Strong Converses using Typical Changes of Measures and Asymptotic Markov Chains

Mustapha Hamad, Michèle Wigger, Mireille Sarkiss

arXiv:2301.06289v1 [cs.IT] 16 Jan 2023

Abstract

The paper presents exponentially-strong converses for source-coding, channel coding, and hypothesis testing problems. More specifically, it presents alternative proofs for the well-known exponentially-strong converse bounds for almost lossless source-coding with side-information and for channel coding over a discrete memoryless channel (DMC). These alternative proofs are solely based on a change of measure argument on the sets of conditionally or jointly typical sequences that result in a correct decision, and on the analysis of these measures in the asymptotic regime of infinite blocklengths. The paper also presents a new exponentially-strong converse for the K-hop hypothesis testing against independence problem with certain Markov chains and a strong converse for the two-terminal L-round interactive compression problem with multiple distortion constraints that depend on both sources and both reconstructions. This latter problem includes as special cases the Wyner-Ziv problem, the interactive function computation problem, and the compression with lossy common reconstruction problem. These new strong converse proofs are derived using similar change of measure arguments as described above and by additionally proving that certain Markov chains involving auxiliary random variables hold in the asymptotic regime of infinite blocklengths. As shown in related publications, the same method also yields converse bounds under expected resource constraints.

Index Terms

Strong converse, change of measure, asymptotic Markov chains, source coding, channel coding, hypothesis testing.

I. Introduction

Strong converse results have a rich history in information theory. They refer to proofs showing that the fundamental performance limit (such as minimum compression rate or maximum rate of communiction) of a specific system does not depend on its allowed error (or excess) probability, as long as this probability is not 1. For example, Wolfowitz' strong converse [2] established that the largest rate of communication over a discrete-memoryless channel (DMC) ensuring a probability of error smaller than δ_n (see Figure 1) always equals capacity, as long as δ_n does not tend to 1 for increasing blocklengths n. Differently stated, for all rates of communication that exceed capacity, the probability of error of any system must tend to 1 as the blocklength n tends to infinity. A stronger result, known as the *exponentially-strong converse* states that for all rates above capacity the probability of decoding error tends to 1 exponentially fast in the blocklength. This result was first established by Csiszár and Körner [3] who presented lower bounds on the error exponents at rates above capacity. Since then, various alternative proofs for the strong or exponentially-strong converse for channel coding over a DMC have been proposed, for example based on the blowing-up lemma [4], [5], by bounding the decoding error probabilities at finite blocklengths [6], [7], or by putting forward geometric and typicality arguments [8]. Various extensions to multi-user communication networks were also derived, see e.g., [5], [9], [10]. In this paper, we present yet-another proof, based on a change of measure argument that restricts to output sequences that are conditionally-typical for one of the possible codewords and lie in the decoding set of this codeword. Related is the converse proof for the wiretap channel by Tyagi and Watanabe [11], which is also based on a similar change of measure argument, but does not restrict to conditionally-typical output sequences. This restriction is however important in our converse proof to relate the rate of the code to the capacity of the DMC.

For (almost) lossless source coding, the strong converse states that any discrete-memoryless source (DMS) cannot be compressed with a rate below the entropy of the source and with reconstruction error probability that stays below 1 asymptotically for infinite blocklengths. This result essentially follows by the asymptotic equipartition property [12], [13]. The exponentially-strong converse for lossless compression [14] states that for all compression rates below entropy, the probability of reconstruction error tends to 1 exponentially fast. Strong converse results also

Part of this material was presented at IEEE Inf. Theory Workshop (ITW) 2022 [1].

M. Hamad used to be with LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France. He is now with Huawei Mathematical and Algorithmic Sciences Lab, Paris Research Center, 92100 Boulogne-Billancourt, France, mustapha.hamad7@gmail.com.

M. Wigger is with LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France, michele.wigger@telecom-paris.fr.

M. Sarkiss is with SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France, mireille.sarkiss@telecom-sudparis.eu.

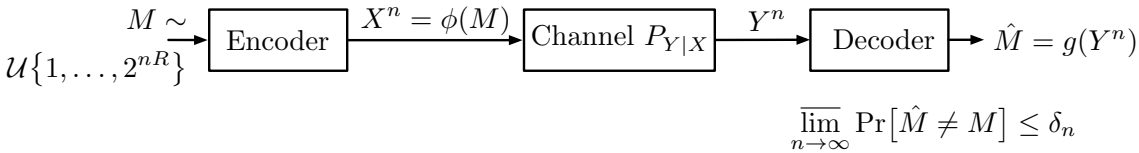


Fig. 1: Channel coding over a DMC subject to a constraint on the probability of error.

extend to lossy compression, where the limit of compression of DMSs is not entropy but the well-known rate-distortion function. The strong converse for lossy compression of DMSs was established by Körner [15], see also the related work by Kieffer [16].

Our focus is on compression scenarios where the decoder has side-information that is correlated with the source as depicted in Figure 2. For memoryless sources, the fundamental limits of compression with side-information were established by Slepian and Wolf [17] for the lossless case and by Wyner and Ziv [18] for the lossy case. Exponentially-strong converses were established by Oohama and Han [19] for the lossless case and by Oohama [20] for the lossy case. Various exponentially-strong converse results were also derived for compression problems in more complicated networks with and without side-information, see e.g., [11], [21]–[24].

In this paper, we reprove the exponentially-strong converse for lossless source coding with side-information. As a new result, we also prove the strong converse for an interactive lossy compression problem (see Figure 3)

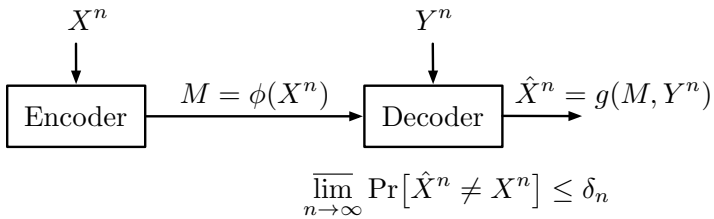
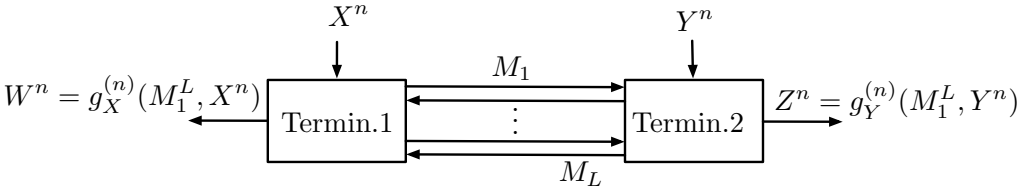


Fig. 2: Almost lossless source coding of a DMS subject to a constraint on the probability of error.

where the distortion measures can depend on the source sequences observed at the two terminals, as well as on the two terminals' produced reconstruction sequences. This problem includes as special cases Kaspi's two-terminal interactive lossy source-coding problem [25], Ma and Ishwar's two-terminal interactive function-computation problem [26], and Steinberg's lossy source coding problem with common reconstructions [27], as well as its extension by Malär et al. [28]. Our proof of the exponentially-strong converse for lossless source coding with side-information is based on a change of measure argument that restricts to source sequences that are jointly typical and result in correct decoding, and on the asymptotic analysis of this new measure. Our strong converse proof for interactive L-round lossy source coding with generalized distortion measures again relies on a change of measure argument on the set of typical sequences that result in the desired distortion and on analysing this new measure asymptotically in the limit of infinite blocklengths. As part of this analysis, we prove in particular that certain Markov chains hold in this asymptotic regime of infinite blocklengths.

The change of measure argument for proving exponentially-strong converses for source coding problems dates back to Gu and Effros [22], [23] and for function computation problems and channel coding to Tyagi and Watanabe [11]. Our proofs are similar to [11], however in our changed measures we restrict to typical sequences, which allows us to circumvent resorting to variational characterizations of the multi-letter and single-letter problems.

In this paper, we also prove a new exponentially-strong converse result for the K-hop hypothesis testing problem in [29], see Figure 4. In this problem, all $K-1$ relays as well as the final receiver guess the hypothesis by testing against independence. The figure of merit is the type-II error exponents that can be achieved at these K terminals subject to rate constraints on the K links and the constraint that the type-I error probabilities at these terminals have to stay below predefined thresholds $\delta_{k,n}$. Specifically, we consider a scenario where $K+1$ terminals observe memoryless source sequences whose underlying joint distribution depends on a binary hypothesis $\mathcal{H} \in \{0,1\}$. The distribution is $P_{Y_0} \prod_{k=1}^{K} P_{Y_k|Y_{k-1}}$ under $\mathcal{H}=0$ and it is $P_{Y_0} \prod_{k=1}^{K} P_{Y_k}$ under $\mathcal{H}=1$. Upon observing its source sequence Y_k^n, each terminal $k=0,\ldots,K-1$ can send a nR_{k+1}-bits message M_{k+1} to the next-following terminal. Terminals $1,\ldots,K$ have to produce a guess of the hypothesis $\hat{\mathcal{H}}_k \in \{0,1\}$ based on their local observations Y_k^n and their received message M_k. The main goal is to maximize their type-II error probability (the probability of error under $\mathcal{H}=1$) under the constraint that for each blocklength n the type-I error probability (the probability of error under the null hypothesis $\mathcal{H}=0$) stays below a given threshold $\delta_{k,n}$.



$$\overline{\lim_{n\to\infty}} \Pr\left[\frac{1}{n}\sum_{i=1}^{n} d_j(X_i, Y_i, W_i, Z_i) > D_j\right] \leq \delta_{j,n}, \qquad j = 1, \ldots, J.$$

Fig. 3: Interactive source coding with multiple distortions depending on the sources and the reconstructions.

For $K = 1$, this problem was solved by Ahlswede and Csiszár [30] for type-I error probabilities $\delta_{1,n}$ that are asymptotically bounded away from 1. In particular, it was shown that the maximum achievable type-II error exponent does not depend on the values $\delta_{1,n}$ as long as they do not tend to 1. For arbitrary $K \geq 2$, the problem was studied under the assumption that all the type-I error probabilities $\delta_{k,n}$ tend to 0 as $n \to \infty$ [29]. In [31], it was shown that for $K = 2$ the result in [29] applies unchanged for sequences of type-I error probabilities $\delta_{1,n}$ and $\delta_{2,n}$ for which each individual sequence as well as the sum of the two sequences is asymptotically bounded away from 1. In this work, we prove the exponentially-strong converse to this result, i.e., we show that for arbitrary $K \geq 1$ and arbitrary type-I error probabilities $\delta_{k,n}$ not vanishing exponentially fast in the blocklength, the result in [29] continues to hold. Notice that the proof of the mentioned special cases with $K = 2$ in [31] used the change of measure argument and variational characterizations proposed by Tyagi and Watanabe [11] and hypercontractivity arguments as in [32]. The proof of the strong converse for $K = 1$ in [30] was based on the blowing-up lemma [4], same as the proof of the strong converse for $K = 1$ when communication is over a DMC and without any rate constraint [33]. The latter work also used the Tyagi-Watanabe change of measure argument combined with variational characterizations. Unlike these works, our general proof does not require any blowing-up lemma or hypercontractivity arguments, nor variational characterizations.

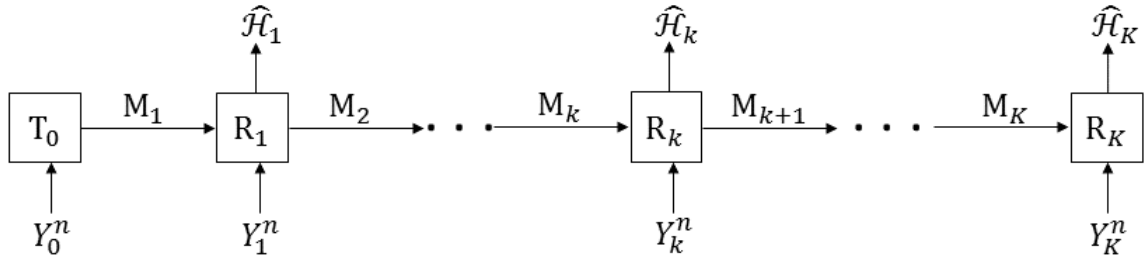


Fig. 4: K-hop hypothesis testing problem.

