Strong Converses using Typical Changes of Measures and Asymptotic Markov Chains

Mustapha Hamad, Michèle Wigger, Mireille Sarkiss

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 new exponentially-strong converses for the K-hop hypothesis testing against independence problem with certain Markov chains and for the two-terminal L-round interactive compression problem with $J \geq 1$ distortion constraints that depend on both sources and both reconstructions. For this latter problem, the exponentially-strong converse result states that whenever the rates lie outside the rate-distortion region with vanishing excess distortion probabilities, then the sum of the J excess distortion probabilities asymptotically exceeds 1 or tends to 1 exponentially fast in the blocklength. (When the sum of the excess distortion probabilities exceeds 1, then a larger rate-distortion region is shown to be achievable.) The considered L-round J-distortion interactive source coding problem includes as special cases the Wyner-Ziv problem, the interactive function computation problem, and the compression with lossy common reconstruction problem. The new strong converse proofs for lossy compression and distributed hypothesis testing are derived using similar change of measure arguments as mentioned earlier and by additionally proving that certain Markov chains involving auxiliary random variables hold in the asymptotic regime of infinite blocklengths.

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 for systems operating beyond their fundamental vanishing-error performance limits, i.e., transmitting at communication rates above capacity or compressing sources at rates below the rate-distortion functions, the probability of error or excess distortion tends to 1 for increasing blocklengths. Exponentially-strong converses refer to proofs showing that this convergence happens exponentially fast in the blocklengths. Different techniques have been proposed in the literature to obtain such strong and exponentially-strong converses. In this paper we present a variation of the change of measure proof techniques [2], [3] and by Tyagi and Watanabe [4] for at hand of four problems:

1) Lossless source coding with side-information (see Figure 1).
2) L-round interactive source coding problem with $J \geq 1$ distortion constraints (see Figure 2).
3) K-hop distributed hypothesis testing problem (see Figure 3).
4) Communication over a discrete memoryless channel (DMC) (see Figure 4).

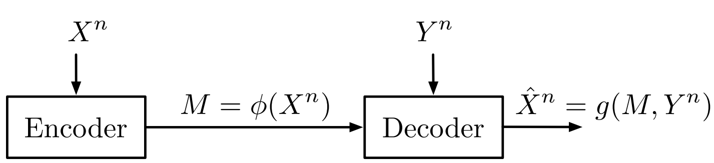

Fig. 1: Almost lossless source coding of a DMS.

The lossless source coding problem with side-information is studied for illustration purposes to highlight the role of the change of measure argument in our proof. Our motivation for choosing Problems 2)–4) is to show that the method can be applied to a wide range of applications. In particular, the interactive lossy compression problem

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.

and the K-hop hypothesis testing problem are complicated and the characterization of their fundamental limits involves several auxiliary random variables. Through these examples we illustrate that our methodology allows to treat even such complicated problems with relatively simple techniques. In fact, in addition to a similar change of measure argument as we used to solve Problem 1, we only require proving that different Markov chains hold in the asymptotic limit of infinite blocklengths. Besides the change of measure argument, the proof of such asymptotic Markov chains is a second important component in our proof.

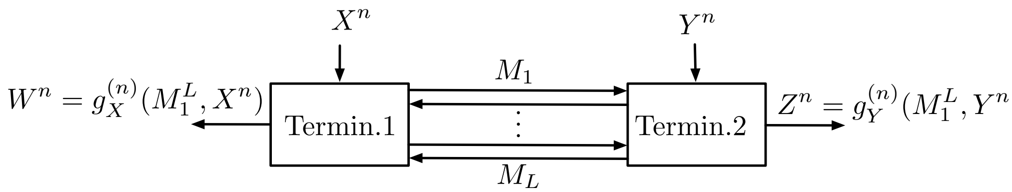

Fig. 2: Two-terminal interactive lossy source coding.

For Problem 2)—the interactive lossy source coding problem—we impose $J \geq 1$ simultaneous distortion constraints. This allows for example to capture a common reconstruction constraint [5], [6] in addition to the standard lossy reconstruction constraint between the source and the decoder's reconstruction. It also allows to have different reconstruction goals at the two interacting terminals, such as each terminal wishing to recover a different function of the two sources or each terminal wishing to reconstruct the other terminal's observations. As we show, in this multi-constraint problem, the exponentially-strong converse only applies to the sum of all excess distortions. This means, we show that when the rates lie outside the rate-distortion region with vanishing excess distortion probabilities, then the sum (over all constraints) of the probabilities of excess distortions either tends to 1 exponentially fast in the blocklength n or exceeds 1 asymptotically as $n \to \infty$. Interestingly, the same statement does not apply for the single excess distortion probabilities, as we show through counter examples. It is interesting also to notice that this sum-of-probability condition naturally shows up in our change of measure argument. In this sense, by tackling Problem 2, we can show how our proof technique is well-suited to treat multi-constraint scenarios. The following-up work [7] indeed showed that our proof method can further be used also to combine different constraints, such as constraints on the excess distortion or detection error probabilities with constraints on the decoding error probability.

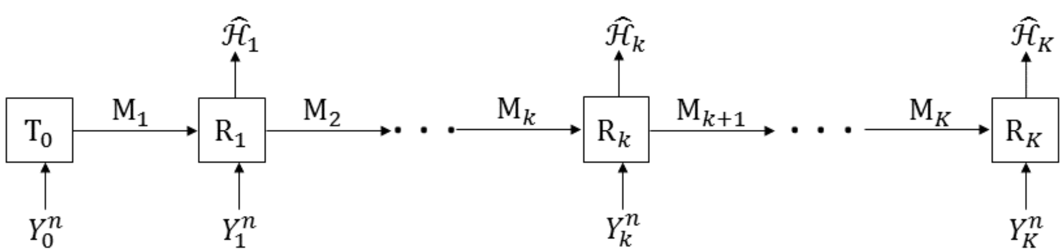

Fig. 3: K-hop hypothesis testing problem.

The scenario of Problem 3—K-hop distributed hypothesis testing—is of interest for low-energy distributed sensor systems that jointly wish to detect alerts. Due to the low-energy conditions, they will not be able to communicate in a all-to-all manner, but communication can only be established to the next-following sensor, i.e., is multi-hop. Stringent delay constraints might also impose that decisions cannot wait until communication has reached the end of the sensor-chain, but intermediate decisions on whether to raise an alert have to be taken. Our interest in studying Problem 3 was to show once more that our method can treat a complicated problem with simple steps. In particular, a previous result [8] for only $K = 2$ hops had to resort to two different proof techniques to bound the two error exponents, while we were able to present a simpler and yet unified proof for an arbitrary number of hops and thus exponents. Moreover, we managed to solve the problem entirely, while [8] left open some special cases.

Finally, we study the classical channel coding Problem 4 to prove that our method also applies to channel coding. For channel coding the change of measure technique is simple, but might not be the very simplest one. However, given the wide applicability of the change of measure proof also to source coding and hypothesis testing, its interest is also in allowing to solve problems that combine channel coding with reconstructions or detection applications. Such problems have recently become very popular in the context of integrated sensing and communication (ISAC)

Fig. 4: Channel coding over a DMC.

systems. In fact, our method recently allowed to derive converse results for ISAC systems [7] by combining the different constraints in a similar way as we combined the $J \geq 1$ distortion constraints to solve our interactive source coding problem.

In the following two subsections, we provide an overview of the previous strong converse results on source coding, hypothesis testing, and channel coding. We then describe the main ingredients of our converse proof technique and describe how it differs from previous related proof techniques.

A. Literature Review on Strong Converses and Contributions

1) Source Coding: 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 [9], [10]. The exponentially-strong converse for lossless compression [11] states that for all compression rates below entropy, the probability of reconstruction error tends to 1 exponentially fast. Strong converse results also 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 [12], see also the related work by Kieffer [13].

Our focus in this paper is on compression scenarios where the decoder has side-information that is correlated with the source as depicted in Figure 1. For memoryless sources, the fundamental limits of compression with side-information were established by Slepian and Wolf [14] for the lossless case and by Wyner and Ziv [15] for the lossy case. Exponentially-strong converses were established by Oohama and Han [16] for the lossless case and by Oohama [17] 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., [2]–[4], [18], [19]. In this paper, we reprove the exponentially-strong converse for lossless source coding with side-information.

Our main focus on source coding is on a $L \geq 1$-round interactive lossy source coding problem (see Figure 2) with $J \geq 1$ distortion constraints, which can depend on the source sequences observed at the two terminals, as well as on the two terminals' reconstruction sequences. This problem includes as special cases Kaspi's two-terminal interactive lossy source-coding problem [20], Ma and Ishwar's two-terminal interactive function-computation problem [21], and Steinberg's lossy source coding problem with common reconstructions [5], as well as its extension by Malär et al. [6].

As a new result, we prove an exponentially-strong converse result for the sum excess-distortion probability of above L-round J-distortions interactive lossy source coding problem. Specifically, we show that whenever the rates lie outside the vanishing-error rate-distortion region, then the sum of the J excess-distortions asymptotically either exceeds 1 or tends to 1 exponentially fast. Obviously for $J = 1$, our result implies the standard strong-converse result for source coding, i.e., that the excess distortion probability has to tend to 1 exponentially fast if the rate lies below the vanishing-error rate-distortion function. This same statement for a single excess distortion probability however does not hold when $J \geq 1$, as we show through a counter-example.

2) Distributed Hypothesis Testing: In this paper, we also prove a new exponentially-strong converse result for the K-hop hypothesis testing problem in [22], see Figure 3. 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 $\delta_{k,n}$.

For $K = 1$, this problem was solved by Ahlswede and Csiszár [23] 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 $n \to \infty$ [22]. In [8], it was shown that for $K = 2$ the result in [22] applies unchanged for sequences of type-I error probabilities $\delta_{1,n}$ and $\delta_{2,n}$ satisfying

$$\lim_{n\to\infty} \delta_{1,n} + \delta_{2,n} \neq 1 \tag{1}$$

$$\lim_{n\to\infty} \delta_{k,n} \neq 1, \qquad k \in \{1, 2\}, \tag{2}$$

In this work, we strengthen this result to obtain exponentially-strong converse and we get rid of Condition (1). That means, 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 [22] continues to hold. Notice that the proof of the mentioned special cases with $K = 2$ in [8] used the change of measure argument and variational characterizations proposed by Tyagi and Watanabe [4] to bound the first error exponent, and it used hypercontractivity arguments as in [24] to bound the second error exponent. The proof of the strong converse for $K = 1$ in [23] was based on the blowing-up lemma [25], [26], same as the proof of the strong converse for $K = 1$ when communication is over a DMC and without any rate constraint [27]. The latter work also used the Tyagi-Watanabe change of measure argument combined with variational characterizations. Unlike these works, our proof does not require any blowing-up lemma or hypercontractivity arguments, nor variational characterizations.

