^R_4 + r_2}, f_{2^R_4 + r_2}) \right)
\]

where last equality holds since \( K_{P_{X|\{E\}}(x|m)e}(e) \) is equal to 1 or 0 (by definition of the random variable \( E \)). Similarly, \( D^*_s(YNF|Y \times T) = \log |N| = R_2 + r_2 \). As a result, from the choices of \( R_3 + r_1 \) and \( R_4 + r_2 \) and Fact A.5, we conclude

\[
\frac{1}{2} \left\| X S'_1 \ldots S'_{2^R_4 + r_1} - X \times S \times \ldots \times S \right\|_1 \leq \delta
\]

\[
\frac{1}{2} \left\| Y T'_1 \ldots T'_{2^R_4 + r_2} - Y \times T \times \ldots \times T \right\|_1 \leq \delta.
\]

(24)

Consider the following inequality:

\[
\frac{1}{2} \left\| \left( S'_1 \ldots S'_{2^R_4 + r_1} | X = x \right) \times \left( T'_1 \ldots T'_{2^R_4 + r_2} | Y = y \right) - S \times \ldots \times S \times T \times \ldots \times T \right\|_1
\]

\[
= \frac{1}{2} \sum_{xyz} p_{X|\{Y \text{ and } Z\}}(x|y) p_{X|\{E\}}(x|e) p_{Y|\{X \text{ and } E\}}(y|x, e) \cdots p_T(n_{j-1}, f_{j-1}) \cdots p_T(n_{j+1}, f_{j+1}) \cdots p_T(n_{2^R_4 + r_2}, f_{2^R_4 + r_2})
\]

\[
\left\| \left( S'_1 \ldots S'_{2^R_4 + r_1} | X = x \right) \times \left( T'_1 \ldots T'_{2^R_4 + r_2} | Y = y \right) - S \times \ldots \times S \times T \times \ldots \times T \right\|_1
\]

\[
\leq \frac{1}{2} \left\| X S'_1 \ldots S'_{2^R_4 + r_1} - X \times S \times \ldots \times S \right\|_1 \leq \delta
\]

\[
\frac{1}{2} \left\| Y T'_1 \ldots T'_{2^R_4 + r_2} - Y \times T \times \ldots \times T \right\|_1 \leq \delta.
\]

(25)

Above, the first equality uses the fact that \( S'_1 S'_2 \ldots S'_{2^R_4 + r_1} J_1 \) and \( T'_1 T'_2 \ldots T'_{2^R_4 + r_2} J_2 \) are independent conditioned on \( XY = xy \) for any \( (x, y) \) in the support of \( XY \). The last inequality uses Eq. (24).
Protocol description: We are now ready to define the protocol (see Figure 3).

Input: Random variables $XYMN$ distributed over $X \times Y \times M \times N$, where $X, Y, M$ and $N$ are finite sets; reals $R_1, R_2, \epsilon, \delta$ satisfying Theorem III.2; $R_3, R_4$ as defined in Eq. (23). Alice, Bob and Charlie are given $x, y$ and $z$, respectively, where $(x, y, z) \sim XYZ$.

Shared resources: Alice and Charlie share $S_1 \ldots S_{2R_3+1}$, which are $2^{R_3+r_1}$ copies of i.i.d. samples of $S$. Bob and Charlie share $T_1 \ldots T_{2R_4+r_2}$, which are $2^{R_4+r_2}$ copies of i.i.d. samples of $T$. Here $S$ and $T$ are defined in Eq. (22).

The protocol:
1) Alice observes a sample $(x, (m_1, e_1), \ldots, (m_{2R_3+1}, e_{2R_3+1}))$ from $XS_1 \ldots S_{2R_3+1}$ and samples $j_1$ from $\begin{pmatrix} X, S_{1}', \ldots, S_{2R_3+1}' \end{pmatrix} = \begin{pmatrix} x, (m_1, e_1), \ldots, (m_{2R_3+1}, e_{2R_3+1}) \end{pmatrix}$.
2) Bob observes a sample $(y, (n_1, f_1), \ldots, (n_{2R_4+2}, f_{2R_4+2}))$ from $YT_1 \ldots T_{2R_4+2}$ and samples $j_2$ from $\begin{pmatrix} Y, T_{1}', \ldots, T_{2R_4+2}' \end{pmatrix} = \begin{pmatrix} y, (n_1, f_1), \ldots, (n_{2R_4+2}, f_{2R_4+2}) \end{pmatrix}$.
3) Alice sends $j_1 \overset{\text{def}}{=} \lceil j_1 \rceil$ to Charlie.
4) Bob sends $j_2 \overset{\text{def}}{=} \lfloor j_2 \rfloor$ to Charlie.
5) Let $\bar{A}$ be the set specified in Theorem IV.4. For every pair $(\bar{j}_1, \bar{j}_2) \in \{(j_1' - 1) \cdot 2^{r_1} + 1, \ldots, (j_1' - 1) \cdot 2^{r_1} + 2^{r_1}\}$ $\times \{(j_2' - 1) \cdot 2^{r_2} + 1, \ldots, (j_2' - 1) \cdot 2^{r_2} + 2^{r_2}\}$, let $A_{\bar{j}_1, \bar{j}_2}$ be the test that accepts a sample $(m_{j_1}, e_{j_1}, n_{j_2}, f_{j_2}, z)$ drawn from $S_{j_1}T_{j_2}Z$, if it belongs to $A$. Charlie performs the tests $A_{\bar{j}_1, \bar{j}_2}$ in the lexicographical order and outputs a sample from $S_{\bar{j}_1}T_{\bar{j}_2}$ for the first pair $(\bar{j}_1, \bar{j}_2)$ where the test succeeds. If no such index exists, Charlie outputs a sample from $S\bar{T}$. Let the output of Charlie be the random variables $M'N'$.

From Eq. (23) the communication cost between Alice and Charlie is $R_1 + 2 \log \frac{3}{\delta} \leq R_1 + 3 \log \frac{1}{\delta}$ and the communication cost between Bob and Charlie is $R_2 + 2 \log \frac{3}{\delta} \leq R_2 + 3 \log \frac{1}{\delta}$. Here, we used $\delta \leq \frac{1}{2}$.

Error analysis: Let $XYZMN''$ be the final random variables if the above protocol was run on $XYZS_1' \ldots S_{2R_3+1}' T_1' \ldots T_{2R_4+2}'$. By the virtue of Eq. (25), which states that the random variables $XYZS_1 \ldots S_{2R_3+1} T_1 \ldots T_{2R_4+2}$ is close to $XYZS_1' \ldots S_{2R_3+1}' T_1' \ldots T_{2R_4+2}'$, and Fact A.3, it holds that

$$\frac{1}{2} \|XYZM''N'' - XYNM'N'\|_1 \leq 2\delta. \quad (26)$$

Next, we bound the error incurred in Step 5 of the protocol, assuming that the protocol was run on $XYZS_1' \ldots S_{2R_3+1}' T_1' \ldots T_{2R_4+2}'$. Our argument will proceed conditioned on the event that the messages from Alice and Bob are $j_1', j_2'$. Let $\tilde{J}_1, \tilde{J}_2$ be the random variables obtained by Charlie in Step 5. We invoke Lemma A.7, viewing $C$ as the random variables $J_1J_2$, $C'$ as the random variable $J_1J_2$ and $H$ as the random variables $XYZS_1' \ldots S_{2R_3+1}' T_1' \ldots T_{2R_4+2}'$.

