Distributed Hypothesis Testing with Variable-Length Coding

Sadaf Salehkalaibar, *IEEE Member* and Michèle Wigger, *IEEE Senior Member*

***Abstract*—The problem of distributed testing against independence with variable-length coding is considered when the *average* and not the *maximum* communication load is constrained as in previous works. The paper characterizes the optimum type-II error exponent of a single-sensor single-decision center system given a maximum type-I error probability when communication is either over a noise-free rate-R link or over a noisy discrete memoryless channel (DMC) with stop-feedback. Specifically, let ϵ denote the maximum allowed type-I error probability. Then the optimum exponent of the system with a rate-R link under a constraint on the average communication load coincides with the optimum exponent of such a system with a rate $R/(1-\epsilon)$ link under a maximum communication load constraint. A strong converse thus does not hold under an average communication load constraint. A similar observation also holds for testing against independence over DMCs. With variable-length coding and stop-feedback and under an average communication load constraint, the optimum type-II error exponent over a DMC of capacity C equals the optimum exponent under fixed-length coding and a maximum communication load constraint when communication is over a DMC of capacity $C(1-\epsilon)^{-1}$.**

I. Introduction

Consider a distributed hypothesis testing problem with a single decision center that aims at identifying the distribution governing the sources observed at the decision center itself and at various sensors. To facilitate this task, the sensors communicate with the decision center over rate-limited links. The focus is on binary hypothesis testing problems where the sources are distributed according to one of only two possible joint distributions, a joint distribution P under the *null hypothesis* ($\mathcal{H} = H_0$) and a different joint distribution Q under the *alternative hypothesis* ($\mathcal{H} = H_1$). The main interest of this paper is in identifying the largest possible Stein-exponent of such systems. That is, the maximum exponential decay of the type-II error probability, i.e., the probability of deciding H_0 when $\mathcal{H} = H_1$, subject to a constraint on the type-I error probability, i.e., on the probability of deciding H_1 when $\mathcal{H} = H_0$. Stein-exponents of distributed hypothesis testing systems have widely been studied in the information-theoretic literature, see for example [1]–[16]. In particular, Ahlswede and Csiszár [1] have characterized the Stein-exponent of a single-sensor system where the sensor communicates with the decision center over a noiseless rate-limited link in the special case of *testing against independence* where Q (the joint distribution under H_1) equals the product of the marginals of P (the distribution under H_0). The Stein exponent of this special case has also been solved in more complicated scenarios with multiple sensors [4], with multiple sensors and cooperation between sensors [7], with a single sensor and successive refinement communication [5], with interactive communication between sensor and decision center [8], with a single sensor and multiple decision centers without and with cooperation [12] and [16], and in a multi-hop environment with multiple sensors and decision centers [11]. In all these works, communication takes place over rate-limited but noiseless links and the maximum allowed type-I error probability $\epsilon \to 0$. Sreekumar and Gündüz [17] identified the Stein exponent of the basic single-sensor single-center system when communication takes place over a discrete memoryless channel (DMC). They showed that the Stein exponent of this setup coincides with the Stein exponent of the scenario with a noiseless link of rate equal to the capacity of the DMC. The Stein exponent thus depends on the DMC's transition law only through its capacity. The extension to multiple sensors that communicate with the single decision center over a discrete memoryless multiple-access channel was presented in [18]. Most of the described results can easily be extended also to generalized testing against independence where the distribution Q under $\mathcal{H}_1$ factorizes into the product of the marginals but not necessarily equal to the marginals of P under $\mathcal{H}_1$ or to testing against conditional independence as introduced in [4], see also [12], [13], [17], [19]. Bounds on the Stein exponents for general distributed hypothesis tests (not necessarily testing against independence or conditional independence) have also been derived for various of the described scenarios. For example, Weinberg and Kochman [6] characterised the Stein-exponent under an optimal detection rule, and Haim and Kochman [20] provided improved exponents for some general tests with binary sources.