We summarize the main contributions of our paper:

- We present alternative exponentially-strong converse proofs for the lossless compression problem with side-information for DMSs and the channel coding problem over a DMCs. The proofs are rather simple and depend only on a change of measure argument and the asymptotic analysis of these new measures.
- We derive new strong converse results for the two-terminal L-round interactive lossy source coding problem for DMSs and with multiple distortion measures that depend on both sources and both reconstructions. This setup includes as special cases the Wyner-Ziv problem, the interactive function computation problem, and the lossy source coding problem with (lossy) common reconstructions. Previous strong converse proofs for the Wyner-Ziv problem and the interactive function computation problem relied on the method of variational characterizations and on a change of measure argument. Our proof also relies on a similar change of measure, which is however restricted to the typical set. This, combined with the asymptotic proofs of specific Markov chains allows us to directly prove the desired results without resorting to variational characterizations.
- We derive new exponentially-strong converse results for a K-hop hypothesis testing problem. Previously, a general weak converse and strong converses in special cases had been derived for this problem. The previous strong converse for $K = 2$ was based on a change of measure argument combined with variational characterizations, and on hypercontractivity arguments. The general exponentially-strong converse proof in this

paper relies on a change of measure argument restricted to the typical set and on an asymptotic analysis of this new measure including the proof of asymptotic Markov chains.

Outline of the Paper: We end this section with remarks on notation. The following Section II presents and proofs two key lemmas used in the rest of the paper. Sections III and IV present our new strong converse proof methods for the almost lossless source coding with side-information problem and for communication over a DMC. These strong converse proofs are solely based on change of measure arguments and on the analysis of these measures in the asymptotic regime of infinite blocklengths. The converse proofs in the next two Sections V and VI are also based on similar change of measure arguments and asymptotic analysis of these measures, but additionally also require proving that certain Markov chains involving auxiliary random variables hold in the asymptotic regime of infinite blocklengths. Specifically, Section V considers the L-round interactive compression problem with distortion functions that can depend on the sources and reconstructions at both terminals. Section VI considers the K-hop hypothesis testing for testing against independence where the observations at the various terminals obey some Markov conditions.

Notation: We mostly follow standard notation where upper-case letters are used for random quantities and lower-case letters for deterministic realizations. Sets are denoted using calligraphic fonts. All random variables are assumed finite and discrete. We abbreviate the n-tuples $(X_1, \ldots, X_n)$ and $(x_1, \ldots, x_n)$ as X^n and x^n and the $n-t$ tuples $(X_{t+1}, \ldots, X_n)$ and $(x_{t+1}, \ldots, x_n)$ as X_{t+1}^n and x_{t+1}^n. We further abbreviate *independent and identically distributed* as *i.i.d.* and *probability mass function* as *pmf.*

Entropy, conditional entropy, and mutual information functionals are written as $H(\cdot)$, $H(\cdot|\cdot)$, and $I(\cdot;\cdot)$, where the arguments of these functionals are random variables and whenever their probability mass function (pmf) is not clear from the context, we add it as a subscript to these functionals. The Kullback-Leibler divergence between two pmfs is denoted by $D(\cdot\|\cdot)$. We shall use $\mathcal{T}_\mu^{(n)}(P_{XY})$ to indicate the jointly strongly-typical set with respect to the pmf P_{XY} on the product alphabet $\mathcal{X} \times \mathcal{Y}$ and parameter μ as defined in [3, Definition 2.8]. Specifically, denoting by $n_{x^n,y^n}(a,b)$ the number of occurrences of the pair (a,b) in sequences (x^n, y^n):

$$n_{x^n,y^n}(a,b) = |\{t\colon (x_t, y_t) = (a,b)\}|\,, \tag{1}$$

a pair (x^n, y^n) lies in $\mathcal{T}_\mu^{(n)}(P_{XY})$ if

$$\left|\frac{n_{x^n,y^n}(a,b)}{n} - P_{XY}(a,b)\right| \leq \mu, \qquad \forall (a,b) \in \mathcal{X} \times \mathcal{Y}, \tag{2}$$

and $n_{x^n,y^n}(a,b) = 0$ whenever $P_{XY}(a,b) = 0$. The conditionally strongly-typical set with respect to a conditional pmf $P_{Y|X}$ from $\mathcal{X}$ to $\mathcal{Y}$, parameter $\mu > 0$, and sequence $x^n \in \mathcal{X}^n$ is denoted $\mathcal{T}_\mu^{(n)}(P_{Y|X}, x^n)$ [3, Definition 2.9]. It contains all sequences $y^n \in \mathcal{Y}^n$ satisfying

$$\left|\frac{n_{x^n,y^n}(a,b)}{n} - \frac{n_{x^n}(a)}{n} P_{Y|X}(b|a)\right| \leq \mu, \qquad \forall (a,b) \in \mathcal{X} \times \mathcal{Y}, \tag{3}$$

and $n_{x^n,y^n}(a,b) = 0$ whenever $P_{Y|X}(b|a) = 0$. Here $n_{x^n}(a)$ denotes the number of occurrences of symbol a in x^n. In this paper, we denote the joint type of (x^n, y^n) by $\pi_{x^n y^n}$, i.e.,

$$\pi_{x^n y^n}(a,b) \triangleq \frac{n_{x^n,y^n}(a,b)}{n}. \tag{4}$$

Accordingly, the marginal type of x^n is written as π_{x^n}. Finally, we use Landau notation $o(1)$ to indicate any function that tends to 0 for blocklengths $n \to \infty$.

II. Auxiliary Lemmas

Lemma 1: Let $\{(X_i, Y_i)\}_{i=1}^\infty$ be a sequence of pairs of i.i.d. random variables according to the pmf P_{XY}. Further let $\{\mu_n\}$ be a sequence of small positive numbers satisfying[1]

$$\lim_{n\to\infty} \mu_n = 0 \tag{5}$$

$$\lim_{n\to\infty} \left(n \cdot \mu_n^2\right)^{-1} = 0 \tag{6}$$

and for each positive integer n let $\mathcal{D}_n$ be a subset of the strongly-typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{XY})$ so that its probability

$$\Delta_n := \Pr[(X^n, Y^n) \in \mathcal{D}_n] \tag{7}$$

[1]Condition (6) ensures that the probability of the strongly typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{XY})$ under $P_{XY}^{\otimes n}$ tends to 1 as $n \to \infty$ [3, Remark to Lemma 2.12].

satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_n = 0. \tag{8}$$

Let further $(\tilde{X}^n, \tilde{Y}^n)$ be random variables of joint pmf

$$P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) = \frac{P_{XY}^{\otimes n}(x^n, y^n)}{\Delta_n} \cdot \mathbb{1}\{(x^n, y^n) \in \mathcal{D}_n\} \tag{9}$$

and T be uniform over $\{1, \ldots, n\}$ independent of all other random variables.

For the distribution in (9), the following limits hold as $n \to \infty$:

$$P_{\tilde{X}_T\tilde{Y}_T} \to P_{XY} \tag{10}$$

$$\left|\frac{1}{n}H(\tilde{X}^n\tilde{Y}^n) - H(\tilde{X}_T\tilde{Y}_T)\right| \to 0 \tag{11}$$

$$\left|\frac{1}{n}H(\tilde{Y}^n) - H(\tilde{Y}_T)\right| \to 0 \tag{12}$$

$$\left|\frac{1}{n}H(\tilde{X}^n|\tilde{Y}^n) - H(\tilde{X}_T|\tilde{Y}_T)\right| \to 0 \tag{13}$$

$$\left|H(\tilde{X}_T\tilde{Y}_T) - H(XY)\right| \to 0 \tag{14}$$

$$\left|H(\tilde{Y}_T) - H(Y)\right| \to 0 \tag{15}$$

$$\left|H(\tilde{X}_T|\tilde{Y}_T) - H(X|Y)\right| \to 0. \tag{16}$$

Proof: Notice that (14)–(16) follow directly from (10) and continuity of entropy. To prove (10), notice that

$$\begin{aligned} P_{\tilde{X}_T\tilde{Y}_T}(x, y) &= \frac{1}{n}\sum_{t=1}^{n} P_{\tilde{X}_t\tilde{Y}_t}(x, y) && (17)\\ &= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{\tilde{X}_t = x, \tilde{Y}_t = y\}\right] && (18)\\ &= \mathbb{E}[\pi_{\tilde{X}^n\tilde{Y}^n}(x, y)], && (19) \end{aligned}$$

where the expectations are with respect to the tuples $\tilde{X}^n$ and $\tilde{Y}^n$. Since by the definition of the typical set,

$$|\pi_{\tilde{X}^n\tilde{Y}^n}(x, y) - P_{XY}(x, y)| \le \mu_n, \tag{20}$$

we conclude that as $n \to \infty$ the probability $P_{\tilde{X}_T\tilde{Y}_T}(x, y)$ tends to $P_{XY}(x, y)$.

To prove (11), notice first that

$$\begin{aligned} &\frac{1}{n}H(\tilde{X}^n\tilde{Y}^n) + \frac{1}{n}D(P_{\tilde{X}^n\tilde{Y}^n}\|P_{XY}^{\otimes n})\\ &= -\frac{1}{n}\sum_{(x^n,y^n)\in\mathcal{D}_n} P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n)\log P_{XY}^{\otimes n}(x^n, y^n) && (21)\\ &= -\frac{1}{n}\sum_{t=1}^{n}\sum_{(x^n,y^n)\in\mathcal{D}_n} P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n)\log P_{XY}(x_t, y_t) && (22)\\ &= -\frac{1}{n}\sum_{t=1}^{n}\sum_{(x_t,y_t)\in\mathcal{X}\times\mathcal{Y}} P_{\tilde{X}_t\tilde{Y}_t}(x_t, y_t)\log P_{XY}(x_t, y_t) && (23)\\ &= -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\left(\frac{1}{n}\sum_{t=1}^{n} P_{\tilde{X}_t\tilde{Y}_t}(x, y)\right)\log P_{XY}(x, y) && (24)\\ &= -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{\tilde{X}_T\tilde{Y}_T}(x, y)\log P_{XY}(x, y) && (25)\\ &= H(\tilde{X}_T\tilde{Y}_T) + D(P_{\tilde{X}_T\tilde{Y}_T}\|P_{XY}), && (26) \end{aligned}$$

where (23) holds by the law of total probability applied to the random variables $\tilde{X}^{t-1}, \tilde{X}_{t+1}^n, \tilde{Y}^{t-1}, \tilde{Y}_{t+1}^n$. Combined with the following two limits (27) and (28) this establishes (11). The first relevant limit is

$$D(P_{\tilde{X}_T\tilde{Y}_T}\|P_{XY}) \to 0, \tag{27}$$

which holds by (10) and because $P_{\tilde{X}_T\tilde{Y}_T}(x,y)=0$ whenever $P_{XY}(x,y)=0$. The second limit is:

$$\frac{1}{n}D(P_{\tilde{X}^n\tilde{Y}^n}\|P_{XY}^{\otimes n})\to 0, \tag{28}$$

and holds because $\frac{1}{n}\log\Delta_n\to 0$ and by the following set of inequalities:

$$\begin{aligned}
0 &\leq \frac{1}{n}D(P_{\tilde{X}^n\tilde{Y}^n}\|P_{XY}^{\otimes n}) \\
&= \frac{1}{n}\sum_{(x^n,y^n)\in\mathcal{D}_n} P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)\log\frac{P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)}{P_{XY}^{\otimes n}(x^n,y^n)} && (29)\\
&= -\frac{1}{n}\sum_{(x^n,y^n)\in\mathcal{D}_n} P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)\log\Delta_n && (30)\\
&= -\frac{1}{n}\log\Delta_n. && (31)
\end{aligned}$$

To prove (12), notice that by the same arguments as we concluded (26), we also have

$$\frac{1}{n}H(\tilde{Y}^n)+\frac{1}{n}D(P_{\tilde{Y}^n}\|P_Y^{\otimes n})=H(\tilde{Y}_T)+D(P_{\tilde{Y}_T}\|P_Y). \tag{32}$$

Moreover, (27) and (28) imply

$$\frac{1}{n}D(P_{\tilde{Y}^n}\|P_Y^{\otimes n})\to 0 \tag{33}$$

$$D(P_{\tilde{Y}_T}\|P_Y)\to 0, \tag{34}$$

which combined with (32) imply (12).

The last limit (13) follows by the chain rule and limits (11) and (12). This concludes the proof. ■

The second lemma dates back to Csiszár and Körner [3].