From Theorem IV.4, for every $(\tilde{j}_1, \tilde{j}_2), (k_1, k_2) \in \{(j_1' - 1) \cdot 2^{r_1} + 1, \ldots, (j_1' - 1) \cdot 2^{r_1} + 2^{r_1}\}$ $\times \{(j_2' - 1) \cdot 2^{r_2} + 1, \ldots, (j_2' - 1) \cdot 2^{r_2} + 2^{r_2}\}$,

$$\Pr_{S \times N F Z} \left[ A_{\tilde{j}_1, \tilde{j}_2} \right] \leq \frac{\delta^2 R_3}{|M|},$$

and

$$\Pr_{S \times T} \left[ A \right] \leq \frac{\delta^2 R_1 + R_2}{|M||N|}.$$

For every $(\tilde{j}_1, \tilde{j}_2)$, the number of pairs $(\tilde{j}_1, \tilde{j}_2)$ with $k_2 \neq \tilde{j}_2$ is less than $2^{r_2}$; the number of pairs $(\tilde{j}_1, k_2)$ with $k_1 \neq \tilde{j}_1$ is less than $2^{r_1}$; and the number of pairs $(k_1, k_2)$ with $k_1 \neq \tilde{j}_1, k_2 \neq \tilde{j}_2$ is less than $2^{r_1 + r_2}$. Lemma A.7 thus implies that

$$\frac{1}{2} \|XYZS_1 \ldots S_{2R_3+1} T_1 \ldots T_{2R_4+2} J_1J_2 - \|XYZS_1 \ldots S_{2R_3+1} T_1 \ldots T_{2R_4+2} \|_1 \leq \epsilon + 5\delta + 2^{r_1 + r_2} \frac{\delta^2 R_1}{|M|} + 2^{r_1 + r_2} \frac{\delta^2 R_2}{|N|} + 2^{r_1 + r_2} \frac{\delta^2 R_1 + R_2}{|M||N|} \leq \epsilon + 8\delta.$$

The last inequality uses Eq. (23). Given the sample $(\tilde{j}_1, \tilde{j}_2)$ from $J_1J_2$, the random variables $S_{\tilde{j}_1} T_{\tilde{j}_2}$ is correlated with $XYZ$. Since Charlie outputs $S_{\tilde{j}_1} T_{\tilde{j}_2}$ given $(\tilde{j}_1, \tilde{j}_2)$ drawn from $\tilde{J}_1 \tilde{J}_2$, we can apply Fact A.3, and conclude

$$\frac{1}{2} \|XYZMN - XYZMN''\|_1 \leq \epsilon + 8\delta + 2\delta \leq \epsilon + 10\delta.$$
Theorem IV.4, to be shown below, gives error bounds on the hypothesis testing part employed by Charlie in the proof of Theorem III.2. It is a bi-variate generalization of Equation 20.

**Theorem IV.4:** Let \( \epsilon \in (0, 1) \) be as given in Theorem III.2 and \( \delta \in (0, \frac{1}{4}) \) satisfy that \( \frac{1}{\delta} \) is an integer greater than 1. Let \( R_1, R_2 \) be as chosen in the statement of Theorem III.2 Equation (3). Let \( E, F \) be random variables, derived from \( XYZMN \), as defined in Eq. (21). Then there exists a set \( \mathcal{A} \subseteq M \times N \times E \times F \times Z \) such that

\[
\Pr_{\mathcal{A} \times \mathcal{T} X T Z}[A] \leq \frac{2^R_2 |M|}{|N|},
\]

and

\[
\Pr_{\mathcal{A} \times \mathcal{S} \times \mathcal{T} \times \mathcal{Z}}[A] \leq \frac{2^{R_1 + R_2} |M|}{|N|}.
\]

**Proof:** We set \( \text{dev} \triangleq \log \max\{\frac{|M|}{|N|}\} \). For any \( (m, n) \in M \times N \), we define \( w_m(x) \triangleq \frac{K_p M|x|}{(m)} \) and \( v_n(y) \triangleq \frac{K_p N|y|}{(n)} \). From our assumption on \( K \), \( w_m(\cdot) \) and \( v_n(\cdot) \) are both integer-valued functions for all \( m, n \). Further define

\[
\text{Good}_1^{m, n, z} \triangleq \{ (x, y) : \frac{p_{M|x|}(m)}{p_{M|z|}(m, n, z)} \leq \delta \cdot 2^{R_1}, \frac{p_{N|y|}(n)}{p_{N|m|}(m, n)} \leq \delta \cdot 2^{R_2} \text{ and } \frac{p_{M|z|}(m, n, z)}{p_{MN}(m, n, z)} \leq \delta \cdot \text{dev} \cdot 2^{R_1 + R_2} \},
\]

and

\[
\text{Good}_2^{m, n, z} \triangleq \{ (x, y) : p_{M|x|}(m) \geq \delta \frac{|M|}{|N|} \text{ and } p_{N|y|}(n) \geq \delta \frac{|N|}{|M|} \}
\]

and

\[
\text{Good}_{m, n, z} \triangleq \text{Good}_1^{m, n, z} \cap \text{Good}_2^{m, n, z}.
\]

We define the new random variables \( X'Y'E'F'MNZ \) obtained by restricting \( XY \) to \( \text{Good}_{m, n, z} \). Conditioned on \( MNZ = (m, n, z) \), namely,

\[
p_{X'Y'E'F'MNZ}(x, y, e, f) \triangleq \begin{cases} p_{XY}(x, y) & \text{if } (x, y) \in \text{Good}_{m, n, z} \text{ and } e \leq w_m(x) \text{ and } f \leq v_n(y) \\ 0 & \text{otherwise.} \end{cases}
\]

From Fact A.2 it holds for any \( m, n, z \) that

\[
\left\| (E'F'X'Y'|MNZ = m, n, z) - (EFXY|MNZ = m, n, z) \right\|_1 \leq 2 \left( 1 - \Pr_{XY|m, n, z}[\text{Good}_{m, n, z}] \right),
\]

where the first equality is from the fact that \( E \) and \( F \) are determined by \( MX \) and \( NY \) due to Eq. (21), respectively; the second equality is from the definition of \( X'Y' \). We have the following claim, shown in Subsection IV-F

**Claim IV.5:**

\[
\sum_{m, n, z} p_{MNZ}(m, n, z) \Pr_{XY|m, n, z}[\text{Good}_{m, n, z}] \geq 1 - \epsilon - 2\delta.
\]

The following Lemma is analogous to Lemma IV.2 and is proved towards the end.

**Lemma IV.6:** Fix some \( m, n, z \). There exists a set \( A_{m, n, z} \subseteq [K] \times [K] \) such that

\[
\Pr_{EF|m, n, z}[A_{m, n, z}] \geq \Pr_{EF|m, n, z}[\text{Good}_{m, n, z}] - 3\delta.
\]

From Eq. (28),

\[
\Pr_{MNZEF}[A] \geq \sum_{m, n, z} p_{MNZ}(m, n, z) \Pr_{XY|m, n, z}[\text{Good}_{m, n, z}] - 3\delta
\]

\[
\geq 1 - \epsilon - 5\delta,
\]

where the last inequality follows from Claim IV.5. Thus, we conclude the first inequality in Theorem IV.4. For the second inequality in Theorem IV.4, use Eq. (29) and consider

\[
\Pr_{MNZEF}[A] = \sum_{m, n, z} p_{MNZ}(m, n, z) \Pr_{EF|m, n, z}[A_{m, n, z}]
\]

\[
\leq \sum_{m, n, z} p_{MNZ}(m, n, z) \frac{1}{|N|} \delta_2^{R_2} p_{N|m|}(m, n) \leq \delta_2^{R_2} \frac{|M|}{|N|}.
\]

Similarly, using Eq. (30),

\[
\Pr_{MNZEF}[A] \leq \frac{\delta_2^{R_1}}{|M|}.
\]

For the last inequality, we apply Eq. 31 to conclude,

\[
\frac{1}{|N|} \sum_{m, n, z} p_{Z}(z) \Pr_{LX|L}[A_{m, n, z}] \leq \sum_{m, n, z} p_{Z}(z) \frac{\delta_2^{R_2} p_{MNZ}(m, n)}{|N|} \leq \delta_2^{R_2} \frac{|M|}{|N|}.
\]

This completes the proof of the theorem. \( \square \)
as follows.

\[
p_{\text{WV}|x,y,e,f,m,n,z}(w,v) \overset{\text{def}}{=} \begin{cases} 
1 & \text{if } w = w_m(x) \land v = v_n(y) \\
0 & \text{otherwise}.
\end{cases}
\]

They are analogous to the random variable \(G\) defined in Subsection IV-D. Now, we sketch a proof outline for Lemma IV.6 before moving to full details.

**Proof outline:** Fix the values \(m, n, z\). We will drop conditioning on \(m, n, z\) from the random variables in this proof outline, for brevity of notation. Recall that \(E\mid (W = w)\) and \(F\mid (V = v)\) are uniform in \([w]\) and \([v]\) respectively. Observe, from the definition of the set \(\text{Good}_{m,n,z}^2\) that for all \(w, v\) drawn from \(\text{WV}\), \(w > \frac{K\delta}{|M|}\) and \(v > \frac{K\delta}{|N|}\). Finally, recall that \(wv = K^2 p_{M|a}(m)p_{N|y}(n)\) for some \(x, y, z\), which implies from the definition of the set \(\text{Good}_{m,n,z}^1\) that \(wv \leq \frac{K\delta^2 R_1 + R_2}{\text{dev}} p_{M|a}(m, n)\).