In above results, the maximum allowed type-I error probability ϵ is taken to 0, which implies that the proofs are built on "weak" converses. In contrast, Ahlswede and Csiszár showed [1] that for single-sensor single-decision center setups with a rate-limited noiseless link a "strong" converse holds, i.e., the maximum type-II error exponent does not depend on ϵ. This result is even more remarkable in that the optimum Stein-exponent is not known for the general hypothesis testing problem with a single noise-less link. Tian and Chen [5] and Cao, Zhou, and Tan [15] proved strong converse results for testing against independence in a single-sensor single-

S. Salehkalaibar is with the Department of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran, s.saleh@ut.ac.ir,

M. Wigger is with LTCI, Telecom Paris, 91120 Palaiseau, Paris, France, michele.wigger@telecom-paristech.fr.

Parts of the material in this paper was presented at *The 2020 Workshop on Resource Allocation, Cooperation and Competition in Wireless Networks (RAWNET)*, June 2020.

decision center setup under noiseless successive refinement communication and in a two-sensor single decision center setup with noiseless multi-hop communication. Two of the main tools for deriving strong converse results are the change of measure approach under the *η-image characterization* [1] and the blowing-up lemma [21], [22] or the hypercontractivity lemma [23].

Another line of works requires that the probability of error decays exponentially under both hypotheses and studies the pair of exponential decays that can simultaneously be achieved. Han and Kobayashi studied the setup with one or multiple sensors that are connected over a noiseless ratelimited link with a single decision center. The extension to DMCs was proposed in [24]. Recently, also a finite blocklength version of this problem was studied in [25]. All these works contain achievability results but no converses.

The described previous results measure communication load in terms of the *maximum* number of transmitted bits or the *maximum* number of channel uses. In this paper, we allow for variable-length coding and consider *average* communication loads. When communication is over a noise-free rate-limited communication link, the average load is simply the expected number of transmitted bits. When communication is over a DMC, then we allow for variable-length coding with stop-feedback from the receiver [26] and communication load is characterised by means of the expected number of channel uses. The feedback signal is assumed to be a function of the channel output but not the source sequence observed at the decision center. This models a setup where the decision center learns its local observations only after the communication is terminated. In this paper, we characterize the optimal Stein-exponents of the single-sensor single-decision center system for testing against independence when variable-length coding is allowed and the *average* communication load is constrained. The derived exponents coincide with the previously obtained exponents with fixed-length coding (and a constraint on the maximum communication load), except that the rates/capacity of the communication links have to be multiplied by the term $(1-\epsilon)^{-1}$ where ϵ denotes the maximum allowed type-I error probability. So, variable-length coding can be seen as boosting the rate/capacity of the communication link by the factor $(1-\epsilon)^{-1}$. Notice that this implies in particular that a strong converse result does not hold under variable-length coding for hypothesis testing. This conclusion is in line with previous related works e.g., [27]–[29] for compression problems with non-zero error probability, which proved that a strong converse established for fixed-length coding can break down when variable-length coding is permitted. Our results further show that the optimal Stein-exponent that is achievable over a DMC depends only on the capacity of the channel but not on other properties of the DMC.

These optimal Stein-exponents can be achieved by simple modifications of the optimal schemes for fixed-length coding, for the latter, see for example [1], [30]. The idea is to identify an event $\mathcal{S}_n$ at the sensor that happens with probability ϵ', for ϵ' slightly smaller than the largest admissible type-I error probability ϵ. In the noiseless link setup, whenever event $\mathcal{S}_n$ occurs, the sensor will send the single bit 0 to the decision center, which then declares $\hat{\mathcal{H}} = H_1$. If the event $\mathcal{S}_n$ does not occur, the sensor acts as in the scheme proposed by Ahlswede and Csiszár [1]. The proposed strategy achieves a smaller type-II error probability than the Ahlswede-Csiszár scheme and its type-I error probability is increased at most by ϵ' (namely the probability of event $\mathcal{S}_n$). The type-II error exponent of the modified scheme thus coincides with Ahlswede-Czsiszár's exponent, and its type-I error probability can be bounded by $\epsilon > \epsilon'$ when the number of observations is sufficiently large. The expected communication rate is decreased by a factor $(1-\epsilon)$ since no rate is required in the event $\mathcal{S}_n$. The main technical contribution in this part is the converse showing that the described simple strategy is optimal. The converse combines Marton's blowing up lemma [22] and a change of measure argument using the η-image characterization similarly to [1] and [5].