Lemma 2: Let A^n and B^n be of arbitrary joint distribution and T be uniform over $\{1,\dots,n\}$ independent of (A^n,B^n). Then:

$$\begin{aligned}
&H(A^n|B^n)-H(B^n|A^n)\\
&=n\left(H(A_T|B_TA^{T-1}B_{T+1}^n)-H(B_T|A_TA^{T-1}B_{T+1}^n)\right).
\end{aligned} \tag{35}$$

The conditional version follows immediately from the definition of conditional entropy:

$$\begin{aligned}
&H(A^n|B^nS)-H(B^n|A^nS)\\
&=n\left(H(A_T|B_TA^{T-1}B_{T+1}^nS)-H(B_T|A_TA^{T-1}B_{T+1}^nS)\right)
\end{aligned} \tag{36}$$

for any random variable S so that T is independent of (A^n,B^n,S).

Proof: Kramer's simple telescoping sum states:

$$0=\sum_{i=1}^n\left[I(A^i;B_{i+1}^n)-I(A^{i-1};B_i^n)\right], \tag{37}$$

which by the chain rule of mutual information implies Csiszár's sum-identity [3]:

$$0=\sum_{i=1}^n\left[I(A_i;B_{i+1}^n|A^{i-1})-I(A^{i-1};B_i|B_{i+1}^n)\right]. \tag{38}$$

where we added and subtracted the mutual informations $\sum_{i=1}^n I(A^{i-1};B_{i+1}^n)$. Similarly, adding and subtracting the mutual informations $\sum_{i=1}^n I(A_i;B_i|A^{i-1},B_{i+1}^n)$ yields:

$$0=\sum_{i=1}^n\left[I(A_i;B_i^n|A^{i-1})-I(A^i;B_i|B_{i+1}^n)\right]. \tag{39}$$

The desired equality in (35) follows then immediately from this last equation (39) and from the chain rule:

$$\begin{aligned}
&H(A^n|B^n)-H(B^n|A^n)\\
&\overset{(a)}{=}H(A^n)-H(B^n)
\end{aligned} \tag{40}$$

$$= \sum_{i=1}^{n} \left[H(A_i|A^{i-1}) - H(B_i|B_{i+1}^n) \right] \tag{41}$$

$$\stackrel{(b)}{=} \sum_{i=1}^{n} \left[H(A_i|A^{i-1}B_i^n) - H(B_i|A^iB_{i+1}^n) \right] \tag{42}$$

$$\stackrel{(c)}{=} n(H(A_T|B_TA^{T-1}B_{T+1}^n) - H(B_T|A_TA^{T-1}B_{T+1}^n)), \tag{43}$$

where (a) is obtained by adding and subtracting $I(A^n; B^n)$, (b) holds by (39); and (c) by the definition of T.

■

III. Lossless Source Coding with Side-Information

This section studies the lossless source coding with side-information setup in Figure 2.

A. Setup and Result

Consider two terminals, an encoder observing the source sequence X^n and a decoder observing the related side-information sequence Y^n, where we assume that

$$(X^n, Y^n) \text{ i.i.d. } \sim P_{XY}, \tag{44}$$

for a given probability mass function P_{XY} on the product alphabet $\mathcal{X} \times \mathcal{Y}$. The encoder uses a function $\phi^{(n)}$ to compress the sequence X^n into a message $M \in \{1, \ldots, 2^{nR}\}$ of given rate $R > 0$:

$$M = \phi^{(n)}(X^n). \tag{45}$$

Based on this message and its own observation Y^n, the decoder is supposed to reconstruct the source sequence X^n with small probability of error. Thus, the decoder applies a decoding function $g^{(n)}$ to (M, Y^n) to produce the reconstruction sequence $\hat{X}^n \in \mathcal{X}^n$:

$$\hat{X}^n = g^{(n)}(M, Y^n). \tag{46}$$

Definition 1: Given a sequence of error probabilities $\{\delta_n\}$, the rate $R > 0$ is said $\{\delta_n\}$-achievable if there exist sequences (in n) of encoding and reconstruction functions $\phi^{(n)}$ and $g^{(n)}$ such that for each blocklength n:

$$\Pr\left[X^n \neq g^{(n)}(\phi^{(n)}(X^n))\right] \leq \delta_n. \tag{47}$$

A standard result in information theory is [19]:

Theorem 1: For any sequence $\{\delta_n\}$ satisfying

$$\lim_{n\to\infty} \frac{1}{n} \log(1 - \delta_n) = 0, \tag{48}$$

any rate $R < H(X|Y)$ is not $\{\delta_n\}$-achievable.

Notice that the theorem in particular implies an exponentially-strong converse, i.e., that for all rates $R < H(X|Y)$ the probability of error approches 1 exponentially fast in n. The result is well known, but we provide an alternative proof in the following subsection.

B. Alternative Strong Converse Proof

Fix a sequence of encoding and decoding functions $\{\phi^{(n)}, g^{(n)}\}_{n=1}^{\infty}$ satisfying (47). Choose a sequence of small positive numbers $\{\mu_n\}$ satisfying

$$\lim_{n\to\infty} \mu_n = 0 \tag{49}$$

$$\lim_{n\to\infty} \left(n \cdot \mu_n^2\right)^{-1} = 0, \tag{50}$$

and select for each n a subset

$$\mathcal{D}_n := \left\{(x^n, y^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{XY}) \colon g^{(n)}\left(\phi^{(n)}(x^n), y^n\right) = x^n\right\}, \tag{51}$$

i.e., the set of all typical (x^n, y^n)-sequences for which the reconstructed sequence $\hat{X}^n$ coincides with the source sequence X^n. Let

$$\Delta_n := \Pr[(X^n, Y^n) \in \mathcal{D}_n], \tag{52}$$

and notice that by (47) and [3, Remark to Lemma 2.12]:

$$\Delta_n \geq 1 - \delta_n - \frac{|\mathcal{X}||\mathcal{Y}|}{4\mu_n^2 n}, \tag{53}$$

and thus by (48) and (50):

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_n = 0. \tag{54}$$

Let further $(\tilde{X}^n, \tilde{Y}^n)$ be random variables of joint pmf

$$P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) = \frac{P_{X^nY^n}(x^n, y^n)}{\Delta_n} \cdot \mathbb{1}\{(x^n, y^n) \in \mathcal{D}_n\}. \tag{55}$$

Let also $\tilde{M} = \phi^{(n)}(\tilde{X}^n)$ and T be uniform over $\{1, \ldots, n\}$ independent of $(\tilde{X}^n, \tilde{Y}^n, \tilde{M})$.

The strong converse is then easily obtained as follows. Similar to the weak converse we have:

$$R \geq \frac{1}{n} H(\tilde{M}) \geq \frac{1}{n} H(\tilde{M}|\tilde{Y}^n) \geq \frac{1}{n} H(\tilde{X}^n|\tilde{Y}^n), \tag{56}$$

where the last inequality in (56) holds because $\tilde{X}^n$ can be obtained as a function of $\tilde{M}$ and $\tilde{Y}^n$, see the definition of the set $\mathcal{D}_n$.

Letting $n \to \infty$, we obtain the desired limit by (13) and (16) in Lemma 1.

IV. Communication over a Memoryless Channel

This section studies communication over a discrete memoryless channel (DMC) as depicted in Figure 1.

A. Setup and Results

Consider a transmitter (Tx) that wishes to communicate to a receiver (Rx) over a DMC parametrized by the finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ and the transition law $P_{Y|X}$. The goal of the communication is that the Tx conveys a message M to the Rx, where M is uniformly distributed over the set $\{1, \ldots, 2^{nR}\}$ with $R > 0$ and $n > 0$ denoting the rate and blocklength of communication, respectively.

For a given blocklength n, the Tx thus produces the n-length sequence of channel inputs

$$X^n = \phi^{(n)}(M) \tag{57}$$

for some choice of the encoding function $\phi^{(n)} \colon \{1, \ldots, 2^{nR}\} \to \mathcal{X}^n$, and the Rx observes the sequence of channel outputs Y^n, where the time-t output Y_t is distributed according to the law $P_{Y|X}(\cdot|x)$ when the time-t input is x, irrespective of the previous and future inputs and outputs.

The receiver attempts to guess message M based on the sequence of channel outputs Y^n:

$$\hat{M} = g^{(n)}(Y^n) \tag{58}$$

using a decoding function of the form $g^{(n)} \colon \mathcal{Y}^n \to \{1, \ldots, 2^{nR}\}$. The goal is to minimize the maximum decoding error probability

$$p^{(n)}(\text{error}) := \max_m \Pr[\hat{M} \neq M | M = m]. \tag{59}$$

Definition 2: The rate $R > 0$ is said $\{\delta_n\}$-achievable over the DMC $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, if there exists a sequence of encoding and decoding functions $\{(\phi^{(n)}, g^{(n)})\}$ such that for each blocklength n the maximum probability of error

$$p^{(n)}(\text{error}) \leq \delta_n. \tag{60}$$

A well-known result in information theory states [3]:

Theorem 2: Any rate $R > C$, where C denotes the capacity

$$C := \max_{P_X} I(X; Y), \tag{61}$$

is not $\{\delta_n\}$-achievable for all sequences $\{\delta_n\}$ satisfying

$$\lim_{n\to\infty} \frac{1}{n} \log(1 - \delta_n) = 0. \tag{62}$$

Above result implies that for all rates above capacity, the probability of error converges exponentially fast to 1. This result is well known, here we present a different converse proof.

B. Alternative Strong Converse Proof

Fix a sequence of encoding and decoding functions $\{(\phi^{(n)}, g^{(n)})\}_{n=1}^{\infty}$ so that (60) holds. Choose a sequence of small positive numbers $\{\mu_n\}$ satisfying

$$\lim_{n\to\infty} \mu_n = 0 \tag{63}$$

$$\lim_{n\to\infty} \left(n \cdot \mu_n^2\right)^{-1} = 0, \tag{64}$$

and define for each message $M = m$ the set

$$\mathcal{D}_m := \left\{ y^n \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y|X=x^n(m)}) \colon g^{(n)}\left(y^n\right) = m \right\} \tag{65}$$

and its probability

$$\Delta_m := \Pr[Y^n \in \mathcal{D}_m | M = m]. \tag{66}$$

(For readability we write the sets $\mathcal{D}_m$ and its probability Δ_m without the subscript n.) By (60) and [3, Remark to Lemma 2.12]:

$$\Delta_m \geq 1 - \delta_n - \frac{|\mathcal{Y}||\mathcal{X}|}{4\mu_n^2 n}, \tag{67}$$

and thus by (62) and (64):

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_m = 0. \tag{68}$$

Let further $(\tilde{M}, \tilde{X}^n, \tilde{Y}^n)$ be random variables so that $\tilde{M}$ is uniform over the set $\{1, \ldots, 2^{nR}\}$ (i.e., it has the same distribution as M),

$$\tilde{X}^n := \phi^{(n)}(\tilde{M}) \tag{69}$$

and

$$P_{\tilde{Y}^n|\tilde{M}}(y^n|m) = \frac{P_{Y|X}^{\otimes n}(y^n|x^n(m))}{\Delta_m} \cdot \mathbb{1}\{y^n \in \mathcal{D}_m\}. \tag{70}$$

Further, let T be independent of $(\tilde{M}, \tilde{X}^n, \tilde{Y}^n)$ and uniform over $\{1, \ldots, n\}$. Notice that since the decoding sets $\{\mathcal{D}_m\}$ are disjoint, by the definition of the new measure $P_{\tilde{Y}^n|\tilde{M}}$ it is possible to determine $\tilde{M}$ from $\tilde{Y}$ with probability 1.

Following similar steps as in the weak converse, we have:

$$\begin{aligned}
R &= \frac{1}{n} H(\tilde{M}) && (71)\\
&\overset{(a)}{=} \frac{1}{n} I(\tilde{M}; \tilde{Y}^n) && (72)\\
&= \frac{1}{n} H(\tilde{Y}^n) - \frac{1}{n} H(\tilde{Y}^n|\tilde{M}) && (73)\\
&\leq \frac{1}{n} \sum_{i=1}^{n} H(\tilde{Y}_i) - \frac{1}{n} H(\tilde{Y}^n|\tilde{M}) && (74)\\
&= H(\tilde{Y}_T|T) - \frac{1}{n} H(\tilde{Y}^n|\tilde{M}) && (75)\\
&\leq H(\tilde{Y}_T) - \frac{1}{n} H(\tilde{Y}^n|\tilde{M}), && (76)
\end{aligned}$$

where (a) holds because $\tilde{M} = g(\tilde{Y}^n)$ as explained above.

By the following lemma, by considering an appropriate subsequence of blocklengths, and by the continuity of the entropy function, we deduce that

$$R \leq I_{P_X P_{Y|X}}(X; Y) \leq C, \tag{77}$$

where the subscript indicates that mutual information is with respect to the joint pmf $P_X P_{Y|X}$ with P_X denoting the pmf mentioned in the lemma. This concludes the proof of the strong converse for channel coding.