Our approach is similar to that in Lemma IV.2. Along the lines of Eq. (19), we expand

\[
p_{EF}(e, f) = \sum_{w, v} p_{\text{WV}}(w, v)p_{E|w}(e)p_{F|v}(f) = \sum_{w, v; w > e \land v > f} p_{\text{WV}}(w, v) \frac{1}{wv} \geq K^2 \delta^2 R_1 + R_2 p_{M|a}(m, n) \sum_{w > e \land v > f} p_{\text{WV}}(w, v).
\]

Thus, we can focus on showing the following claim: with high probability according to \(E\), the quantity

\[
P_{e,f} \overset{\text{def}}{=} \sum_{w, v; w > e \land v > f} p_{\text{WV}}(w, v)
\]

is large. Let us start with a special case where \(\text{WV}\) is completely supported in the region \([a^0 ; a^1] \times [b^0 ; b^1]\) (Figure 4). Further, assume that this region has small size with respect to \([a^1] \times [b^1]\), in the sense that it satisfies \((a^1 - a^0)(b^1 - b^0) \leq \delta\). The support of \(E\) lies in \([a^1] \times [b^1]\), since every \(e, f, w, v\) drawn from \(\text{EFWV}\) are constrained by \(e < w\) and \(f < v\). We divide \([a^1] \times [b^1]\) into four regions, as shown in Figure 4. Let us consider \(P_{e,f}\) in each region. In region 3, \(P_{e,f}\) is equal to 1 since every \(w, v\) satisfies \(w > e\) and \(v > f\). Region 2 has the property that \(e < w\) always, which implies that \(P_{e,f} = \sum_{v} p_{\text{WV}}(v)\). This can be lower bounded using Lemma IV.3. Same argument applies for region 4. The remaining issue is to bound \(P_{e,f}\) in Region 1. But we do not need to consider this case, as the choice of \(a^0, a^1, b^0\) and \(b^1\) ensures that the probability that \(E\) lies in region 1 is at most \(\delta\). Collectively, we are able to show that \(P_{e,f}\) is large, with high probability according to \((e, f)\) drawn from \(EF\). \(A_{m,n,z}\) is then chosen to be the set of all \(e, f\) for which \(P_{e,f}\) has just shown to be large.

In general, \(\text{WV}\) will have support all over the square \([K] \times [K]\). But, we can divide the square \([K] \times [K]\) into several smaller squares such that for any square \([a_2 : a_2] \times [b_2 : b_2]\), it holds that \(\frac{(a_2 - a_1)(b_2 - b_1)}{a_2 b_2} \leq \delta\). This leads to a recursive decomposition given in Figure 5. The number of squares is less than \(O(\log \max\{|M|, |N|\})\), since we have already ensured that \(w \geq \frac{\delta K}{|M|}\) and \(v \geq \frac{\delta K}{|N|}\) (see the definition of the set \(\text{Good}_{m,n,z}^2\)). By the argument outlined in previous paragraph, each such square gives a set that contains a large probability of \(E\). By taking a union over all these sets, we obtain the desired result, with a loss of \(\log \log \max\{|M|, |N|\}\) (due to the union bound) that is reflected in the statement of Theorem III.2.
Proof of Lemma IV.6: Constructing a partition of $[K] \times [K]$; The idea behind the construction is depicted in Figure 5. Let $\delta_1 \equiv \sqrt{\delta}$ and $\epsilon \equiv \frac{\delta}{\log \max\{|M|,|N|\}}$. We may assume that $\delta_1 K$ is an integer. The following set of $\frac{\delta}{\epsilon}$ squares partition the set $[K] \times [K]$:

$$
\left\{(p-1)\delta K + 1: (q-1)\delta K + 1\right\}_{p,q \in [\frac{\delta}{\epsilon}]}.
$$

Let these squares be denoted as $\{C_j\}_{j=1}^{\frac{\delta}{\epsilon}}$ in some order. Let $\alpha C_j$ denote the set $\{(w,v): (\frac{w}{\alpha}, \frac{v}{\alpha}) \in C_j\}$. For all $i \in \{0,\ldots,6\}$, we define

$$
S_i \equiv \left[\delta_i K\times[\delta_i K]\right], T_i \equiv S_i \setminus S_{i+1}.
$$

with $T_e$ being undefined. For $i \in \{0,1,\ldots,6\}$ and $j \in \left[\frac{\delta}{\epsilon}\right]$, we define

$$
T_{i,j} \equiv T_i \cap \alpha C_j.
$$

Then $\{T_{i,j}\}_{i \in \{0,\ldots,6\}, j \in \left[\frac{\delta}{\epsilon}\right]}$ are disjoint sets covering $[K] \times [K] \setminus [S_e]$ and $\{\alpha C_j\}_{j=1}^{\frac{\delta}{\epsilon}}$ equally divide $S_i$ into $\frac{\delta}{\epsilon}$ squares. Thus, $T_{i,j}$ is either a $\delta_1 \delta K \times \delta_1 \delta K$ square or empty. For any $i$, $T_{i,j}$ is non-empty for $\frac{1-\delta_i^2}{\alpha}$ many $j$’s. We relabel the indices $j$’s such that $T_{i,j}$ is non-empty if and only if $1 \leq j \leq \frac{1-\delta_i^2}{\alpha}$. Set

$$
T_{i,j} \equiv \left\{a_{0,j}^i, a_{0,j}^i + 1, \ldots, a_{0,j}^i + 1\right\} \times \left\{b_{0,j}^i, b_{0,j}^i + 1, \ldots, b_{1,j}^i\right\}.
$$

Thus, $T_{i,j}$ is $\delta_1 \delta K \times \delta_1 \delta K$ square and or empty. For any $i, T_{i,j}$ is non-empty for $\frac{1-\delta_i^2}{\alpha}$ many $j$’s. We relabel the indices $j$’s such that $T_{i,j}$ is non-empty if and only if $1 \leq j \leq \frac{1-\delta_i^2}{\alpha}$. Set

$$
T_{i,j} \equiv \left\{a_{0,j}^i, a_{0,j}^i + 1, \ldots, a_{0,j}^i + 1\right\} \times \left\{b_{0,j}^i, b_{0,j}^i + 1, \ldots, b_{1,j}^i\right\}.
$$

Claim IV.8: For any $i, j, m, n, z$, it holds that

$$
\sum_{(e,f) \in \mathcal{A}_{i,j,m,n,z}} p_{E|F|m,n,z,T_{i,j}}(e,f) \leq 2\delta.
$$

Using Claims IV.7, IV.8 and the definition of the set $\mathcal{A}_{i,j,m,n,z}$, we can apply union bound to conclude

$$
\sum_{(e,f) \in \mathcal{A}_{i,j,m,n,z}} p_{E|F|m,n,z,T_{i,j}}(e,f) \geq 1 - 3\delta.
$$

Moreover, for all $(e,f) \in \mathcal{A}_{i,j,m,n,z}$, which ensures that $(e,f) \notin \mathcal{B}_{i,j,m,n,z}$, we have

$$
\sum_{(e,f) \in \mathcal{A}_{i,j,m,n,z}} p_{E|F|m,n,z,T_{i,j}}(e,f) \leq 2\delta.
$$

Let $\mathcal{B}_{i,j,m,n,z}$ refers to the collection of points $(e,f)$ in Regions 2 and 4, where $\sum_{(e,f) \in \mathcal{B}_{i,j,m,n,z}} p_{E|F|m,n,z,T_{i,j}}(e,f)$ has small value.

The second claim upper bounds the errors occurred in Regions 2 and 4 in Figure 4.
Furthermore, Eq. (32) implies
\[
\Pr_{E^F|m,n,z}[A_{m,n,z}^{(1)}] = \sum_{i,j} \Pr_{WV|m,n,z}[T_{i,j}] \Pr_{E^F|m,n,z;T_{i,j}}[A_{m,n,z}^{(1)}] \\
\geq \sum_{i,j} \Pr_{WV|m,n,z}[T_{i,j}] \Pr_{E^F|m,n,z;T_{i,j}}[A_{i,j,m,n,z}] \\
\geq 1 - 3\delta.
\]
(36)
Note that for any \(m, n, z\)
\[
\Pr_{E^F|m,n,z}[A_{m,n,z}^{(2)}] = \Pr_{E^F|m,n,z}[A_{m,n,z}^{(3)}] = 1,
\]
which implies using Eq. (36) that
\[
\Pr_{E^F|m,n,z}[A_{m,n,z}] = \Pr_{E^F|m,n,z}[A_{m,n,z}^{(1)} \cap A_{m,n,z}^{(2)} \cap A_{m,n,z}^{(3)}] \\
\geq 1 - 3\delta.
\]
Combining with Eq. (27), we obtain Eq. (28):
\[
\Pr_{E^F|m,n,z}[A_{m,n,z}] \geq \Pr_{XY|m,n,z}[\text{Good}_{m,n,z}] - 3\delta.
\]
Next, to show Eq. (29), we proceed as follows.
\[
\Pr_{(E|m,z) \times L}[A_{m,n,z}] = \Pr_{(E|m,z) \times L}[A_{m,n,z}^{(1)} \cap A_{m,n,z}^{(2)} \cap A_{m,n,z}^{(3)}] \\
\leq \Pr_{(E|m,z) \times L}[A_{m,n,z}^{(3)}] \\
= \sum_{(e,f): f \leq \delta K^2 R_2 P_{N|z}(n)} \mathbb{1}[P_{E|m,z}(e)] \\
= \sum_{e} \delta^2 K^2 R_2 P_{N|z}(n) \Pr_{E|m,z}(e) \\
= \delta^2 K^2 R_2 P_{N|z}(n).
\]
Eq. (30) is obtained in a similar manner. Finally using Eq. (35), we show Eq. (31).
\[
\Pr_{L \times L}[A_{m,n,z}] = \frac{|A_{m,n,z}|}{K^2} \\
\leq \frac{|A_{m,n,z}^{(1)}|}{K^2} \leq \delta^2 R_1 + R_2 P_{MN|z}(m,n)
\]
This completes the proof.