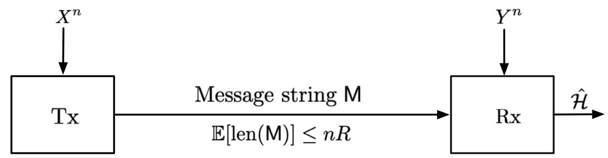

Fig. 1. Variable-length hypothesis testing.

For the DMC, our optimal strategy takes place over two phases. In the first shorter phase, the transmitter sends a dedicated sequence w_0^n if event $\mathcal{S}_n$ occurs and it sends a different sequence w_1^n otherwise. The decision center performs a Neyman-Pearson test to detect which of the two sequences has been transmitted. If it detects w_0^n, it declares directly $\hat{\mathcal{H}} = H_1$ and sends a stop signal. Otherwise, the sensor proceeds to phase 2, where it applies the fixed-length coding scheme proposed in [30] that achieves the optimal Stein-exponent under fixed-length coding for testing against independence over a DMC. In the proposed variable-length strategy the type-II error probability is decreased compared to the fixed-length scheme in [30], the type-I error probability is increased by at most ϵ', and for large numbers of observations, the average number of channel uses is decreased approximately by a factor $(1-\epsilon')$. This last observation holds because the first phase is much smaller than the second phase and transmission stops after the first phase with probability close to ϵ'. We again prove the corresponding converse result. This proof requires some additional steps and considerations concerning the noisy channel law and the stop-feedback compared to the converse for the noise-less link.

The paper is organized as follows. In Section II, the distributed hypothesis testing problem over a noiseless link is studied and the result on the noisy channel is provided in Section III. The proofs of the converses for the noiseless and noisy setups are provided in Sections IV and V, respectively. The paper is concluded in Section VI.

We conclude the introduction with some remarks on notation.

Notation:

Random variables are denoted by capital letters, e.g.,

X, Y, and their realizations by lower-case letters, e.g., x, y. Script symbols such as $\mathcal{X}$ and $\mathcal{Y}$ stand for alphabets of random variables, and $\mathcal{X}^n$ and $\mathcal{Y}^n$ for the corresponding n-fold Cartesian product alphabets. We denote by $\mathcal{X}^\star$ and $\mathcal{Y}^\star$ the sets of all finite-length strings over $\mathcal{X}$ and $\mathcal{Y}$ respectively. The set of real numbers is denoted by $\mathbb{R}$, the set of positive real numbers by $\mathbb{R}_+$, the set of integers by $\mathbb{Z}$, and the set of positive integers by $\mathbb{Z}_+$. Sequences of random variables $(X_i, ..., X_j)$ and realizations $(x_i, \ldots, x_j)$ are abbreviated by X_i^j and x_i^j. When $i = 1$, then we also use X^j and x^j instead of X_1^j and x_1^j.

We write the probability mass function (pmf) of a discrete random variable X as P_X. The conditional pmf of X given Y is written as $P_{X|Y}$. The distributions of X^n, Y^n and (X^n, Y^n) are denoted by P_{X^n}, P_{Y^n} and $P_{X^nY^n}$, respectively. The notation P_{XY}^n denotes the n-fold product distribution.

The term $D(P\|Q)$ stands for the Kullback-Leibler (KL) divergence between two pmfs P and Q over the same alphabet. For a given P_X and a constant $\mu > 0$, the set of sequences with the same type P_X is denoted by $\mathcal{T}^{(n)}(P_X)$. We use $\mathcal{T}_\mu^{(n)}(P_X)$ to denote the set of *μ-typical sequences* in $\mathcal{X}^n$:

$$\mathcal{T}_\mu^{(n)}(P_X) = \left\{ x^n \colon \left| \frac{|\{i \colon x_i = x\}|}{n} - P_X(x) \right| \leq \mu P_X(x),\ \forall x \in \mathcal{X} \right\}, \tag{1}$$

where $|\{i \colon x_i = x\}|$ is the number of positions where the sequence x^n equals x. Similarly, $\mathcal{T}_\mu^{(n)}(P_{XY})$ stands for the set of *jointly μ-typical sequences* whose definition is as in (1) with x replaced by (x, y).