Lemma 3: There exists an increasing subsequence of blocklengths $\{n_i\}_{i=1}^{\infty}$ such that for some pmf P_X:

$$\lim_{i\to\infty} P_{\tilde{Y}_T}(y) = \sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}(y|x) \tag{78}$$

$$\lim_{i\to\infty}\frac{1}{n_i}H(\tilde{Y}^{n_i}|\tilde{M}) = H_{P_X P_{Y|X}}(Y|X), \tag{79}$$

where $H_{P_X P_{Y|X}}(Y|X)$ denotes the conditional entropy of Y given X when the pair $(X,Y)\sim P_X P_{Y|X}$.

Proof: For readability, we will also write $x^n(m)$ and $x^n(\tilde{M})$ to indicate the (random) codewords $\phi^{(n)}(m)$ and $\phi^{(n)}(\tilde{M})$. We have:

$$\begin{aligned}
P_{\tilde{Y}_T}(y) &= \frac{1}{n}\sum_{t=1}^{n} P_{\tilde{Y}_t}(y) && (80)\\
&= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\{\tilde{Y}_t = y\}\right] && (81)\\
&= \mathbb{E}\left[\pi_{\tilde{Y}^n}(y)\right] && (82)\\
&= \sum_{x\in\mathcal{X}}\mathbb{E}\left[\pi_{x^n(\tilde{M})\tilde{Y}^n}(x,y)\right] && (83)\\
&= \sum_{x\in\mathcal{X}}\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\mathbb{E}\left[\pi_{x^n(m)\tilde{Y}^n}(x,y)\middle|\tilde{M}=m\right]. && (84)
\end{aligned}$$

Since $\mathcal{D}_m$ is a subset of the conditional typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{Y|X=x^n(m)})$, for all $m\in\{1,\dots,2^{nR}\}$, all $y^n\in\mathcal{D}_m$, and all $(x,y)\in\mathcal{X}\times\mathcal{Y}$:

$$\left|\pi_{x^n(m)y^n}(x,y) - \pi_{x^n(m)}(x)P_{Y|X}(y|x)\right| \le \mu_n, \tag{85}$$

and if $P_{Y|X}(y|x)=0$ then $\pi_{x^n(m)y^n}(x,y)=0$. Plugging these conditions into (84) we obtain

$$P_{\tilde{Y}_T}(y) \le \sum_{\substack{x\in\mathcal{X}:\\ P_{Y|X}(y|x)>0}}\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\pi_{x^n(m)}(x)P_{Y|X}(y|x) + |\mathcal{X}|\mu_n \tag{86a}$$

and similarly:

$$P_{\tilde{Y}_T}(y) \ge \sum_{\substack{x\in\mathcal{X}:\\ P_{Y|X}(y|x)>0}}\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\pi_{x^n(m)}(x)P_{Y|X}(y|x) - |\mathcal{X}|\mu_n. \tag{86b}$$

Let now $\{n_i\}$ be an increasing subsequence of blocklengths so that the sequence of average types $\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\pi_{x^n(m)}(x)$ converges for each $x\in\mathcal{X}$ and denote the convergence point by $P_X(x)$. Then, since $\mu_n\to 0$ as $n\to\infty$, by (86):

$$\lim_{i\to\infty}P_{\tilde{Y}_T}(y) = \sum_{x\in\mathcal{X}}P_X(x)P_{Y|X}(y|x), \tag{87}$$

establishing the first part of the lemma.

Notice next that by definition

$$\begin{aligned}
&\frac{1}{n}H(\tilde{Y}^n|\tilde{M}=m)\\
&= -\frac{1}{n}\sum_{y^n\in\mathcal{D}_m}P_{\tilde{Y}^n|\tilde{M}=m}(y^n)\log P_{\tilde{Y}^n|\tilde{M}=m}(y^n) && (88)\\
&= -\frac{1}{n}\sum_{y^n\in\mathcal{D}_m}P_{\tilde{Y}^n|\tilde{M}=m}(y^n)\log\frac{P_{Y|X}^{\otimes n}(y^n|x^n(m))}{\Delta_m} && (89)\\
&= -\frac{1}{n}\sum_{t=1}^{n}\sum_{y^n\in\mathcal{D}_m}P_{\tilde{Y}^n|\tilde{M}=m}(y^n)\log P_{Y|X}(y_t|x_t(m))\\
&\quad +\frac{1}{n}\log\Delta_m && (90)\\
&= -\frac{1}{n}\sum_{t=1}^{n}\sum_{y_t\in\mathcal{Y}}P_{\tilde{Y}_t|\tilde{M}=m}(y_t)\log P_{Y|X}(y_t|x_t(m))
\end{aligned}$$

$$+\frac{1}{n}\log\Delta_m \tag{91}$$

$$= -\frac{1}{n}\sum_{t=1}^{n}\sum_{y\in\mathcal{Y}}\mathbb{E}\left[\mathbb{1}\left\{\tilde{Y}_t = y\right\}\Big|\tilde{M}=m\right]\log P_{Y|X}(y|x_t(m))$$

$$+\frac{1}{n}\log\Delta_m \tag{92}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\left\{x_t(m)=x,\tilde{Y}_t=y\right\}\Big|\tilde{M}=m\right]$$
$$\cdot\log P_{Y|X}(y|x)$$

$$+\frac{1}{n}\log\Delta_m \tag{93}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\mathbb{E}\left[\pi_{x^n(m)\tilde{Y}^n}(x,y)\Big|\tilde{M}=m\right]\log P_{Y|X}(y|x)$$

$$+\frac{1}{n}\log\Delta_m. \tag{94}$$

Averaging over all messages $m\in\{1,\dots,2^{nR}\}$, we obtain

$$\frac{1}{n}H(\tilde{Y}^n|\tilde{M}) \tag{95}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\mathbb{E}\left[\pi_{x^n(m)\tilde{Y}^n}(x,y)\Bigg|\tilde{M}=m\right]\log P_{Y|X}(y|x)$$
$$+\frac{1}{n}\log\Delta_m. \tag{96}$$

By (85) and by defining P_X as the convergence point of $\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\pi_{x^n(m)}(x)$ for the sequence of blocklengths $\{n_i\}_{i=1}^{\infty}$, one can follow the same bounding steps as leading to (86) to obtain:

$$\lim_{i\to\infty}\frac{1}{n_i}H(\tilde{Y}^{n_i}|\tilde{M})$$
$$= -\sum_{x\in\mathcal{X}}P_X(x)P_{Y|X}(y|x)\log P_{Y|X}(y|x)$$
$$= H_{P_XP_{Y|X}}(Y|X), \tag{97}$$

which concludes the second part of the proof. ■

V. Interactive Lossy Compression

This section focuses on the interactive lossy compression problem depicted in Figure 3.

A. Setup

Consider two terminals, observing the related source sequences X^n and Y^n, where as in the case of source coding with side-information:

$$(X^n,Y^n)\text{ i.i.d. }\sim P_{XY}, \tag{98}$$

for a given probability mass function P_{XY} on the product alphabet $\mathcal{X}\times\mathcal{Y}$. Communication between the two terminals is over noise-free links and interactive in $L>0$ rounds. The terminal observing X^n starts communication and thus in all *odd rounds* $\ell=1,3,5,\dots$, the message M_ℓ is created as:

$$M_\ell=\phi_\ell^{(n)}(X^n,M_1,\dots,M_{\ell-1}),\qquad \ell=1,3,5,\dots, \tag{99}$$

for an encoding function $\phi_\ell^{(n)}$ on appropriate domains, where each message $M_\ell\in\{1,\dots,2^{nR_\ell}\}$, for given non-negative rates $R_1,\dots,R_L$. (Note that for $\ell=1$, $M_1=\phi_1^{(n)}(X^n)$.) In *even rounds* $\ell=2,4,6,\dots$, the message $M_\ell\in\{1,\dots,2^{nR_\ell}\}$ is created as:

$$M_\ell=\phi_\ell^{(n)}(Y^n,M_1,\dots,M_{\ell-1}),\qquad \ell=2,4,6,\dots. \tag{100}$$

At the end of the L rounds, each terminal produces a reconstruction sequence on a pre-specified alphabet. The terminal observing X^n produces

$$W^n = g_X^{(n)}(X^n, M_1, \dots, M_L) \tag{101}$$

for W^n taking value on the given alphabet $\mathcal{W}^n$. The terminal observing Y^n produces

$$Z^n = g_Y^{(n)}(Y^n, M_1, \dots, M_L) \tag{102}$$

for Z^n taking value on the given alphabet $\mathcal{Z}^n$.

The reconstructions are supposed to satisfy a set of J distortion constraints:

$$\frac{1}{n}\sum_{i=1}^{n} d_j(X_i, Y_i, W_i, Z_i) < D_j, \qquad j \in \{1, \dots, J\}, \tag{103}$$

for given non-negative symbolwise-distortion functions $d_j(\cdot,\cdot,\cdot,\cdot)$.

Definition 3: Given sequences $\{\delta_{j,n}\}$, a rate-tuple $R_1, \dots, R_L \geq 0$ is said $\{\delta_{j,n}\}$-achievable if there exist sequences (in n) of encoding functions $\{\phi_\ell^{(n)}\}_{\ell=1}^L$ and reconstruction functions $g_X^{(n)}$ and $g_Y^{(n)}$ such that the excess distortion probabilities satisfy

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n} d_j(X_i, Y_i, W_i, Z_i) > D_j\right] \leq \delta_{j,n}, \quad j \in \{1, \dots, J\}. \tag{104}$$

Remark 1: Our problem formulation includes various previously studied models as special cases. For example, the Wyner-Ziv problem [18] is included by setting $L = 1$ and choosing a distortion function of the form $d_j(X_i, Z_i)$. Kaspi's interactive source-coding problem is included by restricting to two distortion functions of the form $d_1(X_i, Z_i)$ and $d_2(Y_i, W_i)$. Lossy source coding with side-information and lossy common reconstruction [27], [28] is included by setting $L = 1$. The interactive function computation problem [26] is obtained by choosing $J = 1$, $D_1 = 0$, and distortion function $d_1(X, Y, W, Z) = \mathbb{1}\{Z = W = f(X, Y)\}$ for the desired function f.

Theorem 3: Given any sequences $\{\delta_{j,n}\}$ satisfying

$$\sum_{j=1}^{J} \delta_{j,n} < 1, \qquad n = 1, 2, \dots, \tag{105a}$$

$$\lim_{n\to\infty} \frac{1}{n} \log\left(1 - \sum_{j=1}^{J} \delta_{j,n}\right) = 0, \qquad j \in \{1, \dots, J\}, \tag{105b}$$

a rate-tuple $(R_1, \dots, R_L)$ can only be $\{\delta_{j,n}\}$-achievable if it satisfies the rate-constraints

$$R_\ell \geq I(X; U_\ell | U_1 \cdots U_{\ell-1}), \quad \ell \in \{1, \dots, L\}, \tag{106a}$$

for some auxiliary random variables $U_1, \dots, U_L$ and reconstruction random variables W and Z satisfying the distortion constraints

$$d_j\big(X, Y, W, Z\big) < D_j, \qquad j \in \{1, \dots, J\}, \tag{106b}$$

for $(X, Y) \sim P_{XY}$, and the Markov chains

$$U_\ell \to (X, U_1, \dots, U_{\ell-1}) \to Y, \quad \ell = 1, 3, 5, \dots, \tag{106c}$$

$$U_\ell \to (Y, U_1, \dots, U_{\ell-1}) \to X, \quad \ell = 2, 4, 6, \dots, \tag{106d}$$

$$W \to (X, U_1, \dots, U_L) \to Y, \tag{106e}$$

$$Z \to (Y, U_1, \dots, U_L) \to X, \tag{106f}$$

Remark 2 (A single distortion): For a single distortion constraint $J = 1$, the theorem implies that if the rate-tuple violates the constraints in the theorem, then the probability of excess distortion tends to 1 exponentially fast.

Remark 3 (Vector-valued distortions): Theorem 3 extends in a straightforward manner to vector-valued distortion functions and vector distortions

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{d}_j(X_i, Y_i, W_i, Z_i) < \boldsymbol{D}_j, \qquad j \in \{1, \dots, J\}, \tag{107}$$

where $\boldsymbol{D}_j \in \mathbb{R}^{\nu_j}$ for some positive integer ν_j, distortion functions are non-negative and of the form $\boldsymbol{d}_j \colon \mathcal{X} \times \mathcal{Y} \times \mathcal{W} \times \mathcal{Z} \to \mathbb{R}^{\nu_j}$, and inequality (107) is meant component-wise. The difference between J scalar distortion constraints as in (103) and a single J-valued vector-distortion function as in (107) is that the vector-distortion constraint limits the probability that *any* of the J constraints is violated whereas the J scalar distortion constraints individually limit the probability of each distortion to be violated.