F. Proof of Claims Used in Theorem IV.4 and Lemma IV.6
Proof of Claim IV.5: From the choice of \(\varepsilon\) in Theorem III.2, we have
\[
\sum_{m,n,z} p_{MNZ}(m,n,z) \Pr_{XY|m,n,z}[\text{Good}_{m,n,z}] \\
\geq 1 - \sum_{m,n,z,x,y} \left( p_{MNZXY}(m,n,z,x,y) \cdot \mathbb{1}(x,y) \in \text{Good}_{m,n,z} \right) \\
\geq 1 - \sum_{m,n,z,x,y} \left( p_{MNZXY}(m,n,z,x,y) \cdot \mathbb{1}(x,y) \notin \text{Good}_{m,n,z} \right) \\
= 1 - \varepsilon - \sum_{x,y,z} p_{XYZ}(x,y,z) \sum_{m,n} p_{MN|z}(m) p_{NM|y}(n) \\
\cdot \mathbb{1}(p_{MN|z}(m) \leq \delta) \left( \frac{\delta}{|M|} \right) \\
\geq 1 - \varepsilon - \left( \sum_{x,y,z} p_{XYZ}(x,y,z) \sum_{m,n} p_{MN|z}(m) p_{NM|y}(n) \mathbb{1}(p_{MN|z}(m) \leq \delta) \right) \\
\geq 1 - \varepsilon - 2\delta.
\]
Proof of Claim IV.7: Note that
\[
\sum_{(e,f) \in T_{i,j}} p_{E^F|m,n,z;T_{i,j}}(e,f) \\
= \sum_{(w,v) \in T_{i,j}} (w,v) \cdot p_{WV|m,n,z;T_{i,j}}(w,v) \\
= \sum_{(w,v) \in T_{i,j}} \left( p_{WV|m,n,z;T_{i,j}}(w,v) \cdot \sum_{(e,f) \in T_{i,j}} \mathbb{1}(w,v) \in (w,v) \cdot (w,v) \cdot (w,v) \right) \\
\leq \sum_{(w,v) \in T_{i,j}} \left( \frac{p_{WV|m,n,z;T_{i,j}}(w,v) \cdot (w,v) \cdot (w,v)}{\delta K^2} \right) \\
\leq \frac{\delta^2 K^2}{\delta^2 K^2} = \delta,
\]
where the first inequality follows from the definition of \(T_{i,j}\).
Proof of Claim IV.8: Let \(w^*, v^*\) be the smallest integers such that
\[
\sum_{(w,v) \in T_{i,j}: w \geq w^*} p_{WV|m,n,z;T_{i,j}}(w,v) \leq \delta,
\]
and
\[
\sum_{(w,v) \in T_{i,j}: v \geq v^*} p_{WV|m,n,z;T_{i,j}}(w,v) \leq \delta.
\]
We claim that for any \((e,f) \in \text{Bad}_{i,j,m,n,z}\), either \(e \geq w^*\) or \(f \geq v^*\). Suppose by contradiction that there exists \((e,f) \in \text{Bad}_{i,j,m,n,z}\) such that \(e \leq w^*\) and \(f \leq v^*\). As \((e,f) \notin T_{i,j}\), either \(e \leq a_{i,j}^0\) or \(f \leq b_{i,j}^0\). Suppose \(e \leq a_{i,j}^0\) (corresponding to Region 2 in Figure 4); the other case follows similarly. Then
\[
\sum_{(w,v) \in T_{i,j}: w \geq e \ \& \ v \geq f} p_{WV|m,n,z;T_{i,j}}(w,v) \\
= \sum_{(w,v) \in T_{i,j}: w \geq e \ \& \ v \geq f} p_{WV|m,n,z;T_{i,j}}(w,v) > \delta,
\]
where the equality is from the fact that \(\text{supp}(WV|MNZ = m,n,z,T_{i,j}) \subseteq \{a_{i,j}^0, \ldots, a_{i,j}^1\} \times \{b_{i,j}^0, \ldots, b_{i,j}^1\}\) which ensures that every \(w\) is larger than \(e\); the inequality
follows from the definition of $f^*$. Therefore, we have
\[
\sum_{(e,f) \in \mathcal{H}_{x,y,m,n,z}} p_{E'|F'}|m,n,z,T_{i,j} (e,f) \leq \left( \sum_{(e,f) : e \geq \omega} + \sum_{(e,f) : f \geq \omega} \right) p_{E'|F'}|m,n,z,T_{i,j} (e,f).
\]
We upper bound the first summation on the right hand side, following Lemma IV.3.
\[
\sum_{(e,f) : e \geq \omega} \sum_{(w,v) \in T_{i,j}, w \geq e} p_{WV}|m,n,z,T_{i,j} (w,v) \cdot \frac{1}{wv}
\]
\[
= \sum_{(w,v) \in T_{i,j}, w \geq \omega} \sum_{(e,f) : f \geq \omega} p_{WV}|m,n,z,T_{i,j} (w,v) \cdot \frac{1}{wv} \leq \sum_{(w,v) \in T_{i,j}, w \geq \omega} p_{WV}|m,n,z,T_{i,j} (w,v) \leq \delta.
\]
Similarly,
\[
\sum_{(e,f) : f \geq \omega} \sum_{(w,v) \in T_{i,j}, w \geq \omega} p_{WV}|m,n,z,T_{i,j} (w,v) \cdot \frac{1}{wv} \leq \delta.
\]
Thus, we conclude the claim.

V. LOWER BOUND ON THE EXPECTED COMMUNICATION COST OF ONE-WAY PROTOCOLS FOR SLEPIAN-WOLF TASK

In this section, we revisit the expected communication cost of Slepian-Wolf task, which is a special case of Task B where $M = X$ (originally studied by Slepian and Wolf [2]). The protocol in [9] for Task B implies that there is an interactive protocol achieving the expected communication $H(X|Z) + c \left( \sqrt{H(X|Z)} + \log \frac{1}{N} \right)$, for some constant $c$ independent of $|X|$, $|Z|$. The following theorem shows that the interaction is necessary, by proving a much larger lower bound for one-way protocols.

**Theorem VI.1:** For any integer $N$ and $\epsilon \in (0, \frac{1}{4})$, there exists joint random variables $XZ$ with support $\left[ \{1 - \sqrt{\epsilon} \} N \right] \times [N]$ such that the expected communication cost of any one-way protocol achieving Task B with $M = X$ and error $\epsilon$ is at least
\[
\frac{1}{12\sqrt{\epsilon}} H(X|Z).
\]

**Proof:** Let $\delta \overset{\text{def}}{=} \sqrt{\epsilon}$. We define the random variables $XZ$ as follows. Let $Z$ be a random variable uniformly distributed over $[N]$. For any $z \in [\delta N]$, let $(X|Z = z)$ be the random variable uniformly distributed in $\left[ \{1 - \delta \} N \right]$. For any $z \in \{\delta N + 1, \ldots, N\}$, let $(X|Z = z)$ be distributed as $p_{X|z}(x) = 1 (x = z - \delta N)$. Observe that $p_{X}(x) = \delta (1 - \delta) N + \frac{1}{N} = \frac{1}{(1 - \delta) N}$. Thus $X$ is uniform over $\left[ \{1 - \delta \} N \right]$. Furthermore, $H(X|Z) = \sum_{z = 1}^{\delta N} \frac{1}{N} H(X) = \delta \log (1 - \delta) \leq \delta \log N$.