For any positive integer number $m \geq 1$, we use $\text{string}(m)$ to denote the bit-string of length $\lceil \log_2(m) \rceil$ representing m. We further use sans serif font for finite-length bit-strings, e.g., M for a random bit-string and m for a deterministic bit-string. The function $\text{len}(\mathsf{m})$ returns the length of a given bit-string m.

The Hamming distance between two sequences x^n and y^n is denoted by $d_{\text{H}}(x^n, y^n)$. For any $a, b \in [0, 1]$, we denote the binary entropy function of a by $h_{\text{b}}(a)$ and define $a \star b \triangleq a(1-b) + b(1-a)$.

II. Distributed Hypothesis Testing Over a Positive-Rate Noiseless Link

A. System Model

Consider the distributed hypothesis testing problem with a transmitter and a receiver in Fig. 1. The transmitter observes the source sequence X^n and the receiver observes the source sequence Y^n. Under the null hypothesis

$$\mathcal{H} = H_0 \colon \quad (X^n, Y^n) \sim \text{i.i.d. } P_{XY}, \tag{2}$$

for a given pmf P_{XY}, whereas under the alternative hypothesis

$$\mathcal{H} = H_1 \colon \quad (X^n, Y^n) \sim \text{i.i.d. } P_X \cdot P_Y, \tag{3}$$

where P_X and P_Y denote the marginals of P_{XY}. Upon observing X^n, the transmitter computes the binary message string $\mathsf{M} \in \{0, 1\}^\star$ using a possibly stochastic encoding function

$$\phi^{(n)} \colon \mathcal{X}^n \to \{0, 1\}^\star, \tag{4}$$

so

$$\mathsf{M} = \phi^{(n)}(X^n), \tag{5}$$

in a way that the expected[1] message length satisfies

$$\mathbb{E}\left[\text{len}(\mathsf{M})\right] \leq nR. \tag{6}$$

It then sends the binary message string M over a noise-free bit pipe to the receiver.

The goal of the communication is that the receiver can determine the hypothesis $\mathcal{H}$ based on its observation Y^n and its received message M. Specifically, the receiver produces the guess

$$\hat{\mathcal{H}} = g^{(n)}(Y^n, \mathsf{M}) \tag{7}$$

using a decoding function $g^{(n)} \colon \mathcal{Y}^n \times \{0, 1\}^\star \to \{H_0, H_1\}$. Denoting by $\mathcal{M}$ the set of all realizations of the binary message string M, we can partition the space $\mathcal{M} \times \mathcal{Y}^n$ into an acceptance region for hypothesis H_0

$$\mathcal{A}_n \triangleq \left\{ (\mathsf{m}, y^n) \colon g^{(n)}(y^n, \mathsf{m}) = H_0 \right\}, \tag{8}$$

and the corresponding rejection region

$$\mathcal{R}_n \triangleq (\mathcal{M} \times \mathcal{Y}^n) \backslash \mathcal{A}_n. \tag{9}$$

Definition 1: For any $\epsilon \in [0, 1)$ and for a given rate $R \in \mathbb{R}_+$, a type-II exponent $\theta \in \mathbb{R}_+$ is (ϵ, R)-achievable if there exists a sequence of functions $\{(\phi^{(n)}, g^{(n)})\}_{n \geq 1}$, such that the corresponding acceptance and rejection regions lead to a type-I error probability

$$\alpha_n \triangleq \Pr[(\mathsf{M}, Y^n) \in \mathcal{R}_n | \mathcal{H} = H_0] \tag{10}$$

and a type-II error probability

$$\beta_n \triangleq \Pr[(\mathsf{M}, Y^n) \in \mathcal{A}_n | \mathcal{H} = H_1] \tag{11}$$

satisfying for sufficiently large blocklengths n:

$$\alpha_n \leq \epsilon, \tag{12}$$

and

$$\liminf_{n \to \infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta. \tag{13}$$

The optimal exponent $\theta_\epsilon^*(R)$ is the supremum of all (ϵ, R)-achievable type-II exponents $\theta \in \mathbb{R}_+$.