In the following section, we prove the strong converse, i.e., the non-achievability of any rate-tuple $(R_1, \ldots, R_L)$ not satisfying the above conditions, for any sequences $\{\delta_{j,n}\}$ satisfying (105). Using standard arguments, it can be shown that for any rate-tuple $(R_1, \ldots, R_L)$ satisfying constraints (106) there exist excess probabilities $\{\delta_{j,n}\}$ all tending to 0 as $n \to \infty$ and so that the the rate-tuple $(R_1, \ldots, R_L)$ is $\{\delta_{j,n}\}$-achievable.

B. Strong Converse Proof

Fix a sequence of encoding functions $\{\phi_\ell^{(n)}\}_{\ell=1}^L$ and reconstruction functions $g_X^{(n)}$ and $g_Y^{(n)}$ satisfying (104). Choose a sequence of positive real numbers $\{\mu_n\}$ satisfying

$$\lim_{n\to\infty} \mu_n = 0 \tag{108}$$

$$\lim_{n\to\infty} \left(n \cdot \mu_n^2\right)^{-1} = 0, \tag{109}$$

and the set

$$\begin{aligned}\mathcal{D}_n := \Big\{(x^n, y^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{XY})\colon \\ d_j^{(n)}\left(g_X^{(n)}(x^n, m_1^L),\ g_Y^{(n)}\left(y^n, m_1^L\right)\right) \le D_j, \\ j \in \{1, \ldots, J\}\Big\},\end{aligned} \tag{110}$$

where $m_1^L := (m_1, \ldots, m_L)$ and for odd values of ℓ we have $m_\ell = \phi_\ell^{(n)}(x^n, m_1, \ldots, m_{\ell-1})$ whereas for even values of ℓ we have $m_\ell = \phi_\ell^{(n)}(y^n, m_1, \ldots, m_{\ell-1})$.

Define also the probability

$$\Delta_n := \Pr[(X^n, Y^n) \in \mathcal{D}_n] \tag{111}$$

and notice that by (104) and [3, Remark to Lemma 2.12]:

$$\Delta_n \ge 1 - \sum_{j=1}^J \delta_{j,n} - \frac{|\mathcal{X}||\mathcal{Y}|}{4\mu^2 n}, \tag{112}$$

which by assumptions (105b) and (109) satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_n = 0. \tag{113}$$

Let further $(\tilde{X}^n, \tilde{Y}^n)$ be random variables of joint pmf

$$P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) = \frac{P_{XY}^{\otimes n}(x^n, y^n)}{\Delta_n} \cdot \mathbb{1}\{(x^n, y^n) \in \mathcal{D}_n\}. \tag{114}$$

Let also T be uniform over $\{1, \ldots, n\}$ independent of $(\tilde{X}^n, \tilde{Y}^n)$, and define:

$$\tilde{M}_\ell = \phi_\ell^{(n)}(\tilde{X}^n, \tilde{M}_1, \ldots, \tilde{M}_{\ell-1}), \qquad \ell = 1, 3, 5, \ldots, \tag{115}$$

$$\tilde{M}_\ell = \phi_\ell^{(n)}(\tilde{Y}^n, \tilde{M}_1, \ldots, \tilde{M}_{\ell-1}), \qquad \ell = 2, 4, 6, \ldots. \tag{116}$$

Note that for $\ell = 1$, $\tilde{M}_1 = \phi_1^{(n)}(\tilde{X}^n)$. Define the auxiliary random variables

$$U_1 := (\tilde{X}^{T-1}, \tilde{Y}_{T+1}^n, \tilde{M}_1, T) \tag{117a}$$

$$U_\tau := \tilde{M}_\tau, \qquad \tau \in \{2, \ldots, L\}. \tag{117b}$$

We start with some preliminary observations. For any odd $\ell \ge 1$ observe the following:

$$\begin{aligned}&\frac{1}{n} H(\tilde{X}^n | \tilde{Y}^n \tilde{M}_1 \cdots \tilde{M}_\ell) \\ &\overset{(d)}{=} \frac{1}{n}\big[H(\tilde{X}^n | \tilde{Y}^n \tilde{M}_1 \cdots \tilde{M}_\ell)\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{\substack{\tau\in\{1,\dots,\ell\}:\\ \tau \text{ odd}}} I(\tilde{M}_\tau;\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_{\tau-1}) \\
&+ \sum_{\substack{\tau\in\{2,\dots,\ell-1\}:\\ \tau \text{ even}}} I(\tilde{M}_\tau;\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\tau-1})\Big] \qquad (118)\\
=\ & \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_\ell)-H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_\ell)\Big] \\
&+\frac{1}{n}\Big[H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})-H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})\Big] \\
&+\cdots \\
&+\frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1)-H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1)\Big] \\
&+\frac{1}{n}H(\tilde{Y}^n|\tilde{X}^n) \qquad (119)\\
\overset{(e)}{=}\ & H(\tilde{X}_T|\tilde{Y}_TU_1\cdots U_\ell)-H(\tilde{Y}_T|\tilde{X}_TU_1\cdots U_\ell) \\
&+H(\tilde{Y}_T|\tilde{X}_TU_1\cdots U_{\ell-1})-H(\tilde{X}_T|\tilde{Y}_TU_1\cdots U_{\ell-1}) \\
&+\cdots \\
&+H(\tilde{X}_T|\tilde{Y}_TU_1)-H(\tilde{Y}_T|\tilde{X}_TU_1) \\
&+H(\tilde{Y}_T|\tilde{X}_T)+o(1) \qquad (120)\\
\overset{(f)}{=}\ & H(\tilde{X}_T|\tilde{Y}_TU_1\cdots U_\ell) \\
&+\sum_{\substack{\tau\in\{1,\dots,\ell\}:\\ \tau \text{ odd}}} I(U_\tau;\tilde{Y}_T|\tilde{X}_TU_1\cdots U_{\tau-1}) \\
&+\sum_{\substack{\tau\in\{2,\dots,\ell-1\}:\\ \tau \text{ even}}} I(U_\tau;\tilde{X}_T|\tilde{Y}_TU_1\cdots U_{\tau-1})+o(1) \qquad (121)\\
\overset{(g)}{\geq}\ & H(\tilde{X}_T|\tilde{Y}_TU_1\cdots U_\ell)+o(1), \qquad (122)
\end{aligned}
$$

where, (d) holds because for τ odd, the message $\tilde{M}_\tau$ is a function of $(\tilde{X}^n,\tilde{M}_1,\dots,\tilde{M}_{\tau-1})$ and thus $I(\tilde{M}_\tau;\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_{\tau-1})=0$ whereas for τ even the message $\tilde{M}_\tau$ is a function of $(\tilde{Y}^n,\tilde{M}_1,\dots,\tilde{M}_{\tau-1})$ and thus $I(\tilde{M}_\tau;\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\tau-1})=0$; (e) holds by Lemmas 1 and 2 in Section II and Definitions (117), where for the application of Lemma 1 we also used Equation (113), and it is worth noting here that when $\ell=1$, all the terms before and including the sequential summation "+$\cdots$" do not exist anymore; (f) holds by dividing the entropy terms between sums for τ odd and even and by definition of the mutual information ; and (g) holds by the non-negativity of mutual information.

Following similar steps, we obtain for any even $\ell\geq 2$:

$$
\frac{1}{n}H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_\ell)\geq H(\tilde{Y}_T|\tilde{X}_TU_1\cdots U_\ell)+o(1). \qquad (123)
$$

We now apply bounds (122) and (123) to obtain the desired bounds on the rates and prove validity of some desired asymptotic Markov chains. For any odd $\ell\geq 1$, we have

$$
\begin{aligned}
R_\ell\ \geq\ & \frac{1}{n}H(\tilde{M}_\ell)\geq\frac{1}{n}H(\tilde{M}_\ell|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1}) \qquad (124)\\
=\ & \frac{1}{n}I(\tilde{M}_\ell;\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1}) \\
=\ & \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})-H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_\ell)\Big] \qquad (125)\\
\overset{(h)}{\geq}\ & \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-2}) \\
&\qquad -\sum_{i=1}^{n}H(\tilde{X}_i|\tilde{Y}_i\tilde{X}^{i-1}\tilde{Y}_{i+1}^n\tilde{M}_1\cdots\tilde{M}_\ell)\Big] \qquad (126)\\
\overset{(i)}{\geq}\ & H(\tilde{X}_T|\tilde{Y}_TU_1\cdots U_{\ell-2})-H(\tilde{X}_T|\tilde{Y}_TU_1\cdots U_\ell)
\end{aligned}
$$

$$+o(1) \tag{127}$$
$$= I(U_{\ell-1}U_\ell; \tilde{X}_T|\tilde{Y}_T U_1 \cdots U_{\ell-2}) + o(1) \tag{128}$$
$$\geq I(U_\ell; \tilde{X}_T|\tilde{Y}_T U_1 \cdots U_{\ell-1}) + o(1), \tag{129}$$

where (h) holds because for ℓ odd message $\tilde{M}_{\ell-1}$ is a function of the tuple $(\tilde{Y}^n, \tilde{M}_1, \ldots, \tilde{M}_{\ell-2})$ and because conditioning can only reduce entropy; and (i) holds by (117) and (122). Notice that for $\ell = 1$:

$$R_1 \geq I(U_1; \tilde{X}_T|\tilde{Y}_T) + o(1). \tag{130}$$

For any even $\ell \geq 2$, we have:

$$R_\ell \geq \frac{1}{n} H(\tilde{M}_\ell) \tag{131}$$
$$\geq \frac{1}{n} I(\tilde{M}_\ell; \tilde{Y}^n|\tilde{X}^n \tilde{M}_1 \cdots \tilde{M}_{\ell-1}) \tag{132}$$
$$\overset{(j)}{=} \frac{1}{n}\left[H(\tilde{Y}^n|\tilde{X}^n \tilde{M}_1 \cdots \tilde{M}_{\ell-2}) - H(\tilde{Y}^n|\tilde{X}^n \tilde{M}_1 \cdots \tilde{M}_\ell)\right] \tag{133}$$
$$\overset{(k)}{\geq} H(\tilde{Y}_T|\tilde{X}_T U_1 \cdots U_{\ell-2}) - H(\tilde{Y}_T|\tilde{X}_T U_1 \cdots U_\ell) + o(1) \tag{134}$$
$$\geq I(U_\ell; \tilde{Y}_T|\tilde{X}_T U_1 \cdots U_{\ell-1}) + o(1) \tag{135}$$

where (j) holds because for ℓ even $\tilde{M}_{\ell-1}$ is a function of $(\tilde{X}^n, \tilde{M}_1, \ldots, \tilde{M}_{\ell-2})$ and (k) holds by (117) and (123).

We next notice that for ℓ even (because the message $\tilde{M}_\ell$ is a function of $(\tilde{Y}^n, \tilde{M}_1, \ldots, \tilde{M}_{\ell-1})$):

$$0 = \frac{1}{n} I(\tilde{M}_\ell; \tilde{X}^n|\tilde{Y}^n \tilde{M}_1 \cdots \tilde{M}_{\ell-1})$$
$$= \frac{1}{n}\left[H(\tilde{X}^n|\tilde{Y}^n \tilde{M}_1 \cdots \tilde{M}_{\ell-1}) - H(\tilde{X}^n|\tilde{Y}^n \tilde{M}_1 \cdots \tilde{M}_\ell)\right]$$
$$\overset{(l)}{\geq} H(\tilde{X}_T|\tilde{Y}_T U_1 \cdots U_{\ell-1}) + o(1) - \frac{1}{n}\sum_{i=1}^n H(\tilde{X}_i|\tilde{X}^{i-1}\tilde{Y}_i \tilde{Y}_{i+1}^n \tilde{M}_1 \cdots \tilde{M}_\ell) \tag{136}$$
$$= I(U_\ell; \tilde{X}_T|\tilde{Y}_T U_1 \cdots U_{\ell-1}) + o(1), \tag{137}$$

where (l) holds by (122) and because conditioning can only reduce entropy.