A protocol $\mathcal{P}$ achieving Task $B$ is as follows. Let $R$ be the randomness pre-shared between Alice and Charlie. Conditioning on her input $x$ and the randomness $r$ drawn from $R$, Alice sends a message of length $\ell_{x,r}$ to Charlie. Conditioned on his input $z$, the message from Alice and the randomness $r$, Charlie outputs some $x' \in \{(1 - \sqrt{\epsilon}) N\}$. We denote this random variable by $X'$, which takes a fixed value conditioned on $XZR$. Let $\epsilon_{x,z,r} = 1$ if $x' \neq x$ for a given $x, z, r$ and 0 otherwise. Let $C$ be the expected communication cost of the protocol, that is,
\[
C = \sum_{x,r} p_{X}(x) p_{R}(r) \ell_{x,r}.
\]

Since we have set $M = X$ in Task B, the error is
\[
\epsilon = \frac{1}{2} H(X|X') = \frac{1}{2} \sum_{x,z} |p_{XZ}(z) - p_{XX'}(x, x', z)|
\]
\[
\leq \frac{1}{2} \sum_{x,z} |p_{XZ}(z) - p_{XZ}(x,x',z)p_{X'|x,x,z} (x')|
\]
\[
= \frac{1}{2} \sum_{x,z} \left( p_{XZ}(z) \left( \sum_{x' \neq x} p_{X'|x,x,z} (x') \right) - p_{XZ}(x) p_{R}(r) \left( \sum_{x' \neq x} p_{X'|x,z,r} (x') \right) \right)
\]
\[
= \frac{1}{2} \sum_{x,z} p_{XZ}(z) p_{R}(r) \epsilon_{x,z,r}.
\]

Applying Markov’s inequality to Eq. (37) and Eq. (38), we obtain
\[
\Pr_{r \sim R} \left[ \sum_{x} p_{X}(x) \ell_{x,r} \leq 3C \right] \geq \frac{2}{3}
\]
and
\[
\Pr_{r \sim R} \left[ \sum_{x,z} p_{XZ}(z) \epsilon_{x,z,r} \leq 3 \epsilon \right] \geq \frac{2}{3}.
\]
Thus, there exists some $r_0$ such that
\[
\sum_{x} p_{X}(x) \ell_{x,r_0} \leq 3C, \quad \sum_{x,z} p_{XZ}(z) \epsilon_{x,z,r_0} \leq 3 \epsilon.
\]
Since $Z$ is uniform in $[N]$,
\[
3 \epsilon \geq \sum_{x} p_{Z}(z) \sum_{x} p_{XZ}(x) \epsilon_{x,z,r_0}
\]
\[
\geq \frac{1}{N} \sum_{z = 1}^{\delta N} \sum_{x} p_{XZ}(x) \epsilon_{x,z,r_0}.
\]
Thus, there exists $z_0 \in [\delta N]$ such that $\sum_{x} p_{XZ}(x) \epsilon_{x,z_0,r_0} \leq \frac{3 \epsilon}{N}$. Note that $p_{XZ}(x) = p_{X}(x)$ for $z \in [\delta N]$. We thus
conclude that
\[ \sum_x p_X(x) \epsilon_{x,z_0,r_0} \leq \frac{3e}{\delta} = 3\sqrt{\epsilon}, \sum_x p_X(x) \epsilon_{x,r_0} \leq 3C. \tag{39} \]

Now, recall that \( p_X(x) \) is uniform in \( [(1 - \delta)N] \) and \( \epsilon_{x,z_0,r_0} \) is either 0 or 1. Applying Markov’s inequality to Eq. (39), we conclude that for at most \( \frac{1}{8} \) fraction of \( x \in [(1 - \delta)N] \), \( \epsilon_{x,r_0} \leq 6C \) and for at most \( 3\sqrt{\epsilon} \leq \frac{3}{\delta} \) fraction of \( x \in [(1 - \delta)N] \), \( \epsilon_{x,z_0,r_0} = 0 \). Hence by union bound, for at least \( \frac{1}{2} \) fraction of \( x \in [(1 - \delta)N] \), we have \( \epsilon_{x,r_0} \leq 6C \) and \( \epsilon_{x,z_0,r_0} = 0 \). This implies that in the protocol \( \mathcal{P} \), Alice communicates a different message for each such \( x \). This requires
\[ 6C \geq \log \left( \frac{1-\delta}{8N} \right) \geq \frac{1}{2} \log N \geq \frac{1}{2}\sqrt{\epsilon} H(X|Z). \]

This completes the proof. □

As an application of our result, we obtain a nearly optimal one-shot communication region for Task C and the lossy distributed source coding task, in terms of auxiliary random variables. A utility of our result is that the auxiliary variables involved are of size comparable to that of the input random variables. This feature is often useful from the computational point of view and present in the characterization of communication for various tasks (see [21] for such examples).

We leave open the problem of obtaining a nearly optimal characterization without using the auxiliary random variables. This is not known even for the task of source coding with a helper.

An important question that we do not answer is the formulation of a proper notion of information complexity [20] in the multi-party setting. We believe our compression results will shed light on this, as the notion of information complexity is closely tied to the message compression schemes in the two-party setting [5], [9], [19].

VI. CONCLUSION

In this work, we have studied the problem of message compression in the multi-party setting. We have obtained an achievable communication region that can be viewed as a one-shot analogue of the time sharing region for Task C. Since time-sharing is not possible in the one-shot setting, we have developed a novel hypothesis testing approach to obtain our main result.

As an application of our result, we obtain a nearly optimal one-shot communication region for Task C and the lossy distributed source coding task, in terms of auxiliary random variables. A utility of our result is that the auxiliary variables involved are of size comparable to that of the input random variables. This feature is often useful from the computational point of view and present in the characterization of communication for various tasks (see [21] for such examples).

We leave open the problem of obtaining a nearly optimal characterization without using the auxiliary random variables. This is not known even for the task of source coding with a helper.

An important question that we do not answer is the formulation of a proper notion of information complexity [20] in the multi-party setting. We believe our compression results will shed light on this, as the notion of information complexity is closely tied to the message compression schemes in the two-party setting [5], [9], [19].

APPENDIX A

KNOWN FACTS AND LEMMAS USED IN THE PROOF

The following facts and lemmas from prior works have been used at several places in our proofs.

Fact A.1 ([26], Lemma 12): Let \( \epsilon \in (0, 1) \) and \( \delta \in (0, \epsilon) \). Let \( a \) be a real number that achieves the maximum in the following:
\[ \max \left\{ a' : \Pr_{x \sim X} \left[ \frac{p_X(x)}{p_{X'}(x)} \geq 2^a \right] \geq 1 - \epsilon \right\}. \]

It holds that
\[ \tilde{D}_H(X||X') \geq a \geq \tilde{D}_H - \delta(X||X') + \log \delta. \]

Proof: Consider the set \( S' = \left\{ x : p_X(x) \frac{p_{X'}(x)}{p_X(x)} \geq 2^a \right\} \). Observe that \( \Pr_X[S'] \geq 1 - \epsilon \), by the definition of \( a \). On the other hand, \( \Pr_{X'}[S'] \leq 2^{-a} \Pr_X[S'] \leq 2^{-a} \). This shows that \( \tilde{D}_H(X||X') \geq a \), as \( S' \) is feasible for the optimization in \( \tilde{D}_H(X||X') \). For the lower bound on \( a \), let \( S \) be the set achieving the maximum in the definition of \( \tilde{D}_H^{-\delta}(X||X') \). This implies that
\[ 2^{-\tilde{D}_H^{-\delta}(X||X')} = \sum_{x \in S} p_X(x) = \sum_{x \in S} p_X(x) \frac{p_{X'}(x)}{p_X(x)}, \]
\[ 1 - \epsilon + \delta \leq \sum_{x \in S} p_X(x). \tag{40} \]

Let \( T \subset S \) be the set of all \( x \in S \) for which \( \frac{p_X(x)}{p_{X'}(x)} \geq \frac{1}{2} - \tilde{D}_H^{-\delta}(X||X') \). Then Eq. (40) ensures that \( \sum_{x \in T} p_X(x) \leq \delta \). Thus,
\[ \sum_{x \in S \setminus T} p_X(x) = \sum_{x \in S} p_X(x) - \sum_{x \in T} p_X(x) \geq 1 - \epsilon. \]

To summarize, we have found that for all \( x \in S \setminus T \), \( \Pr_X[S \setminus T] \geq 1 - \epsilon \). This implies that
\[ a \geq \tilde{D}_H^{-\delta}(X||X') + \log \delta, \]

completing the proof. □
Fact A.2: Let $G \subseteq \mathcal{X}$, $X$ be a random variable over $\mathcal{X}$ and $X'$ be the restriction of $X$ over the set $G$. It holds that
\[
\frac{1}{2} \|X - X'\|_1 = 1 - \Pr_X[G].
\]

Proof: Consider
\[
\|X - X'\|_1 = \sum_{x \in G} |p_X(x) - p_X(x)| + \sum_{x \notin G} p_X(x) = \Pr_X[G] \cdot \Pr_X[-G] + \Pr_X[G] = 2 \Pr_X[-G].
\]

Fact A.3: Let $X$ and $X'$ be two random variables distributed over the set $\mathcal{X}$ and $f : \mathcal{X} \rightarrow \mathcal{Y}$ be a function. It holds that
\[
\|f(X) - f(X')\|_1 \leq \|X - X'\|_1.
\]

Fact A.4: Let $XX'$ be joint random variables over the set $\mathcal{X} \times \mathcal{X}$ such that the random variable $X$ satisfies $p_X(x_0) = 1$ for some $x_0 \in \mathcal{X}$. Then
\[
\frac{1}{2} \|X - X'\|_1 = \Pr[X \neq X'].
\]

Proof: Consider
\[
\|X - X'\|_1 = \sum_x |p_X(x) - p_X'(x)| = \sum_{x \neq x_0} |p_X(x)| + \sum_{x = x_0} |p_X'(x)| = 2 \sum_{x \neq x_0} |p_X'(x)|.
\]

Since the last expression is equal to $2 \Pr[X \neq X']$, the proof concludes.