3) Channel Coding: Wolfowitz' strong converse [28] established that for all rates above capacity the probability of communication error over a discrete-memoryless channel (DMC) (see Figure 4) tends to 1 as the blocklength increases. The *exponentially-strong converse* stating that the convergence takes place exponentially fast in the blocklength, was first established by Arimoto and subsequently refined by Csiszár and Körner [29] who provided 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 [25], [26], [30], by bounding the decoding error probabilities at finite blocklengths [31]–[33], or by putting forward geometric and typicality arguments [34]. Various extensions to multi-user communication networks were also derived, see e.g., [30], [35], [36]. 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 [4]. (See our comparison of the two methods in the following subsection.)

B. Relation of our Proof Technique to the Proofs by Gu and Effros [2], [3] and by Tyagi and Watanabe [4]:

Our proof method for exponentially-strong converses has three main components:

- A change of measure argument on the jointly-typical set of the sequences so that there are no errors under the new measure;
- Bounding rates in standard ways, also introducing auxiliary random variables;
- Proof of asymptotic Markov chains of the newly introduced auxiliary random variables and convergence of multi-letter entropy quantities to single-letter entropies.

Changing measures for proving converses has a rich history in information theory. The change of measure proofs by Gu and Effros [2], [3] for source coding networks and by Tyagi and Watanabe [4] for source and channel coding are most related to our method. More precisely, in our work we define the set $\mathcal{D}_n$ of jointly-typical sequences for which there are no errors, and we obtain the new measure by restricting the original i.i.d. measure to this new set. Gu and Effros identified the same sets $\mathcal{D}_n$ but then constructed a new measure that concentrates on $\mathcal{D}_n$ only in the asymptotic limit as $n \to \infty$. This way, their new measure does not inherit from the zero-error property like our new measure, and the error probability has to be taken care of.

The zero-error property was introduced by Tyagi and Watanabe [4], who however do not restrict their new measure on the typical set. Without this restriction, they cannot prove the asymptotic single-letterization of their new measures and instead need to introduce two additional steps in their converse proof that show variational characterizations both for the multi-letter and the single-letter problems. In this sense, their proof technique requires two additional steps compared to ours. Notice that in our approach we have to prove that certain Markov chains hold in the asymptotic limits. This proof is strongly related to a different proof step of the Tyagi-Watanabe paper where they show the single-letterization of their variational characterizations, and thus does not add additional complexity to the proof.

C. *Summary of our Contributions*

We summarize the main contributions of our paper:

- We present an alternative exponentially-strong converse proof for the lossless compression problem with side-information for DMSs. The proof is simple and depends only on a change of measure argument and the asymptotic analysis of this new measure.
- We derive new strong converse results for the two-terminal L-round interactive lossy source coding problem for DMSs under multiple distortion constraints 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. We show an exponentially-strong converse for the sum of the excess distortion probabilities, i.e., we show that when the rates lie outside the rate-distortion region for vanishing excess distortion probabilities, then the sum of the excess-distortion probabilities asymptotically either exceeds 1 or tends to 1 exponentially fast in the blocklength. The constraint on the sum excess-distortion probability arises naturally in our converse proof, and we show that this sum is the right quantity to consider in a strong-converse statement. In fact, a larger rate-distortion region can be achieved when the sum of the excess-distortion probabilities is allowed to exceed 1 asymptotically.

  Our proof of this result relies again on a change of measure argument and the asymptotic proofs of specific Markov chains. It provides a method on how to combine multiple distortion constraints in a strong converse proof, see also the recent following-up work [7].
- We present a new exponentially-strong converse result for the $K \geq 1$-hop testing against independence problem where the observations satisfy a Markov chain. Previously, a strong converse was only known for $K = 2$ and certain assumptions on the missed-detection error probabilities [8]. The previous result also used two different techniques to bound the two error exponents, while in our work we present a simpler and unified method that can be used to bound each of the $K \geq 1$ exponents. Moreover, our proof holds in general and makes no assumption on the type-I error probabilities. Our proof relies on a change of measure argument combined with the proof of asymptotic Markov chains.
- We present an alternative exponentially-strong converse proof for the channel coding problem over a DMC. Our proof depends only on a change of measure argument and the asymptotic analysis of this new measure. This proof method combines well with our converse proof methods for lossy compression and distributed hypothesis testing, and was recently also used to obtain strong converse results for ISAC systems [7].

D. *Outline of the Paper*

We end this section with remarks on notation. The following Section II presents two key lemmas used in the rest of the paper. Section III presents our new strong converse proof for the almost lossless source coding with side-information problem. This strong converse proof is 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 IV and V 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 IV considers the L-round interactive compression problem with $J \geq 1$ distortion functions that can depend on the sources and reconstructions at both terminals. Section V considers the K-hop hypothesis testing for testing against independence where the observations at the various terminals obey some Markov conditions. Section VI presents an alternative proof of the exponentially-strong converse proof for the capacity of a DMC. Section VII finally concludes the paper and provides an outlook.

E. *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 [29, 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):

$$n_{x^n,y^n}(a,b) = \left|\{t \colon (x_t, y_t) = (a,b)\}\right|, \tag{3}$$

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

$$\left|\frac{n_{x^n,y^n}(a,b)}{n} - P_{XY}(a,b)\right| \leq \mu, \qquad \forall (a,b) \in \mathcal{X} \times \mathcal{Y}, \tag{4}$$

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 $\mathcal{T}_\mu^{(n)}(P_{Y|X}, x^n)$ [29, Definition 2.9]. It contains all sequences $y^n \in \mathcal{Y}^n$ satisfying

$$\left|\frac{n_{x^n,y^n}(a,b)}{n} - \frac{n_{x^n}(a)}{n} P_{Y|X}(b|a)\right| \leq \mu, \qquad \forall (a,b) \in \mathcal{X} \times \mathcal{Y}, \tag{5}$$

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.,

$$\pi_{x^n y^n}(a,b) \triangleq \frac{n_{x^n,y^n}(a,b)}{n}. \tag{6}$$

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 $\{\mu_n\}_{n=1}^\infty$ be a sequence of small positive numbers satisfying[1]

$$\lim_{n\to\infty} \mu_n = 0 \tag{7a}$$

$$\lim_{n\to\infty} n \cdot \mu_n^2 = \infty \tag{7b}$$

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

$$\Delta_n := \Pr[(X^n, Y^n) \in \mathcal{D}_n] \tag{8}$$

satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_n = 0. \tag{9}$$

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

$$P_{\tilde{X}^n \tilde{Y}^n}(x^n, y^n) = \frac{P_{XY}^{\otimes n}(x^n, y^n)}{\Delta_n} \cdot \mathbb{1}\{(x^n, y^n) \in \mathcal{D}_n\} \tag{10}$$

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

For the distribution in (10), the following limits hold as $n \to \infty$:

$$P_{\tilde{X}_T \tilde{Y}_T} \to P_{XY} \tag{11}$$

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

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

$$\left|\frac{1}{n} H(\tilde{X}^n | \tilde{Y}^n) - H(\tilde{X}_T | \tilde{Y}_T)\right| \to 0 \tag{14}$$

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

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

$$\left|H(\tilde{X}_T | \tilde{Y}_T) - H(X|Y)\right| \to 0. \tag{17}$$

Proof: See Appendix A. ■

Notice that Inequality (9) states that under the original pmf $P_{XY}^{\otimes n}$ the probability of the set $\mathcal{D}_n$ is sufficiently large and does not decay to 0 exponentially fast. This implies that the new measure $P_{\tilde{X}^n \tilde{Y}^n}$ is not blown up by an exponentially large factor compared to the original i.i.d. measure $P_{XY}^{\otimes n}$.

[1]Condition (7b) 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 $n \to \infty$ [29, Remark to Lemma 2.12].

The second lemma dates back to Csiszár and Körner [29].

Lemma 2: Let A^n, B^n, and S be of arbitrary joint distribution and T be uniform over $\{1,\ldots,n\}$ independent of (A^n, B^n, S). Then: The conditional version follows immediately from the definition of conditional entropy:

$$H(A^n|B^nS) - H(B^n|A^nS) \\ = n\left(H(A_T|B_TA^{T-1}B^n_{T+1}S) - H(B_T|A_TA^{T-1}B^n_{T+1}S)\right). \tag{18}$$

III. Lossless Source Coding with Side-Information

This section studies the lossless source coding with side-information setup in Figure 1.

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

$$(X^n, Y^n) \text{ i.i.d. } \sim P_{XY}, \tag{19}$$

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 $R > 0$:

$$M = \phi^{(n)}(X^n). \tag{20}$$

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 $\hat{X}^n \in \mathcal{X}^n$:

$$\hat{X}^n = g^{(n)}(M, Y^n). \tag{21}$$

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:

$$\Pr\left[X^n \neq g^{(n)}(\phi^{(n)}(X^n))\right] \leq \delta_n. \tag{22}$$

A well-known result in information theory is [16]:

Theorem 1: For any sequence $\{\delta_n\}_{n=1}^{\infty}$ satisfying

$$\lim_{n\to\infty} \frac{1}{n}\log(1-\delta_n) = 0, \tag{23}$$

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

In other words, if $R < H(X|Y)$, then for any sequence of encoding and reconstruction functions $\phi^{(n)}$ and $g^{(n)}$ it holds that the probability of correct reconstruction tends to 0 exponentially fast in the blocklength:

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\log\Pr\left[X^n = g^{(n)}(\phi^{(n)}(X^n))\right] < 0 \tag{24}$$

The theorem thus implies an exponentially-strong converse. 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 (22). Choose a sequence of small positive numbers $\{\mu_n\}_{n=1}^{\infty}$ satisfying (7), and select for each n a subset

$$\mathcal{D}_n := \left\{(x^n, y^n) \in \mathcal{T}^{(n)}_{\mu_n}(P_{XY})\colon g^{(n)}\left(\phi^{(n)}(x^n), y^n\right) = x^n\right\}, \tag{25}$$

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

$$\Delta_n := \Pr[(X^n, Y^n) \in \mathcal{D}_n], \tag{26}$$

and notice that

$$\Delta_n = 1 - \Pr\left[(X^n, Y^n) \notin \mathcal{T}^{(n)}_{\mu_n}(P_{XY}) \text{ or } g^{(n)}\left(\phi^{(n)}(X^n), Y^n\right) \neq X^n\right] \tag{27}$$

$$\geq 1 - \left(\Pr\left[(X^n, Y^n) \notin \mathcal{T}^{(n)}_{\mu_n}(P_{XY})\right] + \Pr\left[g^{(n)}\left(\phi^{(n)}(X^n), Y^n\right) \neq X^n\right]\right) \tag{28}$$

$$\geq 1 - \frac{|\mathcal{X}||\mathcal{Y}|}{4\mu_n^2 n} - \delta_n, \tag{29}$$

where the first inequality holds by the union bound and the second by (22) and [29, Remark to Lemma 2.12]. Therefore, by (23) and (7b):

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_n = 0. \tag{30}$$

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

$$P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) = \frac{P_{X^nY^n}(x^n, y^n)}{\Delta_n} \cdot \mathbb{1}\{(x^n, y^n) \in \mathcal{D}_n\}. \tag{31}$$

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:

$$R \geq \frac{1}{n} H(\tilde{M}) \geq \frac{1}{n} H(\tilde{M}|\tilde{Y}^n) = \frac{1}{n} H(\tilde{X}^n|\tilde{Y}^n), \tag{32}$$

where the equality in (32) holds because under our new measure $P_{\tilde{X}^n\tilde{Y}^n}$ reconstruction errors have zero probability and thus $\tilde{X}^n$ can be obtained as a function of $\tilde{M}$ and $\tilde{Y}^n$. Letting $n \to \infty$, we obtain the desired limit by (14) and (17) in Lemma 1.