Similarly, for $\ell \geq 1$ odd (because the message $\tilde{M}_\ell$ is a function of $(\tilde{X}^n, \tilde{M}_1, \ldots, \tilde{M}_{\ell-1})$):

$$0 = \frac{1}{n} I(\tilde{M}_\ell; \tilde{Y}^n|\tilde{X}^n \cdots \tilde{M}_{\ell-1})$$
$$= \frac{1}{n}\left[H(\tilde{Y}^n|\tilde{X}^n \tilde{M}_1 \cdots \tilde{M}_{\ell-1}) - H(\tilde{Y}^n|\tilde{X}^n \tilde{M}_1 \cdots \tilde{M}_\ell)\right]$$
$$\geq I(U_\ell; \tilde{Y}_T|\tilde{X}_T U_1 \cdots U_{\ell-1}) + o(1). \tag{138}$$

In particular, for $\ell = 1$, message $\tilde{M}_1$ is a function of $\tilde{X}^n$ and we have:

$$0 = \frac{1}{n} I(\tilde{M}_1; \tilde{Y}^n|\tilde{X}^n) \geq I(U_1; \tilde{Y}_T|\tilde{X}_T) + o(1). \tag{139}$$

Let now $\tilde{W}^n := g_X(\tilde{X}^n, \tilde{M}_1, \ldots, \tilde{M}_L)$ and $\tilde{Z}^n := g_Y(\tilde{Y}^n, \tilde{M}_1, \ldots, \tilde{M}_L)$ and notice that by our definition of the set $\mathcal{D}_n$, for any $j \in \{1, \ldots, J\}$:

$$D_j \geq \frac{1}{n}\sum_{i=1}^n d_j\left(\tilde{X}_i, \tilde{Y}_i, \tilde{W}_i, \tilde{Z}_i\right) \tag{140}$$
$$= \mathbb{E}\left[d_j\left(\tilde{X}_T, \tilde{Y}_T, \tilde{W}_T, \tilde{Z}_T\right)\right]. \tag{141}$$

For simplicity, in the sequel we assume that L is even; if L is odd the proof is similar. Similarly to (137) and (138), since $\tilde{W}^n := g_X(\tilde{X}^n, \tilde{M}_1, \ldots, \tilde{M}_L)$ and $\tilde{Z}^n := g_Y(\tilde{Y}^n, \tilde{M}_1, \ldots, \tilde{M}_L)$, we have:

$$0 = \frac{1}{n} I(\tilde{Z}^n; \tilde{X}^n|\tilde{Y}^n \tilde{M}_1 \cdots \tilde{M}_L)$$

$$
\begin{aligned}
&\overset{(m)}{=} \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{L-1}) - H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_L\tilde{Z}^n)\Big]\\
&\overset{(n)}{\geq} H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{L-1}) + o(1)\\
&\quad -\frac{1}{n}\sum_{i=1}^{n} H(\tilde{X}_i|\tilde{X}^{i-1}\tilde{Y}_i\tilde{Y}_{i+1}^n\tilde{M}_1\cdots\tilde{M}_L\tilde{Z}_i) \qquad (142)\\
&= I(U_L\tilde{Z}_T;\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{L-1}) + o(1)\\
&\geq I(\tilde{Z}_T;\tilde{X}_T|\tilde{Y}_T U_1\cdots U_L) + o(1) \qquad (143)
\end{aligned}
$$

and

$$
\begin{aligned}
0 &= \frac{1}{n} I(\tilde{W}^n;\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_L)\\
&= \frac{1}{n}\Big[H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_L) - H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_L\tilde{W}^n)\Big]\\
&\geq I(\tilde{W}_T;\tilde{Y}_T|\tilde{X}_T U_1\cdots U_L) + o(1), \qquad (144)
\end{aligned}
$$

where (m) holds since for even L, message $\tilde{M}_L$ is a function of $(\tilde{Y}^n, \tilde{M}_1, \cdots, \tilde{M}_{L-1})$; and (n) holds by (122) since $L-1$ is odd and because conditioning can only reduce entropy.

The desired rate constraints are then obtained by combining (129), (130), (135), (137), (138), (139), (141), (143), and (144) and by taking $n \to \infty$. Details are as follows. By Carathéodory's theorem [34, Appendix C], there exist auxiliary random variables $U_1, \ldots, U_L$ of bounded alphabets satisfying (129), (130), (135), (137), (138), (139), (141), (143), and (144). We restrict to such auxiliary random variables and invoke the Bolzano-Weierstrass theorem to conclude the existence of a pmf $P^*_{U_1\cdots U_L XYWZ}$, also abbreviated as P^*, and an increasing subsequence of blocklengths $\{n_i\}_{i=1}^{\infty}$ so that

$$
\lim_{i\to\infty} P_{U_1\cdots U_L\tilde{X}\tilde{Y}\tilde{W}\tilde{Z};n_i} = P^*_{U_1\cdots U_L XYWZ}, \qquad (145)
$$

where $P_{U_1\cdots U_L\tilde{X}\tilde{Y}\tilde{W}\tilde{Z};n_i}$ denotes the pmf of the tuple $(U_1\cdots U_L\tilde{X}_T\tilde{Y}_T\tilde{W}_T\tilde{Z}_T)$ at blocklength n_i.

Notice that for any blocklength n_i the pair $\big(\tilde{X}^{n_i}, \tilde{Y}^{n_i}\big)$ lies in the jointly typical set $\mathcal{T}^{(n_i)}_{\mu_{n_i}}(P_{XY})$, i.e., $\big|P_{\tilde{X}\tilde{Y};n_i} - P_{XY}\big| \leq \mu_{n_i}$, and thus since $\mu_n \to 0$ as $n \to \infty$, by the definition of $(\tilde{X}_T, \tilde{Y}_T)$ and by (145), the limiting pmf satisfies $P^*_{XY} = P_{XY}$. We further deduce from (129), (130), (135), (137), (138), (139), (141), (143), and (144) that:

$$
\begin{aligned}
R_\ell &\geq I_{P^*}(X;U_\ell|YU_1\cdots U_{\ell-1}), \qquad \ell = 1,3\ldots && (146a)\\
R_\ell &\geq I_{P^*}(Y;U_\ell|XU_1\cdots U_{\ell-1}), \qquad \ell = 2,4\ldots && (146b)\\
0 &= I_{P^*}(Y;U_\ell|XU_1\cdots U_{\ell-1}), \qquad \ell = 1,3\ldots && (146c)\\
0 &= I_{P^*}(X;U_\ell|YU_1\cdots U_{\ell-1}), \qquad \ell = 2,4\ldots && (146d)\\
0 &= I_{P^*}(Z;X|YU_1\cdots U_L), && (146e)\\
0 &= I_{P^*}(W;Y|XU_1\cdots U_L), && (146f)
\end{aligned}
$$

where the subscript P^* indicates that the mutual information quantities should be computed with respect to P^*.

Combined with (141), which implies

$$
D_j \geq \mathbb{E}_{P^*}[d_j(X,Y,W,Z)], \quad j \in \{1,\ldots,J\}, \qquad (147)
$$

above (in)equalities (146) conclude the desired converse proof.

VI. Testing Against Independence in a K-Hop Network

In this section we focus on the K-hop hypothesis testing setup in Figure 4.

A. Setup

Consider a system with a transmitter T_0 observing the source sequence Y_0^n, $K-1$ relays labelled $\mathrm{R}_1, \ldots, \mathrm{R}_{K-1}$ and observing sequences $Y_1^n, \ldots, Y_{K-1}^n$, respectively, and a receiver R_K observing sequence Y_K^n.

The source sequences $(Y_0^n, Y_1^n, \ldots, Y_K^n)$ are distributed according to one of two distributions depending on a binary hypothesis $\mathcal{H} \in \{0,1\}$:

$$
\begin{aligned}
&\text{if } \mathcal{H}=0: (Y_0^n, Y_1^n, \ldots, Y_K^n) \text{ i.i.d. } \sim P_{Y_0Y_1}P_{Y_2|Y_1}\cdots P_{Y_K|Y_{K-1}}; && (148a)\\
&\text{if } \mathcal{H}=1: (Y_0^n, Y_1^n, \ldots, Y_K^n) \text{ i.i.d. } \sim P_{Y_0}\cdot P_{Y_1}\cdots P_{Y_K}.
\end{aligned}
$$

(148b)

Communication takes place over K hops as illustrated in Figure 4. The transmitter T_0 sends a message $M_1 = \phi_0^{(n)}(Y_0^n)$ to the first relay R_1, which sends a message $M_2 = \phi_1^{(n)}(Y_1^n, M_1)$ to the second relay and so on. The communication is thus described by encoding functions

$$\phi_0^{(n)} \colon \mathcal{Y}_0^n \to \{1, \dots, 2^{nR_1}\}, \tag{149}$$

$$\phi_k^{(n)} \colon \mathcal{Y}_k^n \times \{1, \dots, 2^{nR_k}\} \to \{1, \dots, 2^{nR_{k+1}}\}, \quad k \in \{1, \dots, K-1\}, \tag{150}$$

and messages are obtained as:

$$M_1 = \phi_0^{(n)}(Y_0^n) \tag{151}$$

$$M_{k+1} = \phi_k^{(n)}(Y_k^n, M_k), \quad k \in \{1, \dots, K-1\}. \tag{152}$$

Each relay R_1, …, R_{K-1} as well as the receiver R_K, produces a guess of the hypothesis $\mathcal{H}$. These guesses are described by guessing functions

$$g_k^{(n)} \colon \mathcal{Y}_k^n \times \{1, \dots, 2^{nR_k}\} \to \{0, 1\}, \quad k \in \{1, \dots, K\}, \tag{153}$$

where we request that the guesses

$$\hat{\mathcal{H}}_k = g_k^{(n)}(Y_k^n, M_k), \quad k \in \{1, \dots, K\}, \tag{154}$$

have type-I error probabilities

$$\alpha_{k,n} \triangleq \Pr[\hat{\mathcal{H}}_k = 1 | \mathcal{H} = 0], \quad k \in \{1, \dots, K\}, \tag{155}$$

not exceeding given thresholds, and type-II error probabilities

$$\beta_{k,n} \triangleq \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1], \quad k \in \{1, \dots, K\}, \tag{156}$$

decaying to 0 exponentially fast with largest possible exponents.

Definition 4: Given sequences of allowed type-I error probabilities $\{\delta_{k,n}\}$ and rates $R_1, R_2, \dots, R_K \geq 0$, the exponent tuple $(\theta_1, \theta_2, \dots, \theta_K)$ is called $\{\delta_{k,n}\}$-*achievable* if there exists a sequence of encoding and decision functions $\left\{\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_{K-1}^{(n)}, g_1^{(n)}, g_2^{(n)}, \dots, g_K^{(n)}\right\}_{n \geq 1}$ satisfying for each $k \in \{1, \dots, K\}$ and blocklength n:

$$\alpha_{k,n} \leq \delta_{k,n}, \tag{157a}$$

$$\varliminf_{n \to \infty} \frac{1}{n} \log \frac{1}{\beta_{k,n}} \geq \theta_k. \tag{157b}$$

B. Old and New Results

Definition 5: For any $\ell \in \{1, \dots, K\}$, define the function

$$\eta_\ell \colon \mathbb{R}_0^+ \to \mathbb{R}_0^+ \tag{158}$$

$$R \mapsto \max_{\substack{P_{U|Y_{\ell-1}} : \\ R \geq I(U;Y_{\ell-1})}} I(U; Y_\ell). \tag{159}$$

The described setup was previously studied in [29] and [31], and an extension of the setup under variable-length coding was considered in [35]. While for a general number of users $K \geq 2$ only achievability results and weak converses were presented [29], for $K = 2$ users a strong converse was derived.

Theorem 4 (Theorems 2 and 3 in [31]): Let $K = 2$ and consider fixed allowed type-I error probabilities

$$\delta_{k,n} = \epsilon_k, \quad k \in \{1, 2\}, \tag{160}$$

for given $\epsilon_1, \epsilon_2 \in [0, 1)$ with $\epsilon_1 + \epsilon_2 \neq 1$. An exponent pair (θ_1, θ_2) is (ϵ_1, ϵ_2)-achievable if, and only if,

$$\theta_k \leq \sum_{\ell=1}^{k} \eta_\ell(R_\ell), \qquad k \in \{1, 2\}. \tag{161}$$

Remark 4: In [31], the presentation of Theorem 4 was split into two separate theorems depending on the values of ϵ_1 and ϵ_2. While [31, Theorem 2] considers the case $\epsilon_1 + \epsilon_2 < 1$ and coincides with above formulation, [31, Theorem 3] considers the case $\epsilon_1 + \epsilon_2 > 1$ and is formulated as an optimization problem over three auxiliary

random variables U_1, U_2, V. Without loss in optimality, this optimization can however be restricted to auxiliaries $U_1 = U_2$, and [31, Theorem 3] simplifies to the form presented in above Theorem 4.