B. Optimal Type-II Error Exponent

The following theorem establishes the optimal type-II error exponent $\theta_\epsilon^*(R)$.

Theorem 1: The optimal type-II error exponent with variable-length coding is

$$\theta_\epsilon^*(R) = \max_{\substack{P_{U|X}\colon \\ |\mathcal{U}| \leq |\mathcal{X}|+1 \\ R \geq (1-\epsilon) I(U;X)}} I(U; Y), \tag{14}$$

Proof: Here we only prove achievability. The converse is more technical and proved in Section IV.

[1]The expectation in (6) is with respect to the law of X^n which equals P_X^n under both hypotheses.

Achievability: Fix a large blocklength n, a small number $\mu \in (0, \epsilon)$, and a conditional pmf $P_{U|X}$ such that:

$$R = (1 - \epsilon + \mu)(I(U; X) + \mu). \tag{15}$$

Define the joint pmf

$$P_{UXY} \triangleq P_{U|X} \cdot P_{XY} \tag{16}$$

and randomly generate an n-length codebook $\mathcal{C}_U$ of rate R by picking all entries i.i.d. according to the marginal pmf P_U. The realization of the codebook

$$\mathcal{C}_U \triangleq \left\{ u^n(m) \colon m \in \{1, \ldots, \lfloor 2^{nR} \rfloor\} \right\} \tag{17}$$

is revealed to all terminals.

Finally, choose a subset $\mathcal{S}_n \subseteq \mathcal{T}^{(n)}_{\mu/2}(P_X)$ such that

$$\Pr\left[X^n \in \mathcal{S}_n\right] = \epsilon - \mu. \tag{18}$$

Transmitter: Assume it observes $X^n = x^n$. If

$$x^n \notin \mathcal{S}_n, \tag{19}$$

it looks for an index $m \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$ such that

$$(u^n(m), x^n) \in \mathcal{T}^{(n)}_{\mu/2}(P_{UX}). \tag{20}$$

If successful, it picks one of these indices uniformly at random and sends the binary representation of the chosen index over the noiseless link. So, if the chosen index is $m^* \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$, it sends the corresponding length-nR bit-string

$$\mathsf{M} = \text{string}(m^*). \tag{21}$$

Otherwise it sends the single bit $\mathsf{M} = [0]$.

Receiver: If it receives the single bit $\mathsf{M} = [0]$, it declares $\hat{\mathcal{H}} = H_1$. Otherwise, if the bit string M corresponds to a given index $m \in \{1, \ldots, \lfloor 2^{nR} \rfloor\}$, it checks whether $(u^n(m), y^n) \in \mathcal{T}^{(n)}_{\mu}(P_{UY})$. If successful, it declares $\hat{\mathcal{H}} = H_0$, and otherwise it declares $\hat{\mathcal{H}} = H_1$.

Analysis: The proposed coding scheme is analyzed when averaged over the random code construction. By standard arguments it can then be concluded that the desired exponent is achievable also for at least one realizations of the codebooks.

Since a single bit is sent when $x^n \in \mathcal{S}_n$, the expected message length can be bounded as:

$$\begin{aligned}\mathbb{E}\left[\text{len}(\mathsf{M})\right] &= \Pr[X^n \in \mathcal{S}_n] \cdot \mathbb{E}\left[\text{len}(\mathsf{M}) | X^n \in \mathcal{S}_n\right] \\ &\quad + \Pr[X^n \notin \mathcal{S}_n] \cdot \mathbb{E}\left[\text{len}(\mathsf{M}) | X^n \notin \mathcal{S}_n\right] \quad (22) \\ &\leq (\epsilon - \mu) \cdot 1 + (1 - \epsilon + \mu) \cdot n(I(U; X) + \mu), \quad (23)\end{aligned}$$