The following is a classical version of the convex-split lemma [24] stated in [10].

Fact A.5 (Convex-split lemma [10]): Let $\epsilon \in [0, 1)$ and $\delta \in (0, 1)$, $R$ be a non-negative integer, $XM$ be joint random variables over $\mathcal{X} \times \mathcal{M}$ and $W$ be a random variable over $\mathcal{M}$
\[
R \geq \tilde{D}_\epsilon^\delta(XM\|X \times W) + 2 \log \frac{3}{\delta}.
\]
Let $J$ be a random variable uniformly distributed over $[2^R]$ and the joint random variables $JXM_1 M_2 \ldots M_{2^R}$ be defined to be:
\[
\Pr[JXM_1 \ldots M_{2^R} = (x, m_1, \ldots, m_{2^R})] \triangleq \begin{cases} \Pr_X(x) \Pr_M(m_1) \cdots \Pr_W(m_{2^R}) & J = j \end{cases}
\]
Then
\[
\frac{1}{2} \|X M_1 M_2 \ldots M_{2^R} - X \times W \times W \times \ldots \times W\|_1 \leq \epsilon + \delta.
\]

Its bipartite generalization is as follows, given in [10].

Fact A.6 (Bipartite convex-split lemma [10]): Let $\epsilon \in [0, 1)$ and $\delta \in (0, 1)$. Let $XMN$ (jointly distributed over $\mathcal{X} \times \mathcal{M} \times \mathcal{N}$), $U$ (distributed over $\mathcal{N}$) and $V$ (distributed over $\mathcal{N}$) be random variables. Let $R_1, R_2$ be natural numbers such that
\[
\Pr_{x,m,n \sim XMN}[\sum_{y \in \mathcal{X}} |p_{X}(x,m_{n} = y)| \leq \frac{\delta}{24}, \Pr_{x,m,n \sim XMN}[\sum_{y \in \mathcal{X}} |p_{X}(x,m_{n} = y)| \leq \frac{\delta}{24} \cdot 2^{R_1 + R_2}]
\]

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 $JXM_1 \ldots M_{2^R_1}N_1 \ldots N_{2^R_2}$ be distributed as follows:
\[
\Pr[JXM_1 \ldots M_{2^R_1}N_1 \ldots N_{2^R_2} = (x, m_1, \ldots, m_{2^R_1}, n_1, \ldots, n_{2^R_2})| J = j, K = k] = \Pr_X(x) \Pr_M(m_1) \cdots \Pr_M(m_{2^R_1}) \cdot \Pr_U(m_{2^R_1 + 1}) \cdots \Pr_U(m_{2^R_1 + R_1}) \cdot \Pr_V(n_1) \cdots \Pr_V(n_{2^R_1 - 1}) \cdots \Pr_V(n_{2^R_2}).
\]

Then (below for each $j \in [2^{R_1}]$, $p_{U,j} = p_U$ and for each $k \in [2^{R_2}]$, $p_{V,k} = p_V$)
\[
\frac{1}{2^R_1} \sum_{j \in [2^{R_1}]} \sum_{k \in [2^{R_2}]} \|X M_1 \ldots M_{2^R_1}N_1 \ldots N_{2^R_2} - X \times U_1 \times \ldots \times U_{2^{R_1}} \times V_1 \times \ldots \times V_{2^{R_2}}\|_1 \leq \epsilon + \delta.
\]

We will also need a classical version of position-based decoding [25], which was obtained in [10]. Below we provide a proof for completeness.

Lemma A.7: Given $\epsilon \in [0, 1)$ and joint random variables $CH$ over $[c+1] \times \mathcal{H}$, where $\mathcal{H}$ is a finite set and $C$ is supported in $[c]$, we define $H_i \triangleq (H|C = i)$. For any $i \in [c]$, let $A_i \subset \mathcal{H}$ be a subset satisfying that $\Pr_{H_i}[A_i] \geq 1 - \epsilon$. Consider a protocol $P$ which takes a sample $h \sim H$ and sequentially verifies whether $h \in A_i$ for $i \in [c]$. $P$ terminates and outputs the first $i$ satisfying $h \in A_i$. Otherwise, it outputs $c + 1$. The output is denoted by the random variable $C'$. It holds that
\[
\frac{1}{2} \|HC - HC'\|_1 \leq \sum_{i = 0}^c \sum_{j \neq i} \Pr_{H_i}[A_j] + \epsilon.
\]

Proof: We introduce a new random variable $B$, jointly correlated with $CC'H$, that is equal to 1 if $C' = C$ and 0 otherwise. Then
\[
\|HC - HC'\|_1 \leq \|HCB - HC'B\|_1 = \|p_B(0) \|(HC|B = 0) - (HC'|B = 0)\|_1 + \|p_B(1) \|(HC|B = 1) - (HC'|B = 1)\|_1 \leq 2p_B(0) + p_B(1) \|(HC|B = 1) - (HC'|B = 1)\|_1 .
\]
Consider,
\[
p_B(1) \| (HC|B = 1) - (HC'|B = 1) \|_1 \\
= \sum_h p_B(h) \| (C|BH = (1, h)) - (C'|BH = (1, h)) \|_1 \\
= 2 \sum_h p_B(h) \cdot \Pr[C \neq C'|BH = (1, h)] \\
\leq 2 \sum_{h, i} \Pr[H_i] \cdot \Pr[C \neq C'|B = 1] \\
\leq 2 \Pr[H_i] \cdot \Pr[C \neq C'|B = 1].
\]

Here, & refers to ‘and’. The last equality uses Fact A.4, which applies since conditioned on $h$, the random variable $C'$ takes a fixed value. Finally,
\[
p_B(0) = \sum_i p_C(i) p_B(C = i | B = 0) \\
\leq \sum_i p_C(i) \sum_{j < i} \Pr[A_j] \\
\leq \sum_i p_C(i) \sum_{j \neq i} \Pr[A_j].
\]

Here, the last equality holds since $B = 0$ if $j < i$. Combining both the inequalities above, we conclude the result. □

**APPENDIX B**

**DETERMINISTIC TEST VERSUS PROBABILISTIC TEST IN SMOOTH HYPOTHESIS TESTING DIVERSION**

Without loss of generality, one can consider deterministic tests in Equation 2. This is due to the fact that probabilistic tests do not gain significantly in comparison to the deterministic test. Consider a probabilistic test, which can be viewed as a probabilistic mixture of $m$ deterministic tests with the accepting sets $S_1, \ldots, S_m$. Let $p_i$ be the probability of using the $i$th deterministic test. The probability of accepting $X$ is
\[
\sum_{i=1}^m p_i \Pr[X | S_i] \geq 1 - \epsilon.
\]

Let the probability of accepting $X'$ be
\[
2^{-\ell} \sum_{i=1}^s p_i \Pr[X | S_i].
\]

For a $\delta \in (0, 1)$, let
\[
G \equiv \left\{ i : \Pr[X | S_i] \leq 2^{-\ell} \right\}.
\]

Applying Markov’s inequality, we obtain
\[
\sum_{i \in G} p_i \Pr[X | S_i] \geq 1 - \epsilon - \delta.
\]

This implies
\[
\sum_{i \in G} p_i \Pr[X | S_i] \geq 1 - \epsilon - \delta,
\]

from which we conclude that there exists an $i \in G$ with $\Pr[X | S_i] \geq 1 - \epsilon - \delta$. Since the same test also satisfies $\Pr[X' | S_i] \leq 2^{-\ell} + \delta$, we conclude
\[
\tilde{D}_H^{\ell+\delta} (X \| X') \geq \ell - \log \frac{1}{\delta}.
\]

\[\text{APPENDIX C}\]

**TASK C AND THE SCH TASK**

It is not immediately clear if the task of source coding with a helper is equivalent to Task C with $M = X$ and $N$ is trivial. This is because in the former, the only requirement is that Charlie outputs the correct $X$ with high probability (averaged over $Y$), whereas in the latter it is required that the global random variable is obtained with small error in $\ell_1$-distance. We show here that both definitions are equivalent up to constant factor increase in error and hence Task C with $M = X$ and $N$ trivial is equivalent to the task of source coding with a helper. More precisely, we have the following claim.