Remark 5: The set of pairs (θ_1, θ_2) that are (ϵ_1, ϵ_2) achievable according to Theorem 4 does not depend on the values of ϵ_1 and ϵ_2 (as long as $\epsilon_1 + \epsilon_2 \neq 1$) and forms a rectangular region. In particular, each of the two exponents can be maximized without affecting the other exponent. This result extends to a general number of $K \geq 2$ users, as shown by the achievability result in [29] and by the strong converse result in the following Theorem 5.

Our main result in this section (Theorem 5 ahead) generalizes the strong converse in Theorem 4 to arbitrary $K \geq 2$ and arbitrary $\epsilon_1, \ldots, \epsilon_K \in [0, 1)$. Technically speaking, we prove an exponentially-strong converse result that is a stronger statement. In fact, for any k, an exponent θ_k violating Condition (163) can only be achieved with probabilities $\alpha_{k,n}$ that tend to 1 exponentially fast in the blocklength n.

Theorem 5: Let $\{\delta_{k,n}\}$ be sequences satisfying

$$\lim_{n\to\infty} \frac{1}{n} \log(1 - \delta_{k,n}) = 0, \qquad k \in \{1, \ldots, K\}. \tag{162}$$

Given rates $R_1, \ldots, R_K \geq 0$, the exponent-tuple $(\theta_1, \ldots, \theta_K)$ can only be $\{\delta_{k,n}\}$-achievable, if

$$\theta_k \leq \sum_{\ell=1}^{k} \eta_\ell(R_\ell), \quad k \in \{1, \ldots, K\}. \tag{163}$$

Remark 6: The direct part of this theorem was proved in [29] for some choice of admissible type-I error probabilities $\delta_{k,n} \to 0$, for all k. The strong converse in this theorem thus establishes the optimal exponents for arbitrary $K \geq 2$ and all sequences $\{\delta_{k,n}\}$ that satisfy (162) and do not vanish too quickly.

Remark 7: For all permissible type-I error probabilities $\{\delta_{k,n}\}$ that satisfy (162) and do not vanish too quickly, the set of achievable exponent-tuples $(\theta_1, \ldots, \theta_K)$ form a hypercube, implying that all decision centers, i.e., relays $\mathrm{R}_1, \ldots, \mathrm{R}_{K-1}$ and receiver R_K, can simultaneously achieve their optimal type-II error exponents. To prove the desired converse result in Theorem 5, it thus suffices to show that the bound in (163) holds in a setup where only the single decision center R_k takes a decision.

Remark 8: When one allows for variable-length coding and only limits the expected sizes of the message set but not its maximum sizes, then a tradeoff between the different exponents $\theta_1, \ldots, \theta_K$ arises [35]. Moreover, as also shown in [35], in that case the set of all achievable exponent tuples depends on the asymptotic values of the allowed type-I error probabilities.

C. Strong Converse Proof to Theorem 5

Let $\{\delta_{k,n}\}$ be sequences of allowed type-I error probabilities. Fix a sequence (in n) of encoding and decision functions $\{(\phi_0^{(n)}, \phi_1^{(n)}, \ldots, \phi_{K-1}^{(n)}, g_1^{(n)}, \ldots, g_K^{(n)})\}_{n\geq 1}$ satisfying (157) for $\{\delta_{k,n}\}$ and type-II error exponents $\theta_1, \ldots, \theta_K$.

Choose a sequence (in n) of small positive numbers $\{\mu_n\}$ satisfying

$$\lim_{n\to\infty} \mu_n = 0 \tag{164}$$

$$\lim_{n\to\infty} \left(n \cdot \mu_n^2\right)^{-1} = 0. \tag{165}$$

Fix now an arbitrary $k \in \{1, \ldots, K\}$ and a blocklength n, and let $\mathcal{A}_k$ denote the acceptance region of R_k, i.e.,

$$\mathcal{A}_k := \{(y_0^n, \ldots, y_k^n) \colon g_k^{(n)}(y_k^n, m_k) = 0\}, \tag{166}$$

where we define recursively $m_1 := \phi_0^{(n)}(y_0^n)$ and

$$m_\ell := \phi_{\ell-1}^{(n)}(y_{\ell-1}^n, m_{\ell-1}), \qquad \ell \in \{2, \ldots, k\}. \tag{167}$$

Define also the intersection of this acceptance region with the typical set:

$$\mathcal{D}_k \triangleq \mathcal{A}_k \cap \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 \cdots Y_k}). \tag{168}$$

By [3, Remark to Lemma 2.12], the type-I error probability constraints in (157a), and the union bound:

$$\Delta_k := P_{Y_0^n Y_1^n \cdots Y_k^n}(\mathcal{D}_k) \geq 1 - \delta_{k,n} - \frac{|\mathcal{Y}_0| \cdots |\mathcal{Y}_k|}{4\mu_n^2 n}, \tag{169}$$

and thus

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_k = 0. \tag{170}$$

Let $(\tilde{Y}_0^n, \tilde{Y}_1^n, \ldots, \tilde{Y}_k^n)$ be random variables of joint pmf

$$\begin{aligned} &P_{\tilde{Y}_0^n \tilde{Y}_1^n \cdots \tilde{Y}_k^n}(y_0^n, y_1^n, \ldots, y_k^n) \\ &= \frac{P_{Y_0^n Y_1^n \cdots Y_k^n}(y_0^n, y_1^n, \ldots, y_k^n)}{\Delta_k} \cdot \mathbb{1}\{(y_0^n, y_1^n, \ldots, y_k^n) \in \mathcal{D}_k\}. \end{aligned} \tag{171}$$

Let also $\tilde{M}_\ell = \phi_{\ell-1}^{(n)}(\tilde{M}_{\ell-1}, \tilde{Y}_{\ell-1}^n)$ and T be uniform over $\{1, \ldots, n\}$ independent of $(\tilde{Y}_0^n, \tilde{Y}_1^n, \ldots, \tilde{Y}_k^n, \tilde{M}_1, \ldots, \tilde{M}_k)$.

Notice that for any $\ell \in \{1, \ldots, k\}$:

$$\begin{aligned} R_\ell &\geq \frac{1}{n} H(\tilde{M}_\ell) && (172)\\ &= \frac{1}{n} I(\tilde{M}_\ell; \tilde{Y}_0^n \cdots \tilde{Y}_k^n) && (173)\\ &= \frac{1}{n} H(\tilde{Y}_0^n \cdots \tilde{Y}_k^n) - \frac{1}{n} H(\tilde{Y}_0^n \cdots \tilde{Y}_k^n | \tilde{M}_\ell) && (174)\\ &= H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T}) + o(1) \\ &\quad - \frac{1}{n} \sum_{t=1}^n H(\tilde{Y}_{0,t} \cdots \tilde{Y}_{k,t} | \tilde{M}_\ell \tilde{Y}_0^{t-1} \cdots \tilde{Y}_k^{t-1}) && (175)\\ &= H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T}) + o(1) - H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T} | U_\ell) && (176)\\ &= I(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T}; U_\ell) + o(1) && (177)\\ &\geq I(\tilde{Y}_{\ell-1,T}; U_\ell) + o(1), && (178) \end{aligned}$$

where we defined $U_\ell \triangleq (\tilde{M}_\ell, \tilde{Y}_0^{T-1}, \ldots, \tilde{Y}_k^{T-1}, T)$. Here, (175) holds by extending (11) to k-tuples.

We next upper bound the exponential decay of the type-II error probability. Define:

$$\begin{aligned} Q_{\tilde{M}_k}(m_k) \triangleq \sum_{y_0^n, y_1^n, \ldots, y_{k-1}^n} & P_{\tilde{Y}_0^n}(y_0^n) \cdots P_{\tilde{Y}_{k-1}^n}(y_{k-1}^n) \\ & \cdot \mathbb{1}\{m_k = \phi_k(\phi_{k-1}(\cdots(\phi_1(y_0^n)\cdots)), y_{k-1}^n)\}, \end{aligned} \tag{179}$$

and

$$\begin{aligned} Q_{M_k}(m_k) \triangleq \sum_{y_0^n, y_1^n, \ldots, y_{k-1}^n} & P_{Y_0^n}(y_0^n) \cdots P_{Y_{k-1}^n}(y_{k-1}^n) \\ & \cdot \mathbb{1}\{m_k = \phi_{k-1}(\phi_{k-2}(\cdots(\phi_0(y_0^n)\cdots)), y_{k-1}^n)\}, \end{aligned} \tag{180}$$

and notice that by (171):

$$\begin{aligned} Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k) &\leq Q_{M_k} P_{Y_k^n}(\mathcal{A}_k) \, \Delta_k^{-(k+1)} \\ &= \beta_{k,n} \Delta_k^{-(k+1)}. \end{aligned} \tag{181}$$

Moreover, by (166), the probability $P_{\tilde{M}_k \tilde{Y}_k^n}(\mathcal{A}_k) = 1$, and thus

$$D\left(P_{\tilde{M}_k \tilde{Y}_k^n}(\mathcal{A}_k) \| Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right) = -\log\left(Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right), \tag{182}$$

where on the left-hand side we slightly abused notation and mean the KL divergence of the two binary pmfs induced by $P_{\tilde{M}_k \tilde{Y}_k^n}(\mathcal{A}_k)$ and $1 - P_{\tilde{M}_k \tilde{Y}_k^n}(\mathcal{A}_k)$ and by $Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)$ and $1 - Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)$. Combined with (170), with (181), and with the data-processing inequality, we obtain from (182):

$$-\frac{1}{n} \log \beta_{k,n} \leq -\frac{1}{n} \log\left(Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right) - \frac{(k+1)}{n} \log \Delta_k \tag{183}$$

$$\leq \frac{1}{n} D\left(P_{\tilde{M}_k \tilde{Y}_k^n} \middle\| Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}\right) + o(1). \tag{184}$$

We continue to upper bound the divergence term as

$$\begin{aligned}
&\frac{1}{n} D(P_{\tilde{M}_k \tilde{Y}_k^n} || Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}) \\
&\quad = \frac{1}{n} I(\tilde{M}_k; \tilde{Y}_k^n) + \frac{1}{n} D(P_{\tilde{M}_k} || Q_{\tilde{M}_k}) && (185)\\
&\quad \leq \frac{1}{n} I(\tilde{M}_k; \tilde{Y}_k^n) + \frac{1}{n} D(P_{\tilde{Y}_{k-1}^n \tilde{M}_{k-1}} || P_{\tilde{Y}_{k-1}^n} Q_{\tilde{M}_{k-1}}) && (186)\\
&\quad \leq \frac{1}{n} I(\tilde{M}_k; \tilde{Y}_k^n) + \frac{1}{n} I(\tilde{M}_{k-1}; \tilde{Y}_{k-1}^n) \\
&\qquad\qquad + \frac{1}{n} D(P_{\tilde{Y}_{k-2}^n \tilde{M}_{k-2}} || P_{\tilde{Y}_{k-2}^n} Q_{\tilde{M}_{k-2}}) && (187)\\
&\quad \vdots \\
&\quad \leq \frac{1}{n} \sum_{\ell=1}^{k} I(\tilde{M}_\ell; \tilde{Y}_\ell^n) && (188)\\
&\quad \leq \frac{1}{n} \sum_{\ell=1}^{k} \sum_{t=1}^{n} I(\tilde{M}_\ell \tilde{Y}_0^{t-1} \cdots \tilde{Y}_k^{t-1}; \tilde{Y}_{\ell,t}) && (189)\\
&\quad \leq \sum_{\ell=1}^{k} I(U_\ell; \tilde{Y}_{\ell,T}). && (190)
\end{aligned}$$

Here (186) is obtained by the data processing inequality for KL-divergence and (190) by the definition of U_ℓ and T.