which for sufficiently large n is further bounded as (see (15)):

$$\mathbb{E}\left[\text{len}(M)\right] < nR. \tag{24}$$

To bound the type-I and type-II error probabilities, we notice that when $x^n \notin \mathcal{S}_n$, the scheme coincides with the one proposed by Ahlswede and Csiszár in [1]. When $x^n \in \mathcal{S}_n$, the transmitter sends the single bit $\mathsf{M} = [0]$ and the receiver declares H_1. The type-II error probability of our scheme is thus no larger than the type-II error probability of the Ahlswede-Csiszár scheme in [1], and the type-I error probability is at most $\Pr[X^n \in \mathcal{S}_n] = \epsilon - \mu$ larger than for this Ahlswede-Csiszár scheme. Since the type-I error probability of the Ahlswede-Csiszár scheme tends to 0 as $n \to \infty$ [1], the type-I error probability here is bounded by ϵ, for sufficiently large values of n and all choices of $\mu \in (0, \epsilon)$. Combining these considerations with (24), and letting $n \to \infty$ and $\mu \to 0$ establishes the achievability part of the proof. ■

For comparison, recall the result in [1] which showed that under fixed-length coding, i.e., when instead of the average message length only the maximum message length is constrained by nR, the optimal type-II error exponent equals:

$$\theta^*_{\text{FL}}(R) = \max_{\substack{P_{U|X}: \\ |\mathcal{U}| \leq |\mathcal{X}|+1 \\ R \geq I(U;X)}} I(U; Y). \tag{25}$$

Under fixed-length coding, the optimal type-II error exponent does hence not depend on the maximum allowed type-I error probability ϵ and we say that a "strong converse" holds. Our result in Theorem 1 shows that such a "strong converse" does not hold under variable-length coding and also quantifies the gain in type-II error exponent as a function of the maximum allowed type-I error probability. We thus encounter a similar situation as in source coding with a small positive error probability, where various works [27]–[29] have shown that variable-length coding allows to decrease the required compression rate below the entropy of the source.

We present two examples to further illustrate the gain of variable-length coding compared to fixed-length coding.

Example 1: Suppose that the source alphabets are binary with $P_X = P_Y \sim \text{Bern}(\frac{1}{2})$ and the conditional pmf $P_{Y|X}$ is given by

$$P_{Y|X}(y|x) = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}, \tag{26}$$

where $0 \leq \alpha < \frac{1}{2}$. We can write the following set of inequalities:

$$\begin{aligned}\theta^*_\epsilon(R) &= \max_{\substack{P_{U|X}: \\ |\mathcal{U}| \leq |\mathcal{X}|+1 \\ R \geq (1-\epsilon) I(U;X)}} I(U; Y) \quad (27) \\ &= \max_{\substack{P_{U|X}: \\ |\mathcal{U}| \leq |\mathcal{X}|+1 \\ 1 - h_{\text{b}}(X|U) \leq \frac{R}{1-\epsilon}}} 1 - h_{\text{b}}(Y|U) \quad (28) \\ &= 1 - h_{\text{b}}\left(h_{\text{b}}^{-1}\left(1 - \frac{R}{1-\epsilon}\right) \star \alpha\right) \quad (29)\end{aligned}$$

Notice that the last equality holds by Ms. Gerber's lemma [31, p. 19].

Following similar steps, it can be shown that the optimal type-II error exponent under variable-length coding evaluates to

$$\theta^*_{\text{FL}}(R) = 1 - h_{\text{b}}\left(h_{\text{b}}^{-1}\left(1 - R\right) \star \alpha\right). \tag{30}$$

Fig. 2 shows the optimal error exponents $\theta^*_\epsilon(R)$ and $\theta^*_{\text{FL}}(R)$ in functions of the parameter α for $\epsilon = 0.1$ and $R = 0.8$. The gain of variable-length coding compared to fixed-length coding seems to be particularly pronounced for small values