Claim C.1: Fix $\epsilon \in (0, 1)$. Let $YXX'$ be joint random variables such that $\Pr[X \neq X'] \leq \epsilon$. Then it holds that
\[
\frac{1}{2} \| YXX' - YXX'' \|_1 \leq \epsilon,
\]

where the random variables $YXX''$ are defined as
\[
p_{YXX''}(y, x, x'') = \begin{cases} p_{YX}(y, x) & \text{if } x = x'' \\ 0 & \text{otherwise} \end{cases}
\]

Proof: Consider
\[
\| YXX' - YXX'' \|_1 \\
= \sum_{x, y} p_{XY}(x, y) \| (X|XY = x, y) - (X''|XY = x, y) \|_1 \\
= 2 \sum_{x, y} p_{XY}(x, y) \Pr[S_{X'|XY = x, y} | x' \neq x] \\
= 2 \Pr[X' \neq X] \leq 2\epsilon.
\]

Above, the second equality uses Fact A.4. This completes the proof. □

**APPENDIX D**

**PROOF OF CLAIM IV.1**

Fix $\alpha \in (0, 1)$ and set $\epsilon = \frac{\alpha(1-\alpha)^2}{2}$. Assume $X \equiv \{[X]\}$ and choose $X = Y$. Let
\[
p_X(x) \equiv \begin{cases} \alpha & \text{if } x = 1 \\ 1 - \alpha & \text{otherwise} \end{cases}
\]

Let $M = N = X$ and consider
\[
p_{M|X}(m) \equiv \begin{cases} \alpha & \text{if } m = x \\ 1 - \alpha & \text{otherwise} \end{cases}
\]

\[
p_{N|X}(n) \equiv \begin{cases} \alpha & \text{if } n = x \\ 1 - \alpha & \text{otherwise} \end{cases}
\]

Then
\[
p_{MN}(m, n) = \begin{cases} \alpha^2 + (1-\alpha)^2 & \text{if } m = n = 1 \\ \alpha^2(1-\alpha) & \text{if } m = n \neq 1 \\ \alpha^2(1-\alpha)^2 & \text{if } m \neq n = 1 \end{cases}
\]

\[
\frac{\alpha^2(1-\alpha) + (1-\alpha)^2 + (1-\alpha)^3 ([X] - 2)}{([X] - 1)^2} \\
\frac{\alpha^2(1-\alpha) + (1-\alpha)^2 + (1-\alpha)^3 ([X] - 2)}{([X] - 1)^2} \quad \text{if } m \neq n \\
\frac{\alpha^2(1-\alpha) + (1-\alpha)^2 + (1-\alpha)^3 ([X] - 2)}{([X] - 1)^2} \quad \text{if } m = n \\
\frac{3\alpha(1-\alpha)^2 + (([X] - 3)(1-\alpha)^2)}{([X] - 1)^2} \quad \text{otherwise}.
\]
The probability mass of the set \( \{(x, x, n) : x \neq 1, n \neq x, n \neq 1\} \) under the random variables \( XMN \) is \( \alpha (1 - \alpha)^2 \geq \epsilon. \) Hence, there exists a tuple \( (x, x, n) \in X \times M \times N \) with \( x \neq 1, n \neq x, n \neq 1 \) such that

\[
\epsilon' \geq \log \left( \frac{p_{M|x}(x)p_{N|x}(n)}{p_{MN}(x, n)} \right) \\
\geq \log |X| - \log(1 + \alpha - 2\alpha^2) - c_1(\alpha) \left| \frac{|X|}{|X|} \right|,
\]

for some \( c_1(\alpha) \) that only depends on \( \alpha. \) On the other hand,

\[
H(MN) \\
\leq \alpha^3 \log \frac{1}{\alpha^3} + 3\alpha^2(1 - \alpha) \log \frac{|X|}{3\alpha^2(1 - \alpha)} + (1 - \alpha)^2(1 + 2\alpha) \log \frac{|X|}{(1 - \alpha)^2(1 + 2\alpha)} + c_2(\alpha) \frac{|X|}{|X|} \\
= (1 - \alpha^3) \log |X| + c_2(\alpha) \frac{|X|}{|X|} \\
+ H \left( \{(\alpha^3, 3\alpha^2(1 - \alpha), (1 - \alpha)^2(1 + 2\alpha)) \} \right),
\]

for some \( c_2(\alpha) \) that only depends on \( \alpha. \) Thus, for large \( |X|, \) we find

\[
\frac{H(MN)}{\epsilon'} \leq 1 - \alpha^3 + c_3(\alpha) \frac{|X|}{|X|},
\]

for some \( c_3(\alpha) \) that only depends on \( \alpha. \) We can solve \( \alpha \) in terms of \( \epsilon \) to conclude that either \( \alpha \leq 4\epsilon \) or \( 1 - \alpha \leq \sqrt{4\epsilon}. \)

Using the second bound, we obtain that

\[
\frac{H(MN)}{\epsilon'} \leq 6\sqrt{\epsilon} + c_3(\alpha) \frac{|X|}{|X|} \leq 7\sqrt{\epsilon},
\]

giving the desired upper bound.

**APPENDIX E**

**PROOF OF CLAIM III.11**

Proof: We will prove the first inequality. The second inequality follows along similar lines. Let \( k_1 \overset{\text{def}}{=} \bar{D}_S^\delta(XM|X \times S). \) Then

\[
\Pr_{x,m \sim XM} \left[ 2^{k_1} \leq \frac{p_{XM}(x, m)}{p_X(x)p_S(m)} \right] p_{M|x}(m) = \frac{1}{3} \leq \delta.
\]

Using \( \Pr_{x,m \sim XM} \left[ p_{M|x}(m) = \frac{1}{3} \right] = \Pr[X = M] = \frac{1}{3}, \) this implies

\[
\frac{1}{3} \cdot \Pr_{x,m \sim XM} \left[ 2^{k_1} \leq \frac{p_{M|x}(m)}{p_S(m)} \right] p_{M|x}(m) = \frac{1}{3} \leq \delta.
\]

As a result, we conclude that

\[
\Pr_{m \sim M} \left[ p_S(m) \leq \frac{2^{-k_1}}{3} \right] \leq 3\delta.
\]

Similarly, let \( k_2 = \bar{D}_H^\delta(MNZ|S \times NZ). \) Then the condition \( M \times N \) ensures that

\[
\Pr_{m,n,x \sim MNZ} \left[ 2^{k_1} \geq \frac{p_{MNz}(m, n, x)}{p_S(m)p_Nz(x, n)} \right] p_{M|x}(m) = \frac{1}{3} \leq \epsilon_1.
\]

Using

\[
\Pr_{m,n,x \sim MNZ} \left[ p_{M|x}(m) = \frac{2}{3(|X| - 1)} \right] = \Pr[Z \neq M] = \frac{2}{3}
\]

(recall that \( Z = X \)), we similarly have

\[
\frac{2}{3} \Pr_{x,m \sim MZ} \left[ 2^{k_2} \geq \frac{p_{M|x}(m)}{p_S(m)} \right] p_{M|x}(m) = \frac{2}{3(|X| - 1)}
\]

\[
= \Pr_{x,m \sim MZ} \left[ 2^{k_2} \geq \frac{p_{M|x}(m)}{p_S(m)} \right] p_{M|x}(m) = \frac{2}{3(|X| - 1)}
\]

\[
\leq \Pr_{x,m \sim MZ} \left[ 2^{k_2} \geq \frac{p_{M|x}(m)}{p_S(m)} \right] \leq \epsilon_1.
\]

Thus,

\[
\Pr_{m \sim M} \left[ p_S(m) \geq \frac{2 \cdot 2^{-k_2}}{3(|X| - 1)} \right]
\]

\[
= \Pr_{m \sim M} \left[ 2^{k_2} \geq \frac{p_{M|x}(m)}{p_S(m)} \right] p_{M|x}(m) = \frac{2}{3(|X| - 1)}
\]

\[
\leq \frac{3\epsilon_1}{2}.
\]

Now, define

\[
\ell_1 \overset{\text{def}}{=} \min \left\{ \ell : \Pr_{m \sim M} \left[ p_S(m) \leq 2^{-\ell} \right] \leq 3\delta \right\}
\]

and

\[
\ell_2 \overset{\text{def}}{=} \max \left\{ \ell : \Pr_{m \sim M} \left[ p_S(m) \geq 2^{-\ell} \right] \leq \frac{3\epsilon_1}{2} \right\}.
\]

Equations 42 and 43 imply that

\[
k_1 \geq \ell_1 - \log(3), \quad k_2 \leq \ell_2 - \log(|X| - 1) - \log \frac{3}{2}.
\]

Moreover, since \( 3\delta < \frac{1}{2} \) and \( \frac{3\epsilon_1}{2} < \frac{1}{2}, \) there exists a \( m \) such that \( 2^{-k_2} > p_S(m) > 2^{-k_1}. \) This implies that \( \ell_1 > \ell_2. \) Hence, we conclude that

\[
k_1 - k_2 \geq \ell_1 - \ell_2 + \log(|X| - 1) - \log(2) \geq \log(|X|) - 2.
\]

This concludes the proof.