Above lossless source coding example well illustrates the idea of change of measure strong converse proofs and why we would like to restrict the new measure to the jointly typical set. With the new measure we transform the problem into a zero-error problem (because the set $\mathcal{D}_n$ only includes source sequences leading to perfect reconstructions), thus avoiding cumbersome error terms and immediately obtaining the equality in (32). Moreover, the conditional entropy of the new measure tends to the one of the original i.i.d. measure, because the former is restricted to the jointly typical set and compared to the original i.i.d. measure is boosted at most by the factor Δ_n^{-1}, which does not scale exponentially in the blocklength.

Gu and Effros gave a related strong converse proof in [2]. Their proof however does not follow the bounding steps (32), but instead bounds the size of the number of source sequences x^n for which (x^n, y^n) lies in $\mathcal{D}_n$ for a specific sequence y^n, which by the zero-error property of set $\mathcal{D}_n$ lower-bounds 2^{nR}.

IV. Interactive Lossy Compression

This section focuses on the interactive lossy compression problem depicted in Figure 2.

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:

$$(X^n, Y^n) \text{ i.i.d. } \sim P_{XY}, \tag{33}$$

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* $\ell = 1, 3, 5, \ldots$, the message M_ℓ is created as:

$$M_\ell = \phi_\ell^{(n)}(X^n, M_1, \ldots, M_{\ell-1}), \qquad \ell = 1, 3, 5, \ldots, \tag{34}$$

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

$$M_\ell = \phi_\ell^{(n)}(Y^n, M_1, \ldots, M_{\ell-1}), \qquad \ell = 2, 4, 6, \ldots. \tag{35}$$

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

$$W^n = g_X^{(n)}(X^n, M_1, \ldots, M_L) \tag{36}$$

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

$$Z^n = g_Y^{(n)}(Y^n, M_1, \ldots, M_L) \tag{37}$$

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

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

$$\frac{1}{n} \sum_{i=1}^{n} d_j(X_i, Y_i, W_i, Z_i) < D_j, \qquad j \in \{1, \ldots, J\}, \tag{38}$$

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

Definition 2: Given sequences $\{\delta_{j,n}\}$, a rate-tuple $R_1,\ldots,R_L \geq 0$ 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

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n d_j(X_i,Y_i,W_i,Z_i) > D_j\right] \leq \delta_{j,n}, \quad j \in \{1,\ldots,J\}. \tag{39}$$

Remark 1: Our problem formulation includes various previously studied models as special cases. For example:

- *The Wyner-Ziv problem [15]* is included by restricting to a single interaction round $L=1$, to a single distortion function $J=1$, and by choosing a distortion function of the form
$$d_1(X_i,Y_i,W_i,Z_i) = \tilde{d}_1(X_i,Z_i) \tag{40}$$
- *Kaspi's interactive source-coding problem* is included by restricting to $J=2$ distortion functions of the form
$$d_1(X_i,Y_i,W_i,Z_i) = \tilde{d}_1(X_i,Z_i) \tag{41}$$
$$d_2(X_i,Y_i,W_i,Z_i) = \tilde{d}_2(Y_i,W_i). \tag{42}$$
- Lossy source coding with side-information and lossy common reconstruction [5], [6] is included by restricting to a single interaction round $L=1$ and the $J=2$ distortion measures
$$d_1(X_i,Z_i) = \tilde{d}_1(X_i,Z_i) \tag{43}$$
$$d_2(X_i,Y_i,W_i,Z_i) = \tilde{d}_2(W_i,Z_i). \tag{44}$$
- The interactive function computation problem [21] is obtained by choosing $J=1$, $D_1=0$, and distortion function
$$d_1(X,Y,W,Z) = \mathbb{1}\{Z = W = f(X,Y)\} \tag{45}$$
for the desired function f.

Theorem 2: Given sequences $\left\{\{\delta_{j,n}\}_{n=1}^\infty\right\}_{j=1}^J$ satisfying

$$\sum_{j=1}^J \delta_{j,n} < 1, \qquad n = 1,2,\ldots, \tag{46}$$

$$\lim_{n\to\infty}\frac{1}{n}\log\left(1-\sum_{j=1}^J \delta_{j,n}\right) = 0, \qquad j \in \{1,\ldots,J\}, \tag{47}$$

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

$$R_\ell \geq I(X;U_\ell|U_1\cdots U_{\ell-1}Y), \quad \ell \in \{1,\ldots,L\}, \qquad \ell \text{ odd} \tag{48a}$$
$$R_\ell \geq I(Y;U_\ell|U_1\cdots U_{\ell-1}X), \quad \ell \in \{1,\ldots,L\}, \qquad \ell \text{ even}, \tag{48b}$$

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

$$\mathbb{E}[d_j(X,Y,W,Z)] < D_j, \qquad j \in \{1,\ldots,J\}, \tag{48c}$$

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

$$U_\ell \to (X,U_1,\ldots,U_{\ell-1}) \to Y, \quad \ell = 1,3,5,\ldots, \tag{48d}$$
$$U_\ell \to (Y,U_1,\ldots,U_{\ell-1}) \to X, \quad \ell = 2,4,6,\ldots, \tag{48e}$$
$$W \to (X,U_1,\ldots,U_L) \to Y, \tag{48f}$$
$$Z \to (Y,U_1,\ldots,U_L) \to X, \tag{48g}$$

We can thus conclude that if a rate-tuple $(R_1,\ldots,R_L)$ satisfies (48), then the sum of the excess distortions asymptotically either exceeds 1 or it approaches 1 exponentially fast in the blocklength.

Remark 2 (A single distortion): For a single distortion constraint $J = 1$, above 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 (Condition (47) *cannot be relaxed):* Condition (47) in Theorem 2 cannot be relaxed and Remark 2 does not apply for $J > 1$, as we explain in the following. For simplicity, consider the case $J = 2$, $L = 1$, and

$$\delta_{1,n} = \delta_{2,n} = 1/2 + \epsilon, \tag{49}$$

for a positive $\epsilon \in (0, 1/2)$. Then, the following rate is achievable

$$R_1 \geq \min_{P_{U_1|X}, P_{U_1'|X}} \max \{I(X; U_1|Y),\ I(Y; U_1'|X)\} \tag{50}$$

where the minimum is over all conditional pmfs for which there exist reconstruction random variables $W = g_X(X)$ and $Z = g_Y(Y, U_1)$ and $W' = g'_X(X)$ and $Z' = g'_Y(Y, U_1')$ satisfying the distortion constraints

$$\mathbb{E}[d_1(X, Y, W, Z)] < D_1 \tag{51}$$

$$\mathbb{E}[d_2(X, Y, W', Z')] < D_2. \tag{52}$$

Notice that the rate in (159) can violate the conditions in above Theorem 2, because the theorem would force $U_1 = U_1'$ and $W' = W$ and $Z' = Z$.

The rate in (159) is achieved by a randomized scheme, where with probability $1/2$ the encoder sends a first flagbit 0 followed by a Wyner-Ziv message using auxiliary distribution $P_{U_1|X}$ and it applies reconstruction function $g_X(\cdot)$, and with probability $1/2$ it sends the flagbit 1 followed by a Wyner-Ziv message using the auxiliary distribution $P_{U_1'|X}$ and applies reconstruction function g'_X. Upon receiving flagbit 0, the decoder uses the Wyner-Ziv decoder for $P_{U_1|X}$ and reconstruction function g_Y, and upon receiving flagbit 1, the decoder uses the Wyner-Ziv decoder for $P_{U_1'|X}$ and reconstruction function g'_Y. Since Wyner-Ziv codes can achieve vanishing probabilities of excess distortions, our system satisfies both distortion constraints with probability $1/2$, which by (49) is below $\delta_{1,n}$ and $\delta_{2,n}$.

Remark 4 (Vector-valued distortions): Theorem 2 extends in a straightforward manner to vector-valued distortion functions and vector distortions

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{d}_j(X_i, Y_i, W_i, Z_i) < \boldsymbol{D}_j, \qquad j \in \{1, \dots, J\}, \tag{53}$$

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 (53) is meant component-wise. The difference between J scalar distortion constraints as in (38) and a single J-valued vector-distortion function as in (53) 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 $(R_1, \dots, R_L)$ not satisfying the above conditions, for any sequences $\{\delta_{j,n}\}$ satisfying (47). Using standard arguments, it can be shown that for any rate-tuple $(R_1, \dots, R_L)$ satisfying constraints (48) there exist excess probabilities $\{\delta_{j,n}\}$ all tending to 0 as $n \to \infty$ and so that the the rate-tuple $(R_1, \dots, R_L)$ 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 (39). Choose a sequence of positive real numbers $\{\mu_n\}$ satisfying (7),

and the set

$$\begin{aligned}\mathcal{D}_n := \Big\{ &(x^n, y^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{XY}) \colon \\ &d_j^{(n)} \left(x^n, y^n,\ g_X^{(n)}(x^n, m_1^L),\ g_Y^{(n)}\left(y^n, m_1^L\right)\right) \leq D_j, \\ &\qquad\qquad\qquad j \in \{1, \dots, J\} \Big\},\end{aligned} \tag{54}$$

where we define

$$d_j^{(n)}\left(x^n, y^n, w^n, z^n\right) := \frac{1}{n}\sum_{i=1}^{n} d_j(x_i, y_i, w_i, z_i), \tag{55}$$