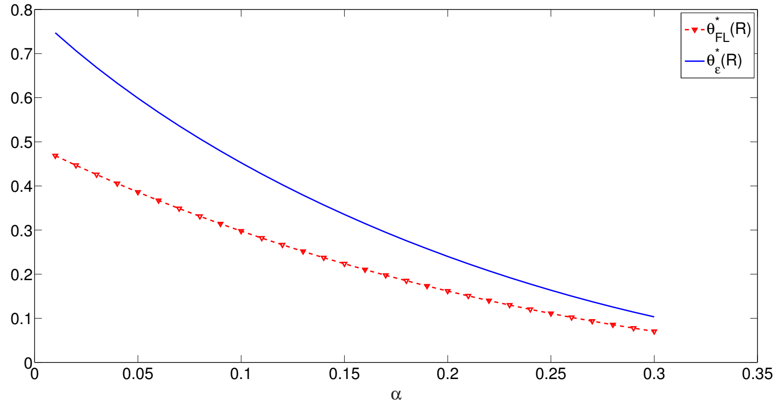

Fig. 2. Comparison of fixed-length and variable-length codings for Example 1.

of α, where the sources are highly correlated under the null hypothesis H_0.

Though we have proved Theorem 1 only for finite alphabets, we will evaluate it for a Gaussian example, as it is often done in the information-theoretic literature.

Example 2: Given $\rho \in [0,1]$, define the two covariance matrices

$$\mathbf{K}_{XY}^0 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{K}_{XY}^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{31}$$

Under the null hypothesis,

$$\mathcal{H} = H_0\colon \qquad (X,Y) \sim \mathcal{N}(0, \mathbf{K}_{XY}^0), \tag{32}$$

and under the alternative hypothesis,

$$\mathcal{H} = H_1\colon \qquad (X,Y) \sim \mathcal{N}(0, \mathbf{K}_{XY}^1). \tag{33}$$

The above setup can model a communication scenario with a jammer. Under the null hypothesis, the jammer interferes with the communication and the observations at the transmitter and receiver are correlated with each other where the correlation is modelled by the parameter ρ. Under the alternative hypothesis, the jammer remains silent and the observations X^n and Y^n are independent of each other. The goal of the system is to detect the presence of the jammer.

To characterize the optimal type-II error exponent in the above example, notice that under $\mathcal{H} = H_0$, one can write $Y = \rho X + Z$ with Z a zero-mean Gaussian random variable of variance $1-\rho^2$ and independent of X. Consider the following set of equalities:

$$\theta_{\text{G},\epsilon}^*(R) = \max_{\substack{P_{U|X}\colon \\ R \geq (1-\epsilon) I(U;X)}} I(U;Y) \tag{34}$$

$$= \frac{1}{2} \log \left(\frac{1}{1 - \rho^2 + \rho^2 \cdot 2^{-\frac{2R}{1-\epsilon}}} \right), \tag{35}$$

where (35) is a well-known inequality based on the entropy-power inequality (EPI) [31, pp. 22]. Notice that the maximum in (34) is achieved by jointly Gaussian (U,X).

Following similar steps, one can show that the optimal type-II error exponent under fixed-length coding evaluates to:

$$\theta_{\text{G,FL}}^*(R) = \frac{1}{2} \log \left(\frac{1}{1 - \rho^2 + \rho^2 \cdot 2^{-2R}} \right). \tag{36}$$

Fig. 3 shows the optimal error exponents $\theta_{\text{G},\epsilon}^*(R)$ and $\theta_{\text{G,FL}}^*(R)$ in function of the parameter ρ for $\epsilon = 0.1$ and $R = 0.8$. For large values of the parameter ρ where the sources are highly correlated under the null hypothesis, variable-length coding outperforms fixed-length coding.

III. Testing Over a Discrete Memoryless Channel (DMC)

A. System Model

Consider a hypothesis testing system with a single transmitter and a single receiver where communication is over a discrete memoryless channel (DMC) with input alphabet $\mathcal{W}$, output alphabet $\mathcal{V}$, and transition law $\Gamma_{V|W}(\cdot|\cdot)$. The number of channel uses is a random quantity, because the transmitter stops transmission after receiving a feedback signal from the receiver. This stop feedback-signal is without error or delay.