**ACKNOWLEDGMENT**

The authors thank R. Jain and N. A. Warsi for helpful discussions, and especially R. Jain for pointing out the connection between Wyner’s result [34] and the convex-split method. The authors also thank anonymous referees for helpful comments and simplifications in various parts of the proofs.
REFERENCES

[1] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, Jul. 1948.
[2] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” IEEE Trans. Inf. Theory, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
[3] T. S. Han, Information-Spectrum Methods in Information Theory. Berlin, Germany: Springer-Verlag, 2003.
[4] R. Jain, J. Radhakrishnan, and P. Sen, “A direct sum theorem in communication complexity via message compression,” in Proc. 30th Int. Conf. Automata, Lang. Program. (ICALP), Berlin, Germany: Springer-Verlag, 2003, pp. 300–315.
[5] P. Harsha, R. Jain, D. Mc Allester, and J. Radhakrishnan, “The communication complexity of correlation,” IEEE Trans. Inf. Theory, vol. 56, no. 1, pp. 438–449, Jan. 2010.
[6] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, “The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels,” IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2926–2959, May 2014.
[7] P. Cuff, “Distributed channel synthesis,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.
[8] D. Huffman, “A method for the construction of minimum-redundancy codes,” Proc. IRE, vol. 40, no. 9, pp. 1098–1101, 1952.
[9] M. Braverman and A. Rao, “Information equals amortized communication,” in Proc. 52nd Symp. Found. Comput. Sci. (FOCS), Washington, DC, USA, 2011, pp. 748–757.
[10] A. Anshu, R. Jain, and N. A. Warsi, “A unified approach to source and message compression,” 2017, arXiv:1707.03619. [Online]. Available: https://arxiv.org/abs/1707.03619
[11] T. Uteymatsu and T. Matsutsta, “Revisiting the Slepain–Wolf coding problem for general sources: A direct approach,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Honolulu, HI, USA, Jun. 2014.
[12] V. Y. F. Tan and O. Kosut, “On the dispersions of three network information theory problems,” IEEE Trans. Inf. Theory, vol. 60, no. 2, pp. 881–903, Feb. 2014.
[13] N. A. Warsi, “Simple one-shot bounds for various source coding problems using smooth Rényi quantities,” Probl. Inf. Transmiss., vol. 52, no. 1, pp. 39–65, 2016.
[14] A. Wyner, “On source coding with side information at the decoder,” IEEE Trans. Inf. Theory, vol. 21, no. 3, pp. 294–300, May 1975.
[15] S. Miyake and F. Kanaya, “Coding theorems on correlated general sources,” IEICE Trans. Fundam., vol. E78-A, no. 9, pp. 1063–1070, Sep. 1995.
[16] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, “Nonasymptotic and second-order achievability bounds for coding with side-information,” IEEE Trans. Inf. Theory, vol. 61, no. 4, pp. 1574–1605, Apr. 2015.
[17] S. Verdú, “Non-asymptotic achievability bounds in multiserver information theory,” in Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton), Oct. 2012, pp. 1–8.
[18] T. Uteymatsu and T. Matsutsta, “Source coding with side information at the decoder revisited,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Hong Kong, Jun. 2015.
[19] B. Barak, M. Braverman, X. Chen, and A. Rao, “How to compress interactive communication,” SIAM J. Comput., vol. 42, no. 3, pp. 1327–1363, Jan. 2013.
[20] M. Braverman, “Interactive information complexity,” SIAM J. Comput., vol. 44, no. 6, pp. 1698–1739, 2015.
[21] A. E. Gamal and Y.-H. Kim, Network Information Theory. New York, NY, USA: Cambridge Univ. Press, 2012.
[22] S. Y. Tung, “Source coding,” Ph.D. dissertation, Cornell Univ., Ithaca, NY, USA, 1978.
[23] T. Berger, “Multiterminal source coding,” in The Information Theory Approach to Communications. New York, NY, USA: Springer-Verlag, 1978.
[24] A. Anshu, V. K. Devabathini, and R. Jain, “Quantum communication using coherent rejection sampling,” Phys. Rev. Lett., vol. 119, Sep. 2017, Art. no. 120506.
[25] A. Anshu, R. Jain, and N. A. Warsi, “Building blocks for communication over noisy quantum networks,” IEEE Trans. Inf. Theory, vol. 65, no. 2, pp. 1287–1306, Feb. 2019.
[26] M. Tomamichel and M. Hayashi, “A hierarchy of information quantities for finite block length analysis of quantum tasks,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7693–7710, Nov. 2013.
[27] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” IEEE Trans. Inf. Theory, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
[28] K. Iwata and J. Muramatsu, “An information-spectrum approach to rate-distortion function with side information,” IEICE Trans. Fundam. Electron., Commun. Comput. Sci., vol. E85-A, no. 6, pp. 1387–1395, 2002.
[29] M. H. Yassaei, M. R. Aref, and A. Gohari, “A technique for deriving one-shot achievability results in network information theory,” in Proc. IEEE Int. Symp. Inf. Theory, Jul. 2013, pp. 1287–1291.
[30] V. Kostina and S. Verdú, “Fixed-length lossy compression in the finite blocklength regime,” IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.
[31] V. Kostina, “Lossy data compression: Nonasymptotic fundamental limits,” Ph.D. dissertation, Princeton Univ., Princeton, NJ, USA, 2013.
[32] T. Berger and R. Yeung, “Multiterminal source encoding with one distortion criterion,” IEEE Trans. Inf. Theory, vol. 35, no. 2, pp. 228–236, Mar. 1989.
[33] A. B. Wagner, B. G. Kelly, and Y. Altug, “Distributed rate-distortion with common components,” IEEE Trans. Inf. Theory, vol. 57, no. 7, pp. 4055–4057, Jul. 2011.
[34] A. Wyner, “The common information of two dependent random variables,” IEEE Trans. Inf. Theory, vol. IT-21, no. 2, pp. 163–179, Mar. 1975.
[35] I. Newman, “Private vs. Common random bits in communication complexity,” Inf. Process. Lett., vol. 39, no. 2, pp. 67–71, Jul. 1991.
[36] J. von Neumann, “Various techniques used in connection with random digits,” in Monte Carlo Method (National Bureau of Standards Applied Mathematics Series), vol. 12, A. S. Householder, G. E. Forsythe, and H. H. Germond, Eds. Washington, DC, USA: US Government Printing Office, 1951, ch. 13, pp. 36–38.

Anurag Anshu received the Ph.D. degree from the Centre for Quantum Technologies, NUS, Singapore, in August 2018. He is currently a Joint Post-Doctoral Researcher with the Institute for Quantum Computing, University of Waterloo, and the Perimeter Institute for Theoretical Physics. He is interested in quantum Shannon theory, quantum communication complexity, and Hamiltonian complexity.

Penghui Yao received the Ph.D. degree in computer science from the Centre for Quantum Technologies (CQT), National University of Singapore, in 2013. He spent one year at QCT as a Research Associate; one year at the Centrum Wiskunde and Informatica, the Netherlands, as a Postdoc; one year at the Institute for Quantum Computing, University of Waterloo, Waterloo, ON, Canada, as a Postdoc; and one and a half years at the Joint Center for Quantum Information and Computer Science, University of Maryland, MD, USA, as a Hartree Post-Doctoral Fellow. He is currently an Associate Professor with the Department of Computer Science, Nanjing University, China. His research interests are in the areas of communication complexity, information theory, computational complexity, and Fourier analysis.