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

Define also the probability

$$\Delta_n := \Pr[(X^n, Y^n) \in \mathcal{D}_n] \tag{56}$$

and notice that by the union bound and the two bounds (39) and [29, Remark to Lemma 2.12]:

$$\Delta_n \geq 1 - \sum_{j=1}^{J} \delta_{j,n} - \frac{|\mathcal{X}||\mathcal{Y}|}{4\mu^2 n}, \tag{57}$$

which by assumptions (47) and (7b) satisfies

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_n = 0. \tag{58}$$

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

$$P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) = \frac{P_{XY}^{\otimes n}(x^n, y^n)}{\Delta_n} \cdot \mathbb{1}\{(x^n, y^n) \in \mathcal{D}_n\}. \tag{59}$$

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

$$\tilde{M}_\ell = \phi_\ell^{(n)}(\tilde{X}^n, \tilde{M}_1, \ldots, \tilde{M}_{\ell-1}), \qquad \ell = 1, 3, 5, \ldots, \tag{60}$$

$$\tilde{M}_\ell = \phi_\ell^{(n)}(\tilde{Y}^n, \tilde{M}_1, \ldots, \tilde{M}_{\ell-1}), \qquad \ell = 2, 4, 6, \ldots. \tag{61}$$

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

$$U_1 := (\tilde{X}^{T-1}, \tilde{Y}_{T+1}^n, \tilde{M}_1, T) \tag{62a}$$

$$U_\tau := \tilde{M}_\tau, \qquad \tau \in \{2, \ldots, L\}. \tag{62b}$$

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

$$\begin{aligned}
&\frac{1}{n} H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_\ell) \\
&\overset{(d)}{=} \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_\ell) \\
&\qquad + \sum_{\substack{\tau\in\{1,\ldots,\ell\}:\\ \tau \text{ odd}}} I(\tilde{M}_\tau;\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_{\tau-1}) \\
&\qquad + \sum_{\substack{\tau\in\{2,\ldots,\ell-1\}:\\ \tau \text{ even}}} I(\tilde{M}_\tau;\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\tau-1})\Big] && (63) \\
&= \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_\ell) - H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_\ell)\Big] \\
&\quad + \frac{1}{n}\Big[H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1}) - H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})\Big] \\
&\quad + \cdots \\
&\quad + \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1) - H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1)\Big] \\
&\quad + \frac{1}{n} H(\tilde{Y}^n|\tilde{X}^n) && (64) \\
&\overset{(e)}{=} H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_\ell) - H(\tilde{Y}_T|\tilde{X}_T U_1\cdots U_\ell) \\
&\quad + H(\tilde{Y}_T|\tilde{X}_T U_1\cdots U_{\ell-1}) - H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{\ell-1}) \\
&\quad + \cdots \\
&\quad + H(\tilde{X}_T|\tilde{Y}_T U_1) - H(\tilde{Y}_T|\tilde{X}_T U_1) \\
&\quad + \frac{1}{n} H(\tilde{Y}^n|\tilde{X}^n) && (65) \\
&\overset{(f)}{=} H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_\ell) - H(\tilde{Y}_T|\tilde{X}_T U_1\cdots U_\ell) \\
&\quad + H(\tilde{Y}_T|\tilde{X}_T U_1\cdots U_{\ell-1}) - H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{\ell-1}) \\
&\quad + \cdots \\
&\quad + H(\tilde{X}_T|\tilde{Y}_T U_1) - H(\tilde{Y}_T|\tilde{X}_T U_1)
\end{aligned}$$

$$
\begin{aligned}
&+H(\tilde{Y}_T|\tilde{X}_T)+o(1) &(66)\\
&\stackrel{(g)}{=} H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_\ell)\\
&\quad+\sum_{\substack{\tau\in\{1,\dots,\ell\}:\\ \tau \text{ odd}}} I(U_\tau;\tilde{Y}_T|\tilde{X}_T U_1\cdots U_{\tau-1})\\
&\quad+\sum_{\substack{\tau\in\{2,\dots,\ell-1\}:\\ \tau \text{ even}}} I(U_\tau;\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{\tau-1})+o(1) &(67)\\
&\stackrel{(h)}{\geq} H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_\ell)+o(1), &(68)
\end{aligned}
$$

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 Lemma 2 in Section II and Definitions (62);
- (f) holds by Lemma 1 in Section II, where we also used Equation (58);
- (g) holds by dividing the entropy terms between sums for τ odd and even and by definition of the mutual information; and
- (h) holds by the non-negativity of mutual information.

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

$$
\frac{1}{n}H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_\ell)\geq H(\tilde{Y}_T|\tilde{X}_T U_1\cdots U_\ell)+o(1). \tag{69}
$$

We now apply bounds (68) and (69) 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

$$
\begin{aligned}
R_\ell &\geq \frac{1}{n}H(\tilde{M}_\ell)\geq\frac{1}{n}H(\tilde{M}_\ell|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1}) &(70)\\
&= \frac{1}{n}I(\tilde{M}_\ell;\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})\\
&= \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})-H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_\ell)\Big] &(71)\\
&\stackrel{(h)}{\geq} \frac{1}{n}\big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-2})\\
&\qquad -\sum_{i=1}^{n}H(\tilde{X}_i|\tilde{Y}_i\tilde{X}^{i-1}\tilde{Y}_{i+1}^n\tilde{M}_1\cdots\tilde{M}_\ell)\big] &(72)\\
&\stackrel{(i)}{\geq} H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{\ell-2})-H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_\ell)\\
&\quad+o(1) &(73)\\
&= I(U_{\ell-1}U_\ell;\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{\ell-2})+o(1) &(74)\\
&\geq I(U_\ell;\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{\ell-1})+o(1), &(75)
\end{aligned}
$$

where (h) holds because for ℓ odd message $\tilde{M}_{\ell-1}$ is a function of the tuple $(\tilde{Y}^n,\tilde{M}_1,\dots,\tilde{M}_{\ell-2})$ and because conditioning can only reduce entropy; and (i) holds by (62) and (68). Notice that for $\ell=1$:

$$
R_1\geq I(U_1;\tilde{X}_T|\tilde{Y}_T)+o(1). \tag{76}
$$

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

$$
\begin{aligned}
R_\ell &\geq \frac{1}{n}H(\tilde{M}_\ell) &(77)\\
&\geq \frac{1}{n}I(\tilde{M}_\ell;\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1}) &(78)\\
&\stackrel{(j)}{=} \frac{1}{n}\Big[H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_{\ell-2})-H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_\ell)\Big] &(79)\\
&\stackrel{(k)}{\geq} H(\tilde{Y}_T|\tilde{X}_T U_1\cdots U_{\ell-2})-H(\tilde{Y}_T|\tilde{X}_T U_1\cdots U_\ell)
\end{aligned}
$$

$$
\begin{aligned}
&+o(1) && (80)\\
&\geq I(U_\ell;\tilde{Y}_T|\tilde{X}_T U_1\cdots U_{\ell-1})+o(1) && (81)
\end{aligned}
$$

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 (62) and (69).

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

$$
\begin{aligned}
0 &= \frac{1}{n} I(\tilde{M}_\ell;\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})\\
&= \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})-H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_\ell)\Big]\\
&\overset{(l)}{\geq} H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{\ell-1})+o(1)\\
&\quad -\frac{1}{n}\sum_{i=1}^{n} H(\tilde{X}_i|\tilde{X}^{i-1}\tilde{Y}_i\tilde{Y}_{i+1}^n\tilde{M}_1\cdots\tilde{M}_\ell) && (82)\\
&= I(U_\ell;\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{\ell-1})+o(1), && (83)
\end{aligned}
$$

where (l) holds by (68) and because conditioning can only reduce entropy.

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

$$
\begin{aligned}
0 &= \frac{1}{n} I(\tilde{M}_\ell;\tilde{Y}^n|\tilde{X}^n\cdots\tilde{M}_{\ell-1})\\
&= \frac{1}{n}\Big[H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_{\ell-1})-H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_\ell)\Big]\\
&\geq I(U_\ell;\tilde{Y}_T|\tilde{X}_T U_1\cdots U_{\ell-1})+o(1). && (84)
\end{aligned}
$$

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

$$
0=\frac{1}{n}I(\tilde{M}_1;\tilde{Y}^n|\tilde{X}^n)\geq I(U_1;\tilde{Y}_T|\tilde{X}_T)+o(1). \tag{85}
$$

Let now $\tilde{W}^n := g_X(\tilde{X}^n,\tilde{M}_1,\ldots,\tilde{M}_L)$ and $\tilde{Z}^n := g_Y(\tilde{Y}^n,\tilde{M}_1,\ldots,\tilde{M}_L)$. Since the set $\mathcal{D}_n$ only contains sequences satisfying all J distortion constraints, the quadruple $(\tilde{X}^n,\tilde{Y}^n,\tilde{W}^n,\tilde{Z}^n)$ satisfies each of the J distortion constraints with probability 1. Therefore, we have for any $j\in\{1,\ldots,J\}$:

$$
\begin{aligned}
D_j &\geq \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[d_j\left(\tilde{X}_i,\tilde{Y}_i,\tilde{W}_i,\tilde{Z}_i\right)\right] && (86)\\
&= \mathbb{E}\left[d_j\left(\tilde{X}_T,\tilde{Y}_T,\tilde{W}_T,\tilde{Z}_T\right)\right], && (87)
\end{aligned}
$$

where the equality holds simply by the definition of T and the total law of expectation.

For simplicity, in the sequel we assume that L is even; if L is odd the proof is similar. Similarly to (83) and (84), 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:

$$
\begin{aligned}
0 &= \frac{1}{n} I(\tilde{Z}^n;\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_L)\\
&\overset{(m)}{=} \frac{1}{n}\Big[H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_{L-1})-H(\tilde{X}^n|\tilde{Y}^n\tilde{M}_1\cdots\tilde{M}_L\tilde{Z}^n)\Big]\\
&\overset{(n)}{\geq} H(\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{L-1})+o(1)\\
&\quad -\frac{1}{n}\sum_{i=1}^{n} H(\tilde{X}_i|\tilde{X}^{i-1}\tilde{Y}_i\tilde{Y}_{i+1}^n\tilde{M}_1\cdots\tilde{M}_L\tilde{Z}_i) && (88)\\
&= I(U_L\tilde{Z}_T;\tilde{X}_T|\tilde{Y}_T U_1\cdots U_{L-1})+o(1)\\
&\geq I(\tilde{Z}_T;\tilde{X}_T|\tilde{Y}_T U_1\cdots U_L)+o(1) && (89)
\end{aligned}
$$

and

$$
\begin{aligned}
0 &= \frac{1}{n} I(\tilde{W}^n;\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_L)\\
&= \frac{1}{n}\Big[H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_L)-H(\tilde{Y}^n|\tilde{X}^n\tilde{M}_1\cdots\tilde{M}_L\tilde{W}^n)\Big]\\
&\geq I(\tilde{W}_T;\tilde{Y}_T|\tilde{X}_T U_1\cdots U_L)+o(1), && (90)
\end{aligned}
$$

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 (68) since $L-1$ is odd and because conditioning can only reduce entropy.