Combined with (184), we obtain

$$-\frac{1}{n} \log \beta_{k,n} \leq \sum_{\ell=1}^{k} I(U_\ell; \tilde{Y}_{\ell,T}) + o(1). \tag{191}$$

Finally, we proceed to prove that for any $\ell \in \{1, \ldots, k\}$ the Markov chain $U_\ell \to \tilde{Y}_{\ell-1,T} \to \tilde{Y}_{\ell,T}$ holds in the limit as $n \to \infty$. We start by noticing the Markov chain $\tilde{M}_1 \to \tilde{Y}_0^n \to (\tilde{Y}_1^n, \cdots, \tilde{Y}_k^n)$, and thus:

$$\begin{aligned}
0 &= \frac{1}{n} I(\tilde{M}_1; \tilde{Y}_1^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n) && (192)\\
&= \frac{1}{n} H(\tilde{Y}_1^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n) - \frac{1}{n} H(\tilde{Y}_1^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \tilde{M}_1) && (193)\\
&= H(\tilde{Y}_{1,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T}) + o(1) - \frac{1}{n} H(\tilde{Y}_1^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \tilde{M}_1) && (194)\\
&\geq H(\tilde{Y}_{1,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T}) + o(1) \\
&\quad - H(\tilde{Y}_{1,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \tilde{Y}_0^{T-1} \cdots \tilde{Y}_k^{T-1} \tilde{Y}_{0,T+1}^n \tilde{M}_1 T) && (195)\\
&\geq I(\tilde{Y}_{1,T} \cdots \tilde{Y}_{k,T}; U_1 | \tilde{Y}_{0,T}) + o(1) \geq 0, && (196)
\end{aligned}$$

where (194) is obtained by extending (13) to multiple random variables. We thus conclude that

$$\lim_{n \to \infty} I(\tilde{Y}_{1,T} \cdots \tilde{Y}_{k,T}; U_1 | \tilde{Y}_{0,T}) = 0. \tag{197}$$

Notice next that for any $\ell \in \{2, \ldots, k\}$:

$$\begin{aligned}
I(U_\ell; \tilde{Y}_{\ell,T} | \tilde{Y}_{\ell-1,T}) &\leq I(U_\ell \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-2,T}; \tilde{Y}_{\ell,T} | \tilde{Y}_{\ell-1,T}) && (198)\\
&= I(U_\ell; \tilde{Y}_{\ell,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) \\
&\quad + I(\tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-2,T}; \tilde{Y}_{\ell,T} | \tilde{Y}_{\ell-1,T}) && (199)\\
&= I(U_\ell; \tilde{Y}_{\ell,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1), && (200)
\end{aligned}$$

where the last equality can be proved by extending (11) and (13) to multiple random variables and by noting the factorization $P_{Y_0} P_{Y_1|Y_0} \cdots P_{Y_K|Y_{K-1}}$.

Following similar steps to (192)–(196), we further obtain:

$$
\begin{aligned}
0 &= \frac{1}{n} I(\tilde{M}_\ell; \tilde{Y}_\ell^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \cdots \tilde{Y}_{\ell-1}^n) && (201)\\
&= \frac{1}{n} H(\tilde{Y}_\ell^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \cdots \tilde{Y}_{\ell-1}^n) \\
&\quad - \frac{1}{n} H(\tilde{Y}_\ell^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \cdots \tilde{Y}_{\ell-1}^n \tilde{M}_\ell) && (202)\\
&= H(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) \\
&\quad - \frac{1}{n} H(\tilde{Y}_\ell^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \cdots \tilde{Y}_{\ell-1}^n \tilde{M}_\ell) && (203)\\
&\geq H(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) \\
&\quad - \frac{1}{n} \sum_{t=1}^{n} H(\tilde{Y}_{\ell,t} \cdots \tilde{Y}_{k,t} | \tilde{Y}_{0,t} \cdots \tilde{Y}_{\ell-1,t} \tilde{Y}_0^{t-1} \cdots \tilde{Y}_k^{t-1} \tilde{M}_\ell) \\
& && (204)\\
&= H(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) \\
&\quad - H(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T} \tilde{Y}_0^{T-1} \cdots \tilde{Y}_k^{T-1} \tilde{M}_\ell T) \\
& && (205)\\
&= I(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T}; U_\ell | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) && (206)\\
&\geq I(\tilde{Y}_{\ell,T}; U_\ell | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) \geq 0. && (207)
\end{aligned}
$$

We thus conclude that

$$
I(U_\ell; \tilde{Y}_{\ell,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) = o(1), \tag{208}
$$

which combined with (200) proves

$$
I(U_\ell; \tilde{Y}_{\ell,T} | \tilde{Y}_{\ell-1,T}) = o(1). \tag{209}
$$

The converse is then concluded by taking $n \to \infty$, as we explain in the following. By Carathéodory's theorem [34, Appendix C], for each n there must exist random variables $U_1, \ldots, U_k$ satisfying (209), (191), and (178) over alphabets of sizes

$$
|\mathcal{U}_\ell| \leq |\mathcal{Y}_{\ell-1}| \cdot |\mathcal{Y}_\ell| + 2, \qquad \ell \in \{1, \ldots, k\}. \tag{210}
$$

We thus restrict to random variables of above (bounded) supports and invoke the Bolzano-Weierstrass theorem to conclude for each $\ell \in \{1, \ldots, k\}$ the existence of pmfs $P^{(\ell)}_{Y_{\ell-1} Y_\ell U_\ell}$ over $\mathcal{Y}_{\ell-1} \times \mathcal{Y}_\ell \times \mathcal{U}_\ell$, also abbreviated as $P^{(\ell)}$, and an increasing subsequence of positive numbers $\{n_i\}_{i=1}^{\infty}$ satisfying

$$
\lim_{i \to \infty} P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell U_\ell; n_i} = P^{(\ell)}_{Y_{\ell-1} Y_\ell U_\ell}, \quad \ell \in \{1, \ldots, k\}, \tag{211}
$$

where $P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell U_\ell; n_i}$ denotes the pmf at blocklength n_i.

By the monotone continuity of mutual information for discrete random variables, we can then deduce that

$$
R_\ell \geq I_{P^{(\ell)}}(U_\ell; Y_{\ell-1}), \quad \ell \in \{1, \ldots, k\}, \tag{212}
$$

$$
\theta_k \leq \sum_{\ell=1}^{k} I_{P^{(\ell)}}(U_\ell; Y_\ell), \tag{213}
$$

where the subscripts indicate that mutual informations should be computed according to the indicated pmfs.

Since for any blocklength n_i the pair $\left(\tilde{Y}_{\ell-1}^{n_i}, \tilde{Y}_\ell^{n_i}\right)$ lies in the jointly typical set $\mathcal{T}^{(n_i)}_{\mu_{n_i}}(P_{Y_{\ell-1} Y_\ell})$, we have $\left|P_{Y_{\ell-1} Y_\ell; n_i} - P_{Y_{\ell-1} Y_\ell}\right| \leq \mu_{n_i}$ and thus the limiting pmfs satisfy $P^{(\ell)}_{Y_{\ell-1} Y_\ell} = P_{Y_{\ell-1} Y_\ell}$. By similar continuity considerations and by (209), for all $\ell \in \{1, \ldots, k\}$ the Markov chain

$$
U_\ell \to Y_{\ell-1} \to Y_\ell, \tag{214}
$$

holds under $P^{(\ell)}_{Y_{\ell-1} Y_\ell U_\ell}$. This concludes the proof.

VII. Conclusions and Outlook

This paper presented new strong converse proofs for source and channel coding setups and for hypothesis testing. Most of the presented converses are exponentially-strong converses and allow to conclude that the error probabilities tend to 1 exponentially fast whenever the rates (or exponents) violate certain conditions. The proofs for the standard almost lossless source coding with side-information problem and for communication over discrete memoryless channels (DMC) are solely based on change of measure arguments as inspired by [11], [22], [23] and by asymptotic analysis of the distributions implied by these changes of measure. Notice in particular that the restriction to strongly-typical and conditionally strongly-typical sets allows to simplify the proofs and circumvent additional arguments based on variational characterizations as in [11]. The converse results for almost lossless source coding and communication over DMC have been known for long, and our contribution here is to present a simple alternative proof.

In contrast, our converse proofs for the L-round interactive compression and the K-hop hypothesis testing setups are novel contributions in this article. Only special cases had been reported previously. Our proofs for these setups use similar change of measure arguments as in almost lossless source coding, but additionally also rely on the proofs of Markov chains that hold in the asymptotic regime of infinite blocklengths. These Markov chains are required to conclude existence of the desired auxiliary random variables. Strong converses of several special cases of our L-round interactive compression had been reported previously, in particular see [11]. A strong converse proof for the 2-hop hypothesis testing setup was already presented in [31]; the proof in [31] however does not apply to the case $\epsilon_1 + \epsilon_2 = 1$ and seems more complex since it is based on the blowing-up lemma and hypercontractivity arguments.

In related publications we have used the presented proof method to derive fundamental limits of hypothesis testing systems under expected rate constraints [35], [36] and (expected) secrecy constraints [37]. In contrast to the results presented in this paper, the fundamental limits of distributed hypothesis testing under these expectation constraints depend on the allowed type-I error probabilities. It turns out that the proof technique based on change of measure arguments as proposed in the present paper naturally captures this dependence.

Acknowledgement

The authors would like to thank R. Graczyk for useful discussions on the strong converse proof for channel coding.

References

[1] M. Hamad, M. Wigger, and M. Sarkiss, "Strong converses using change of measure and asymptotic Markov chains." [Online]. Available: https://arxiv.org/, May 2022.
[2] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, vol. 1, no. 4, pp. 591 – 606, 1957.
[3] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
[4] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.
[5] G. Dueck, "The strong converse coding theorem for the multiple-access channel," *Journal of Combinatorics, Information & System Sciences*, vol. 6, no. 3, pp. 187–196, 1981.
[6] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," (Prague, Czech Replublic), pp. 689–723, 962.
[7] H. V. P. Y. Polyanskiy and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
[8] R. Graczyk, "An elementary proof of the strong converse of the channel coding theorem." Nov. 2022.
[9] S. L. Fong and V. Y. Tan, "A proof of the strong converse theorem for gaussian multiple access channels," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4376–4394, 2016.
[10] S. L. Fong and V. Y. Tan, "A proof of the strong converse theorem for gaussian broadcast channels via the gaussian poincaré inequality," *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7737–7746, 2017.
[11] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 689–703, 2019.
[12] A. I. Khinchin, "The entropy concept in probability theory," *Usp. Mat. Nay*, vol. 8, pp. 3–20, 1953.
[13] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, vol. 24, pp. 196–219, 1953.
[14] I. Csiszár and G. Longo, "On error exponent for source coding and for testing simple statistical hypotheses," *Studia Sci. Math. Hungar.*, pp. 181–191, 1977.
[15] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Transactions of the 6th Prague conference on Information Theory, etc.*, (Prague, Czech Republic), pp. 411–425, 1973.
[16] J. Kieffer, "Strong converses in source coding relative to a fidelity criterion," *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 257–262, 1991.
[17] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.
[18] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.
[19] Y. Oohama and T. S. Han, "Universal coding for the slepian-wolf data compression system and the strong converse theorem," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1908–1919, 1994.
[20] Y. Oohama, "Exponential strong converse for source coding with side information at the decoder," *Entropy*, vol. 20, no. 5, p. 352, 2018.
[21] Y. Oohama, "Exponential strong converse for one helper source coding problem," *Entropy*, vol. 21, no. 6, p. 567, 2019.

[22] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *2009 IEEE International Symposium on Information Theory*, pp. 2316–2320, 2009.
[23] W. Gu and M. Effros, "A strong converse in source coding for super-source networks," in *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 395–399, 2011.
[24] O. Kosut and J. Kliewer, "Strong converses are just edge removal properties," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3315–3339, 2018.
[25] A. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 735–740, 1985.
[26] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6180–6195, 2011.
[27] Y. Steinberg, "Coding and common reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4995–5010, 2009.
[28] A. Lapidoth, A. Malär, and M. Wigger, "Constrained source-coding with side information," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3218–3237, 2014.
[29] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing in multi-hop networks." [Online]. Available: https://arxiv.org/abs/1708.05198v1, 2017.
[30] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.
[31] D. Cao, L. Zhou, and V. Y. F. Tan, "Strong converse for hypothesis testing against independence over a two-hop network," *Entropy (Special Issue on Multiuser Information Theory II)*, vol. 21, Nov. 2019.
[32] J. Liu, R. Van Handel, and S. Verdú, "Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 943–947, IEEE, 2017.
[33] S. Sreekuma and D. Gündüz, "Strong converse for testing against independence over a noisy channel," 2020.
[34] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
[35] M. Hamad, M. Wigger, and M. Sarkiss, "Multi-hop network with multiple decision centers under expected-rate constraints," 2022.
[36] M. Hamad, M. Sarkiss, and M. Wigger, "Benefits of rate-sharing for distributed hypothesis testing," in *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 2714–2719, 2022.
[37] S. Faour, M. Hamad, M. Sarkiss, and M. Wigger, "Testing against independence with an eavesdropper," 2022.