The desired rate constraints are then obtained by combining (75), (76), (81), (83), (84), (85), (87), (89), and (90) and by taking $n \to \infty$. Details are as follows. By Carathéodory's theorem [37, Appendix C], there exist auxiliary random variables $U_1, \ldots, U_L$ of bounded alphabets satisfying (75), (76), (81), (83), (84), (85), (87), (89), and (90). 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

$$\lim_{i \to \infty} P_{U_1 \cdots U_L \tilde{X}\tilde{Y}\tilde{W}\tilde{Z};n_i} = P^*_{U_1 \cdots U_L XYWZ}, \tag{91}$$

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 $\left(\tilde{X}^{n_i}, \tilde{Y}^{n_i}\right)$ lies in the jointly typical set $\mathcal{T}^{(n_i)}_{\mu_{n_i}}(P_{XY})$, i.e., $\left|P_{\tilde{X}\tilde{Y};n_i} - P_{XY}\right| \leq \mu_{n_i}$, and thus since $\mu_n \to 0$ as $n \to \infty$, by the definition of $(\tilde{X}_T, \tilde{Y}_T)$ and by (91), the limiting pmf satisfies $P^*_{XY} = P_{XY}$. We further deduce from (75), (76), (81), (83), (84), (85), (87), (89), and (90) that:

$$R_\ell \geq I_{P^*}(X; U_\ell | Y U_1 \cdots U_{\ell-1}), \qquad \ell = 1, 3 \ldots \tag{92a}$$
$$R_\ell \geq I_{P^*}(Y; U_\ell | X U_1 \cdots U_{\ell-1}), \qquad \ell = 2, 4 \ldots \tag{92b}$$
$$0 = I_{P^*}(Y; U_\ell | X U_1 \cdots U_{\ell-1}), \qquad \ell = 1, 3 \ldots \tag{92c}$$
$$0 = I_{P^*}(X; U_\ell | Y U_1 \cdots U_{\ell-1}), \qquad \ell = 2, 4 \ldots \tag{92d}$$
$$0 = I_{P^*}(Z; X | Y U_1 \cdots U_L), \tag{92e}$$
$$0 = I_{P^*}(W; Y | X U_1 \cdots U_L), \tag{92f}$$

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

Combined with (87), which implies

$$D_j \geq \mathbb{E}_{P^*}[d_j(X, Y, W, Z)], \quad j \in \{1, \ldots, J\}, \tag{93}$$

above (in)equalities (92) conclude the desired converse proof.

V. Testing Against Independence in a K-Hop Network

In this section we focus on the K-hop hypothesis testing setup in Figure 3.

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 $\mathcal{H} \in \{0, 1\}$:

$$\text{if } \mathcal{H} = 0 : (Y_0^n, Y_1^n, \ldots, Y_K^n) \text{ i.i.d. } \sim P_{Y_0 Y_1} P_{Y_2|Y_1} \cdots P_{Y_K|Y_{K-1}}; \tag{94a}$$
$$\text{if } \mathcal{H} = 1 : (Y_0^n, Y_1^n, \ldots, Y_K^n) \text{ i.i.d. } \sim P_{Y_0} \cdot P_{Y_1} \cdots P_{Y_K}. \tag{94b}$$

Communication takes place over K hops as illustrated in Figure 3. 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

$$\phi_0^{(n)} \colon \mathcal{Y}_0^n \to \{1, \ldots, 2^{nR_1}\}, \tag{95}$$
$$\phi_k^{(n)} \colon \mathcal{Y}_k^n \times \{1, \ldots, 2^{nR_k}\} \to \{1, \ldots, 2^{nR_{k+1}}\}, \qquad k \in \{1, \ldots, K-1\}, \tag{96}$$

and messages are obtained as:

$$M_1 = \phi_0^{(n)}(Y_0^n) \tag{97}$$
$$M_{k+1} = \phi_k^{(n)}(Y_k^n, M_k), \quad k \in \{1, \ldots, K-1\}. \tag{98}$$

Each relay $\mathrm{R}_1, \ldots, \mathrm{R}_{K-1}$ as well as the receiver R_K, produces a guess of the hypothesis $\mathcal{H}$. These guesses are described by guessing functions

$$g_k^{(n)} \colon \mathcal{Y}_k^n \times \{1, \ldots, 2^{nR_k}\} \to \{0, 1\}, \quad k \in \{1, \ldots, K\}, \tag{99}$$

where we request that the guesses

$$\hat{\mathcal{H}}_k = g_k^{(n)}(Y_k^n, M_k), \quad k \in \{1, \ldots, K\}, \tag{100}$$

have type-I error probabilities

$$\alpha_{k,n} \triangleq \Pr[\hat{\mathcal{H}}_k = 1 | \mathcal{H} = 0], \quad k \in \{1, \ldots, K\}, \tag{101}$$

not exceeding given thresholds, and type-II error probabilities

$$\beta_{k,n} \triangleq \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1], \quad k \in \{1, \ldots, K\}, \tag{102}$$

decaying to 0 exponentially fast with largest possible exponents.

Definition 3: Given sequences of allowed type-I error probabilities $\{\delta_{k,n}\}$ and rates $R_1, R_2, \ldots, R_K \geq 0$, the exponent tuple $(\theta_1, \theta_2, \ldots, \theta_K)$ is called $\{\delta_{k,n}\}$-*achievable* if there exists a sequence of encoding and decision functions $\{\phi_0^{(n)}, \phi_1^{(n)}, \ldots, \phi_{K-1}^{(n)}, g_1^{(n)}, g_2^{(n)}, \ldots, g_K^{(n)}\}_{n \geq 1}$ satisfying for each $k \in \{1, \ldots, K\}$ and blocklength n:

$$\alpha_{k,n} \leq \delta_{k,n}, \tag{103a}$$

$$\varliminf_{n \to \infty} \frac{1}{n} \log \frac{1}{\beta_{k,n}} \geq \theta_k. \tag{103b}$$

B. Old and New Results

Definition 4: For any $\ell \in \{1, \ldots, K\}$, define the function

$$\eta_\ell \colon \mathbb{R}_0^+ \to \mathbb{R}_0^+ \tag{104}$$

$$R \mapsto \max_{\substack{P_{U|Y_{\ell-1}}: \\ R \geq I(U; Y_{\ell-1})}} I(U; Y_\ell). \tag{105}$$

The described setup was previously studied in [22] and [8], and an extension of the setup under variable-length coding was considered in [38]. While for a general number of users $K \geq 2$ only achievability results and weak converses were presented [22], for $K = 2$ users a strong converse was derived.

Theorem 3 (Theorems 2 and 3 in [8]): Let $K = 2$ and consider fixed allowed type-I error probabilities

$$\delta_{k,n} = \epsilon_k, \quad k \in \{1, 2\}, \tag{106}$$

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,

$$\theta_k \leq \sum_{\ell=1}^{k} \eta_\ell(R_\ell), \qquad k \in \{1, 2\}. \tag{107}$$

Remark 5: In [8], the presentation of Theorem 3 was split into two separate theorems (Theorems 2 and 3 in [8]) depending on the values of ϵ_1 and ϵ_2. While [8, Theorem 2] considers the case $\epsilon_1 + \epsilon_2 < 1$ and coincides with above formulation, [8, 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 [8, Theorem 3] simplifies to the form presented in above Theorem 3. This observation is important to note that our main result in this section, Theorem 4 ahead, is not only more general, but also consistent with the existing results in [8].

Remark 6: The set of pairs (θ_1, θ_2) that are (ϵ_1, ϵ_2) achievable according to Theorem 3 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 [22] and by the strong converse result in the following Theorem 4.

Our main result in this section (Theorem 4 ahead) generalizes the strong converse in Theorem 3 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 (109) can only be achieved with probabilities $\alpha_{k,n}$ that tend to 1 exponentially fast in the blocklength n.

Theorem 4: Let $\{\delta_{k,n}\}$ be sequences satisfying

$$\lim_{n \to \infty} \frac{1}{n} \log(1 - \delta_{k,n}) = 0, \qquad k \in \{1, \ldots, K\}. \tag{108}$$

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

$$\theta_k \leq \sum_{\ell=1}^{k} \eta_\ell(R_\ell), \quad k \in \{1, \ldots, K\}. \tag{109}$$

Remark 7: The direct part of this theorem was proved in [22] 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 (108) and do not vanish too quickly.

Remark 8: For all permissible type-I error probabilities $\{\delta_{k,n}\}$ that satisfy (108) 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 4, it thus suffices to show that the bound in (109) holds in a setup where only the single decision center R_k takes a decision.

Remark 9: 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 [38]. Moreover, as also shown in [38], 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 4

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 (103) for $\{\delta_{k,n}\}$ and type-II error exponents $\theta_1, \ldots, \theta_K$.

Choose a sequence of small positive numbers $\{\mu_n\}_{n=1}^{\infty}$ satisfying (7). 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.,

$$\mathcal{A}_k := \left\{(y_0^n, \ldots, y_k^n) \colon g_k^{(n)}(y_k^n, m_k) = 0\right\}, \tag{110}$$

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

$$m_\ell := \phi_{\ell-1}^{(n)}(y_{\ell-1}^n, m_{\ell-1}), \qquad \ell \in \{2, \ldots, k\}. \tag{111}$$

Define also the law under $\mathcal{H} = 0$:

$$P_{Y_0 \cdots Y_k} = P_{Y_0 Y_1} P_{Y_2|Y_1} \cdots P_{Y_k|Y_{k-1}}, \tag{112}$$

and the intersection of this acceptance region with the typical set:

$$\mathcal{D}_k \triangleq \mathcal{A}_k \cap \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 \cdots Y_k}). \tag{113}$$

By [29, Remark to Lemma 2.12], the type-I error probability constraints in (103a), and the union bound:

$$\Delta_k := P_{Y_0^n Y_1^n \cdots Y_k^n}(\mathcal{D}_k) \geq 1 - \delta_{k,n} - \frac{|\mathcal{Y}_0| \cdots |\mathcal{Y}_k|}{4\mu_n^2 n}, \tag{114}$$

and thus

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_k = 0. \tag{115}$$

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

$$\begin{aligned} &P_{\tilde{Y}_0^n \tilde{Y}_1^n \cdots \tilde{Y}_k^n}(y_0^n, y_1^n, \ldots, y_k^n) \\ &= \frac{P_{Y_0^n Y_1^n \cdots Y_k^n}(y_0^n, y_1^n, \ldots, y_k^n)}{\Delta_k} \cdot \mathbb{1}\{(y_0^n, y_1^n, \ldots, y_k^n) \in \mathcal{D}_k\}, \end{aligned} \tag{116}$$

and notice that for each $\ell \in \{0, 1, \ldots, k\}$:

$$P_{\tilde{Y}_\ell^n}(y_\ell^n) = \frac{P_{Y_\ell^n}(y_\ell^n)}{\Delta_k} \tag{117}$$

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 $\ell \in \{1, \ldots, k\}$:

$$
\begin{aligned}
R_\ell &\geq \frac{1}{n} H(\tilde{M}_\ell) && (118)\\
&= \frac{1}{n} I(\tilde{M}_\ell; \tilde{Y}_0^n \cdots \tilde{Y}_k^n) && (119)\\
&= \frac{1}{n} H(\tilde{Y}_0^n \cdots \tilde{Y}_k^n) - \frac{1}{n} H(\tilde{Y}_0^n \cdots \tilde{Y}_k^n | \tilde{M}_\ell) && (120)\\
&= H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T}) + o(1) \\
&\quad - \frac{1}{n} \sum_{t=1}^{n} H(\tilde{Y}_{0,t} \cdots \tilde{Y}_{k,t} | \tilde{M}_\ell \tilde{Y}_0^{t-1} \cdots \tilde{Y}_k^{t-1}) && (121)\\
&= H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T}) + o(1) - H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T} | U_\ell) && (122)\\
&= I(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T}; U_\ell) + o(1) && (123)\\
&\geq I(\tilde{Y}_{\ell-1,T}; U_\ell) + o(1), && (124)
\end{aligned}
$$

where we defined $U_\ell \triangleq (\tilde{M}_\ell, \tilde{Y}_0^{T-1}, \ldots, \tilde{Y}_k^{T-1}, T)$. Here, (121) holds by extending (12) to k-tuples.

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

$$
\begin{aligned}
Q_{\tilde{M}_k}(m_k) \triangleq \sum_{y_0^n, y_1^n, \ldots, y_{k-1}^n} & P_{\tilde{Y}_0^n}(y_0^n) \cdots P_{\tilde{Y}_{k-1}^n}(y_{k-1}^n) \\
& \cdot \mathbb{1}\{m_k = \phi_k(\phi_{k-1}(\cdots(\phi_1(y_0^n)\cdots)), y_{k-1}^n)\},
\end{aligned} \tag{125}
$$

and

$$
\begin{aligned}
Q_{M_k}(m_k) \triangleq \sum_{y_0^n, y_1^n, \ldots, y_{k-1}^n} & P_{Y_0^n}(y_0^n) \cdots P_{Y_{k-1}^n}(y_{k-1}^n) \\
& \cdot \mathbb{1}\{m_k = \phi_{k-1}(\phi_{k-2}(\cdots(\phi_0(y_0^n)\cdots)), y_{k-1}^n)\},
\end{aligned} \tag{126}
$$

and notice that by (117) and (125):

$$
\begin{aligned}
Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k) &\leq Q_{M_k} P_{Y_k^n}(\mathcal{A}_k)\, \Delta_k^{-(k+1)} \\
&= \beta_{k,n} \Delta_k^{-(k+1)}.
\end{aligned} \tag{127}
$$

Moreover, by (110), the probability $P_{\tilde{M}_k \tilde{Y}_k^n}(\mathcal{A}_k) = 1$, and thus

$$
D\left(P_{\tilde{M}_k \tilde{Y}_k^n}(\mathcal{A}_k) \| Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right) = -\log\left(Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right), \tag{128}
$$

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 (115), with (127), and with the data-processing inequality, we obtain from (128):

$$
\begin{aligned}
-\frac{1}{n} \log \beta_{k,n} &\leq -\frac{1}{n} \log\left(Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right) - \frac{(k+1)}{n} \log \Delta_k && (129)\\
&\leq \frac{1}{n} D\left(P_{\tilde{M}_k \tilde{Y}_k^n} \middle\| Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}\right) + o(1). && (130)
\end{aligned}
$$

We continue to upper bound the divergence term as

$$
\begin{aligned}
&\frac{1}{n} D(P_{\tilde{M}_k \tilde{Y}_k^n} || Q_{\tilde{M}_k} P_{\tilde{Y}_k^n}) \\
&\quad = \frac{1}{n} I(\tilde{M}_k; \tilde{Y}_k^n) + \frac{1}{n} D(P_{\tilde{M}_k} || Q_{\tilde{M}_k}) && (131)\\
&\quad \leq \frac{1}{n} I(\tilde{M}_k; \tilde{Y}_k^n) + \frac{1}{n} D(P_{\tilde{Y}_{k-1}^n \tilde{M}_{k-1}} || P_{\tilde{Y}_{k-1}^n} Q_{\tilde{M}_{k-1}}) && (132)\\
&\quad \leq \frac{1}{n} I(\tilde{M}_k; \tilde{Y}_k^n) + \frac{1}{n} I(\tilde{M}_{k-1}; \tilde{Y}_{k-1}^n) \\
&\qquad\qquad + \frac{1}{n} D(P_{\tilde{Y}_{k-2}^n \tilde{M}_{k-2}} || P_{\tilde{Y}_{k-2}^n} Q_{\tilde{M}_{k-2}}) && (133)
\end{aligned}
$$

$$\vdots$$

$$\leq \frac{1}{n}\sum_{\ell=2}^{k} I(\tilde{M}_\ell;\tilde{Y}_\ell^n) + \frac{1}{n}D(P_{\tilde{Y}_1^n\tilde{M}_1}||P_{\tilde{Y}_1^n}Q_{\tilde{M}_1}) \tag{134}$$

$$= \frac{1}{n}\sum_{\ell=1}^{k} I(\tilde{M}_\ell;\tilde{Y}_\ell^n) \tag{135}$$

$$\leq \frac{1}{n}\sum_{\ell=1}^{k}\sum_{t=1}^{n} I(\tilde{M}_\ell\tilde{Y}_0^{t-1}\cdots\tilde{Y}_k^{t-1};\tilde{Y}_{\ell,t}) \tag{136}$$

$$\leq \sum_{\ell=1}^{k} I(U_\ell;\tilde{Y}_{\ell,T}). \tag{137}$$

Here

- (132) is obtained by the data processing inequality for KL-divergence and because $\tilde{M}_k$ is a function of $\tilde{M}_{k-1}$ and $\tilde{Y}_k^n$;
- (133) is obtained by applying the same arguments as leading to (131) and (132), but now to the pair $(\tilde{M}_{k-1},\tilde{Y}_{k-1})$ instead of $(\tilde{M}_k,\tilde{Y}_k)$;
- (134) is obtained by iteratively applying the same arguments as leading to (131) and (132) to the pairs $(\tilde{M}_{k-2},\tilde{Y}_{k-2}),\ldots,(\tilde{M}_2,\tilde{Y}_2)$;
- (135) holds because $P_{\tilde{M}_1} = Q_{\tilde{M}_1}$ and thus $D(P_{\tilde{Y}_1^n\tilde{M}_1}||P_{\tilde{Y}_1^n}Q_{\tilde{M}_1}) = I(\tilde{Y}_1;\tilde{M}_1)$; and
- (137) holds by the definition of U_ℓ and T.

Combined with (130), we obtain

$$-\frac{1}{n}\log\beta_{k,n} \leq \sum_{\ell=1}^{k} I(U_\ell;\tilde{Y}_{\ell,T}) + o(1). \tag{138}$$

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:

$$\begin{aligned}
0 &= \frac{1}{n}I(\tilde{M}_1;\tilde{Y}_1^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n) && (139)\\
&= \frac{1}{n}H(\tilde{Y}_1^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n) - \frac{1}{n}H(\tilde{Y}_1^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n\tilde{M}_1) && (140)\\
&= H(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}) + o(1) - \frac{1}{n}H(\tilde{Y}_1^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n\tilde{M}_1) && (141)\\
&\geq H(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}) + o(1)\\
&\quad - H(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}\tilde{Y}_0^{T-1}\cdots\tilde{Y}_k^{T-1}\tilde{Y}_{0,T+1}^n\tilde{M}_1T) && (142)\\
&\geq I(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T};U_1|\tilde{Y}_{0,T}) + o(1) \geq 0, && (143)
\end{aligned}$$

where (141) is obtained by extending (14) to multiple random variables. We thus conclude that

$$\lim_{n\to\infty} I(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T};U_1|\tilde{Y}_{0,T}) = 0. \tag{144}$$

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

$$\begin{aligned}
I(U_\ell;\tilde{Y}_{\ell,T}|\tilde{Y}_{\ell-1,T}) &\leq I(U_\ell\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-2,T};\tilde{Y}_{\ell,T}|\tilde{Y}_{\ell-1,T}) && (145)\\
&= I(U_\ell;\tilde{Y}_{\ell,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-1,T})\\
&\quad + I(\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-2,T};\tilde{Y}_{\ell,T}|\tilde{Y}_{\ell-1,T}) && (146)\\
&= I(U_\ell;\tilde{Y}_{\ell,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-1,T}) + o(1), && (147)
\end{aligned}$$

where the last equality can be proved by extending (12) and (14) 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 (139)–(143), we further obtain for each $\ell \in \{1,\ldots,k\}$:

$$0 = \frac{1}{n}I(\tilde{M}_\ell;\tilde{Y}_\ell^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n\cdots\tilde{Y}_{\ell-1}^n) \tag{148}$$

$$
\begin{aligned}
&= \frac{1}{n} H(\tilde{Y}_\ell^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \cdots \tilde{Y}_{\ell-1}^n) \\
&\quad - \frac{1}{n} H(\tilde{Y}_\ell^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \cdots \tilde{Y}_{\ell-1}^n \tilde{M}_\ell) && (149)\\
&= H(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) \\
&\quad - \frac{1}{n} H(\tilde{Y}_\ell^n \cdots \tilde{Y}_k^n | \tilde{Y}_0^n \cdots \tilde{Y}_{\ell-1}^n \tilde{M}_\ell) && (150)\\
&\geq H(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) \\
&\quad - \frac{1}{n} \sum_{t=1}^{n} H(\tilde{Y}_{\ell,t} \cdots \tilde{Y}_{k,t} | \tilde{Y}_{0,t} \cdots \tilde{Y}_{\ell-1,t} \tilde{Y}_0^{t-1} \cdots \tilde{Y}_k^{t-1} \tilde{M}_\ell) && (151)\\
&= H(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) \\
&\quad - H(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T} \tilde{Y}_0^{T-1} \cdots \tilde{Y}_k^{T-1} \tilde{M}_\ell T) && (152)\\
&= I(\tilde{Y}_{\ell,T} \cdots \tilde{Y}_{k,T}; U_\ell | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) && (153)\\
&\geq I(\tilde{Y}_{\ell,T}; U_\ell | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) + o(1) \geq 0. && (154)
\end{aligned}
$$

We thus conclude that

$$
I(U_\ell; \tilde{Y}_{\ell,T} | \tilde{Y}_{0,T} \cdots \tilde{Y}_{\ell-1,T}) = o(1), \tag{155}
$$

which combined with (147) proves

$$
I(U_\ell; \tilde{Y}_{\ell,T} | \tilde{Y}_{\ell-1,T}) = o(1). \tag{156}
$$

The converse is then concluded by taking $n \to \infty$, as we explain in the following. By Carathéodory's theorem [37, Appendix C], for each n there must exist random variables $U_1, \ldots, U_k$ satisfying (156), (138), and (124) over alphabets of sizes

$$
|\mathcal{U}_\ell| \leq |\mathcal{Y}_{\ell-1}| \cdot |\mathcal{Y}_\ell| + 2, \qquad \ell \in \{1, \ldots, k\}. \tag{157}
$$

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

$$
\lim_{i \to \infty} P_{\tilde{Y}_{\ell-1} \tilde{Y}_\ell U_\ell; n_i} = P^{(\ell)}_{Y_{\ell-1} Y_\ell U_\ell}, \quad \ell \in \{1, \ldots, k\}, \tag{158}
$$

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

$$
R_\ell \geq I_{P^{(\ell)}}(U_\ell; Y_{\ell-1}), \quad \ell \in \{1, \ldots, k\}, \tag{159}
$$

$$
\theta_k \leq \sum_{\ell=1}^{k} I_{P^{(\ell)}}(U_\ell; Y_\ell), \tag{160}
$$

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 $\mathcal{T}_{\mu_{n_i}}^{(n_i)}(P_{Y_{\ell-1} Y_\ell})$, 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 (156), for all $\ell \in \{1, \ldots, k\}$ the Markov chain

$$
U_\ell \to Y_{\ell-1} \to Y_\ell, \tag{161}
$$

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

VI. Communication over a Memoryless Channel

This section studies communication over a discrete memoryless channel (DMC) as depicted in Figure 4.

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 $\mathcal{M} := \{1, \ldots, 2^{nR}\}$ 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

$$X^n = \phi^{(n)}(M) \tag{162}$$

for some choice of the encoding function $\phi^{(n)} \colon \mathcal{M} \to \mathcal{X}^n$, 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:

$$\hat{M} = g^{(n)}(Y^n) \tag{163}$$

using a decoding function of the form $g^{(n)} \colon \mathcal{Y}^n \to \mathcal{M}$. The goal is to minimize the average decoding error probability

$$p^{(n)}(\text{error}) := \Pr\left[\hat{M} \neq M\right]. \tag{164}$$

Definition 5: 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

$$p^{(n)}(\text{error}) \leq \delta_n. \tag{165}$$

A well-known result in information theory states [29]:

Theorem 5: Any rate $R > C$, where C denotes the capacity

$$C := \max_{P_X} I(X; Y), \tag{166}$$

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

$$\lim_{n\to\infty} \frac{1}{n} \log(1 - \delta_n) = 0. \tag{167}$$

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 (165) holds. Choose a sequence of small positive numbers $\{\mu_n\}$ satisfying (7) and define the set

$$\mathcal{D}_n := \left\{(m, y^n) \colon y^n \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y|X=x^n(m)}) \text{ and } g^{(n)}(y^n) = m\right\} \tag{168}$$

and its probability

$$\Delta_n := \Pr[(M, Y^n) \in \mathcal{D}_n]. \tag{169}$$

By the union bound followed by (165) and the bound on the probability of the typical set derived in [29, Remark to Lemma 2.12], we have:

$$\Delta_n \geq 1 - \delta_n - \frac{|\mathcal{Y}||\mathcal{X}|}{4\mu_n^2 n}, \tag{170}$$

and thus by (167) and (7b):

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_n = 0. \tag{171}$$

Let further $(\tilde{M}, \tilde{X}^n, \tilde{Y}^n)$ be random variables so that

$$P_{\tilde{M}\tilde{X}\tilde{Y}^n}(m, x^n, y^n) = \frac{1}{2^{nR}} \cdot \frac{P_{Y|X}^{\otimes n}(y^n|\phi^n(m))}{\Delta_n} \cdot \mathbb{1}\{(m, y^n) \in \mathcal{D}_n\} \cdot \mathbb{1}\left\{x^n = \phi^{(n)}(m)\right\}. \tag{172}$$

Further, let T be independent of $(\tilde{M}, \tilde{X}^n, \tilde{Y}^n)$ and uniform over $\{1, \ldots, n\}$. By above definition

$$P_{\tilde{M}}(m) \leq \frac{1}{2^{nR}} \cdot \frac{1}{\Delta_n}, \tag{173}$$

and thus

$$\frac{1}{n}H(\tilde{M}) \geq R + \frac{1}{n}\log \Delta_n. \tag{174}$$

Moreover, since decoding sets are disjoint, by the definition of the new measure $P_{\tilde{M}\tilde{X}^n\tilde{Y}^n}$ it is possible to determine $\tilde{M}$ from $\tilde{Y}$ with probability 1. We combine these observations with similar steps as in the weak converse to:

$$\begin{aligned}
R &\leq \frac{1}{n}H(\tilde{M}) - \frac{1}{n}\log \Delta_n &(175)\\
&\overset{(a)}{=} \frac{1}{n}I(\tilde{M};\tilde{Y}^n) - \frac{1}{n}\log \Delta_n &(176)\\
&= \frac{1}{n}H(\tilde{Y}^n) - \frac{1}{n}H(\tilde{Y}^n|\tilde{M}) - \frac{1}{n}\log \Delta_n &(177)\\
&\leq \frac{1}{n}\sum_{i=1}^{n} H(\tilde{Y}_i) - \frac{1}{n}H(\tilde{Y}^n|\tilde{M}) - \frac{1}{n}\log \Delta_n &(178)\\
&= H(\tilde{Y}_T|T) - \frac{1}{n}H(\tilde{Y}^n|\tilde{M}) - \frac{1}{n}\log \Delta_n &(179)\\
&\leq H(\tilde{Y}_T) - \frac{1}{n}H(\tilde{Y}^n|\tilde{M}) - \frac{1}{n}\log \Delta_n, &(180)
\end{aligned}$$

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

By (171) and the following lemma, by considering an appropriate subsequence of blocklengths, and by the continuity of the entropy function, we deduce that

$$R \leq I_{P_X P_{Y|X}}(X;Y) \leq C, \tag{181}$$

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:[2]

$$\lim_{i\to\infty} P_{\tilde{Y}_T}(y) = \sum_{x\in\mathcal{X}} P_X(x)P_{Y|X}(y|x) \tag{182}$$

$$\lim_{i\to\infty} \frac{1}{n_i}H(\tilde{Y}^{n_i}|\tilde{M}) = H_{P_X P_{Y|X}}(Y|X), \tag{183}$$

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:

$$\begin{aligned}
P_{\tilde{Y}_T}(y) &= \frac{1}{n}\sum_{t=1}^{n} P_{\tilde{Y}_t}(y) &(184)\\
&= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{\tilde{Y}_t = y\}\right] &(185)\\
&= \mathbb{E}\left[\pi_{\tilde{Y}^n}(y)\right] &(186)\\
&= \sum_{x\in\mathcal{X}} \mathbb{E}\left[\pi_{x^n(\tilde{M})\tilde{Y}^n}(x,y)\right] &(187)
\end{aligned}$$

where the fourth equality holds because for any pair of sequences x^n, y^n we have $\sum_{x\in\mathcal{X}} \pi_{x^n y^n}(x,y) = \pi_{y^n}(y)$, and by exchanging sum and expectation. By the way we defined the set $\mathcal{D}_n$, we have for all $(m, y^n) \in \mathcal{D}_n$ that

$$\left|\pi_{x^n(m)y^n}(x,y) - \pi_{x^n(m)}(x)P_{Y|X}(y|x)\right| \leq \mu_n, \tag{188}$$

and if $P_{Y|X}(y|x) = 0$ then $\pi_{x^n(m)y^n}(x,y) = 0$. Plugging these conditions into (187) we obtain

$$P_{\tilde{Y}_T}(y) \leq \sum_{\substack{x\in\mathcal{X}:\\ P_{Y|X}(y|x)>0}} \mathbb{E}\left[\pi_{x^n(\tilde{M})}(x)\right] \cdot P_{Y|X}(y|x) + |\mathcal{X}|\mu_n \tag{189a}$$

[2]Recall that the random variable $\tilde{Y}_T$ depends on the blocklength n_i, and thus taking the limit $i \to \infty$ is well-defined.

and similarly:

$$P_{\tilde{Y}_T}(y) \geq \sum_{\substack{x\in\mathcal{X}:\\ P_{Y|X}(y|x)>0}} \mathbb{E}\left[\pi_{x^n(\tilde{M})}(x)\right] \cdot P_{Y|X}(y|x) - |\mathcal{X}|\mu_n. \tag{189b}$$

Let now $\{n_i\}$ be an increasing subsequence of blocklengths so that the sequence of expected types $\mathbb{E}\left[\pi_{x^n(\tilde{M})}(x)\right]$ 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 (190):

$$\lim_{i\to\infty} P_{\tilde{Y}_T}(y) = \sum_{x\in\mathcal{X}} P_X(x) P_{Y|X}(y|x), \tag{190}$$

establishing the first part of the lemma.

Notice next that by definition

$$\begin{aligned}
&\frac{1}{n}H(\tilde{Y}^n|\tilde{M}=m)\\
&= -\frac{1}{n}\sum_{y^n\in\mathcal{D}_m} P_{\tilde{Y}^n|\tilde{M}=m}(y^n)\log P_{\tilde{Y}^n|\tilde{M}=m}(y^n) && (191)\\
&= -\frac{1}{n}\sum_{y^n\in\mathcal{D}_m} P_{\tilde{Y}^n|\tilde{M}=m}(y^n)\log \frac{P_{Y|X}^{\otimes n}(y^n|x^n(m))}{\Delta_m} && (192)\\
&= -\frac{1}{n}\sum_{t=1}^{n}\sum_{y^n\in\mathcal{D}_m} P_{\tilde{Y}^n|\tilde{M}=m}(y^n)\log P_{Y|X}(y_t|x_t(m))\\
&\quad +\frac{1}{n}\log\Delta_m && (193)\\
&= -\frac{1}{n}\sum_{t=1}^{n}\sum_{y_t\in\mathcal{Y}} P_{\tilde{Y}_t|\tilde{M}=m}(y_t)\log P_{Y|X}(y_t|x_t(m))\\
&\quad +\frac{1}{n}\log\Delta_m && (194)\\
&= -\frac{1}{n}\sum_{t=1}^{n}\sum_{y\in\mathcal{Y}} \mathbb{E}\left[\mathbb{1}\left\{\tilde{Y}_t=y\right\}\Big|\tilde{M}=m\right]\log P_{Y|X}(y|x_t(m))\\
&\quad +\frac{1}{n}\log\Delta_m && (195)\\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\mathbb{1}\left\{x_t(m)=x,\tilde{Y}_t=y\right\}\Big|\tilde{M}=m\right]\\
&\qquad\qquad \cdot\log P_{Y|X}(y|x)\\
&\quad +\frac{1}{n}\log\Delta_m && (196)\\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbb{E}\left[\pi_{x^n(m)\tilde{Y}^n}(x,y)\Big|\tilde{M}=m\right]\log P_{Y|X}(y|x)\\
&\quad +\frac{1}{n}\log\Delta_m. && (197)
\end{aligned}$$

Taking expectation with respect to $P_{\tilde{M}}$, we obtain

$$\begin{aligned}
&\frac{1}{n}H(\tilde{Y}^n|\tilde{M}) && (198)\\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbb{E}\left[\pi_{x^n(\tilde{M})\tilde{Y}^n}(x,y)\right]\log P_{Y|X}(y|x)\\
&\quad +\frac{1}{n}\log\Delta_m. && (199)
\end{aligned}$$

By (189) and by recalling the definition of P_X as the convergence point of $\mathbb{E}\left[\pi_{x^n(\tilde{M})}(x)\right]$ for the sequence of blocklengths $\{n_i\}_{i=1}^{\infty}$, one can follow the same bounding steps as leading to (190) to obtain:

$$\lim_{i\to\infty}\frac{1}{n_i}H(\tilde{Y}^{n_i}|\tilde{M})$$

$$= -\sum_{x\in\mathcal{X}} P_X(x)P_{Y|X}(y|x)\log P_{Y|X}(y|x)$$
$$= H_{P_X P_{Y|X}}(Y|X), \tag{200}$$

which concludes the second part of the proof. ■

VII. Conclusions and Outlook

This paper presented new exponentially-strong converse proofs for source and channel coding setups and for hypothesis testing, i.e., our results allow to conclude that either the decoding or detection error probabilities or the sum of the excess-distortion probabilities tend to 1 exponentially fast whenever the rates (or error 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 [2]–[4] 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 proof steps establishing variational characterizations of multi-letter and single-letter expressions as in [4].

The results 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 [4]. A strong converse proof for the 2-hop hypothesis testing setup was already presented in [8], but not for the fully general setup and using different techniques to bound the two exponents. In our work we presented a simplified and unified proof that applies to all exponents and without further assumptions.

In related publications we show that the proof technique presented in this paper can be extended to more complicated setups including either additional expectation constraints (e.g., constraints on the expected rate or expected equivocation in a secrecy setup) or setups with mixed channel coding, reconstruction, and detection constraints. For example, in [38], [39] and [40], we used the presented proof method to derive fundamental limits of hypothesis testing systems under expected rate constraints and (expected) secrecy constraints. In contrast to the results presented in this paper, these fundamental limits depend on the allowed type-I error probabilities. It turns out that the proposed proof technique based on change of measure arguments naturally captures the dependence between the allowed error probabliities and the fundamental limits under expectation constraints. In another related work [7], we considered an integrated sensing and communication (ISAC) systems that combines channel coding with either source coding or detection. With appropriate modifications, the proof method presented in this paper could be used to derive exponentially-strong converse results also for ISAC.

Acknowledgement

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

Appendix A
Proof of Lemma 1

Notice that (15)–(17) follow directly from (11) and continuity of entropy. To prove (11), notice that

$$P_{\tilde{X}_T\tilde{Y}_T}(x,y) = \frac{1}{n}\sum_{t=1}^{n} P_{\tilde{X}_t\tilde{Y}_t}(x,y) \tag{201}$$
$$= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{\tilde{X}_t = x, \tilde{Y}_t = y\}\right] \tag{202}$$
$$= \mathbb{E}[\pi_{\tilde{X}^n\tilde{Y}^n}(x,y)], \tag{203}$$

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

$$|\pi_{\tilde{X}^n\tilde{Y}^n}(x,y) - P_{XY}(x,y)| \le \mu_n, \tag{204}$$

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 (12), notice first that

$$\begin{aligned}
&\frac{1}{n}H(\tilde{X}^n\tilde{Y}^n)+\frac{1}{n}D(P_{\tilde{X}^n\tilde{Y}^n}\|P_{XY}^{\otimes n}) \\
&=-\frac{1}{n}\sum_{(x^n,y^n)\in\mathcal{D}_n}P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)\log P_{XY}^{\otimes n}(x^n,y^n) && (205)\\
&=-\frac{1}{n}\sum_{t=1}^{n}\sum_{(x^n,y^n)\in\mathcal{D}_n}P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)\log P_{XY}(x_t,y_t) && (206)\\
&=-\frac{1}{n}\sum_{t=1}^{n}\sum_{(x_t,y_t)\in\mathcal{X}\times\mathcal{Y}}P_{\tilde{X}_t\tilde{Y}_t}(x_t,y_t)\log P_{XY}(x_t,y_t) && (207)\\
&=-\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\left(\frac{1}{n}\sum_{t=1}^{n}P_{\tilde{X}_t\tilde{Y}_t}(x,y)\right)\log P_{XY}(x,y) && (208)\\
&=-\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}}P_{\tilde{X}_T\tilde{Y}_T}(x,y)\log P_{XY}(x,y) && (209)\\
&=H(\tilde{X}_T\tilde{Y}_T)+D(P_{\tilde{X}_T\tilde{Y}_T}\|P_{XY}), && (210)
\end{aligned}$$

where (208) 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 (212) and (213) this establishes (12). The first relevant limit is

$$D(P_{\tilde{X}_T\tilde{Y}_T}\|P_{XY})\to 0, \tag{211}$$

which holds by (11) and because $P_{\tilde{X}_T\tilde{Y}_T}(x,y)=0$ whenever $P_{XY}(x,y)=0$. The second limit is:

$$\frac{1}{n}D(P_{\tilde{X}^n\tilde{Y}^n}\|P_{XY}^{\otimes n})\to 0, \tag{212}$$

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

$$\begin{aligned}
0 &\leq \frac{1}{n}D(P_{\tilde{X}^n\tilde{Y}^n}\|P_{XY}^{\otimes n}) \\
&= \frac{1}{n}\sum_{(x^n,y^n)\in\mathcal{D}_n}P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)\log\frac{P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)}{P_{XY}^{\otimes n}(x^n,y^n)} && (213)\\
&= -\frac{1}{n}\sum_{(x^n,y^n)\in\mathcal{D}_n}P_{\tilde{X}^n\tilde{Y}^n}(x^n,y^n)\log\Delta_n && (214)\\
&= -\frac{1}{n}\log\Delta_n. && (215)
\end{aligned}$$

To prove (13), notice that by the same arguments as we concluded (211), we also have

$$\frac{1}{n}H(\tilde{Y}^n)+\frac{1}{n}D(P_{\tilde{Y}^n}\|P_Y^{\otimes n})=H(\tilde{Y}_T)+D(P_{\tilde{Y}_T}\|P_Y). \tag{216}$$

Moreover, (212) and (213) imply

$$\frac{1}{n}D(P_{\tilde{Y}^n}\|P_Y^{\otimes n})\to 0 \tag{217}$$

$$D(P_{\tilde{Y}_T}\|P_Y)\to 0, \tag{218}$$

which combined with (217) imply (13).

The last limit (14) follows by the chain rule and limits (12) and (13). This concludes the proof.

References

[1] M. Hamad, M. Wigger, and M. Sarkiss, "Strong converses using change of measure and asymptotic markov chains," in *2022 IEEE Information Theory Workshop (ITW)*, pp. 535–540, 2022.

[2] 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, IEEE, 2009.

[3] 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, IEEE, 2011.

[4] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 689–703, 2019.

[5] Y. Steinberg, "Coding and common reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4995–5010, 2009.

[6] 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.
[7] M. Ahmadipour, M. Wigger, and S. Shamai, "Strong converses for memoryless bi-static isac." [Online]. Available: https://arxiv.org/abs/2303.06636, 2023.
[8] 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.
[9] A. I. Khinchin, "The entropy concept in probability theory," *Usp. Mat. Nay*, vol. 8, pp. 3–20, 1953.
[10] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, vol. 24, pp. 196–219, 1953.
[11] 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.
[12] 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.
[13] 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.
[14] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.
[15] 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.
[16] 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.
[17] Y. Oohama, "Exponential strong converse for source coding with side information at the decoder," *Entropy*, vol. 20, no. 5, p. 352, 2018.
[18] Y. Oohama, "Exponential strong converse for one helper source coding problem," *Entropy*, vol. 21, no. 6, p. 567, 2019.
[19] 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.
[20] A. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 735–740, 1985.
[21] 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.
[22] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing in multi-hop networks." [Online]. Available: https://arxiv.org/abs/1708.05198v1, 2017.
[23] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.
[24] 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.
[25] P. G. R. Ahlswede and J. Körner, "Bounds on conditional probabilities with applications in multiuser communication," *Z. Wahrscheinlichkeitstheorie und verw. Geb.*, vol. 34, pp. 157–177, 1976.
[26] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.
[27] S. Sreekumar and D. Gündüz, "Strong converse for testing against independence over a noisy channel," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1283–1288, 2020.
[28] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, vol. 1, no. 4, pp. 591 – 606, 1957.
[29] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
[30] 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.
[31] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," (Prague, Czech Replublic), pp. 689–723, 962.
[32] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (corresp.)," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 357–359, 1973.
[33] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
[34] R. Graczyk, "An elementary proof of the strong converse of the channel coding theorem." Nov. 2022.
[35] S. L. Fong and V. Y. F. 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.
[36] S. L. Fong and V. Y. F. 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.
[37] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
[38] M. Hamad, M. Wigger, and M. Sarkiss, "Multi-hop network with multiple decision centers under expected-rate constraints," *IEEE Transactions on Information Theory*, vol. 69, no. 7, pp. 4255–4283, 2023.
[39] 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.
[40] S. Faour, M. Hamad, M. Sarkiss, and M. Wigger, "Testing against independence with an eavesdropper," in *2023 IEEE Information Theory Workshop (ITW)*, pp. 277–282, 2023.