As in the previous section, the transmitter observes the source sequence X^n and the receiver observes the side-information sequence Y^n, where

$$\text{under } \mathcal{H} = H_0\colon \quad (X^n, Y^n) \sim \text{i.i.d. } P_{XY}, \tag{37}$$

and

$$\text{under } \mathcal{H} = H_1\colon \quad (X^n, Y^n) \sim \text{i.i.d. } P_X \cdot P_Y. \tag{38}$$

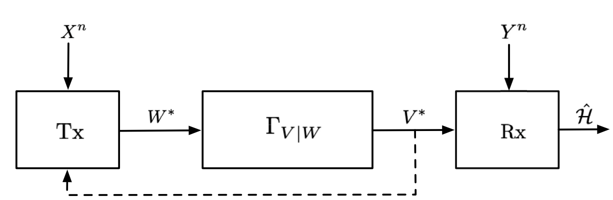

Fig. 4. Hypothesis testing over a noisy channel with variable-length coding and stop feedback.

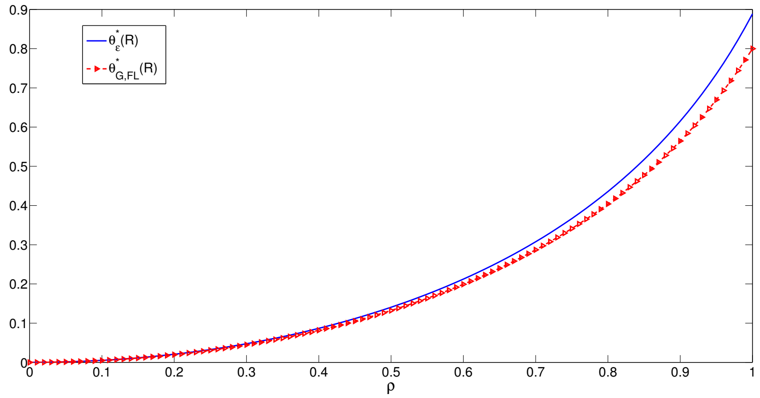

Fig. 3. Comparison of fixed-length and variable-length codings for Example 2.

Based on the source sequence X^n, the transmitter generates an infinite-length stream

$$W'^{\infty}(X^n) = W'_1, W'_2, \ldots \tag{39}$$

and for each channel use prior to the stop-feedback, it sends the corresponding symbol of the sequence $W'^{\infty}(X^n)$ over the channel. For each time-instant k, let $L_k = 1$ indicate that the receiver has not yet sent the stop-symbol, and $L_k = 0$ otherwise. We then have for the time-k channel input W_k:

$$W_k = W'_k, \qquad \text{if } L_k = 1, \text{ for } k = 1, 2, \ldots. \tag{40}$$

Let τ_n denote the transmission duration, i.e.,

$$\tau_n := \min\{k \geq 1 \colon L_k = 0\}. \tag{41}$$

The receiver observes the random channel outputs $V_1, V_2, \ldots, V_{\tau_n}$ corresponding to the inputs $W_1, W_2, \ldots, W_{\tau_n}$ fed to the given DMC $\Gamma_{V|W}$. At each time $k = 1, 2, \ldots$, the receiver decides whether the communication should continue ($L_k = 1$) or not ($L_k = 0$). For simplicity, we assume that the decision L_k is only a function of the first $k-1$ channel outputs $V_1, \ldots, V_{k-1}$ but not of Y^n, in which case τ_n is a stopping time of the filtration $\sigma\{V^k\}_{k\geq 1}$. The described setup models for example a situation where the receiver learns the side-information Y^n only after the communication has terminated. Thus, in our scenario:

$$L_k = e_k^{(n)}(V^k), \tag{42}$$

for each $k = 1, 2, \ldots$ and some stopping function $e_k \colon \mathcal{V}^k \to \{0, 1\}$. The stopping functions determine the set of all output strings for which the receiver stops the transmission:

$$\mathcal{V}_{\text{stop}} \triangleq \left\{ v^\tau \in \mathcal{V}^\star \colon e_\tau^{(n)}(v^\tau) = 0 \text{ and } e_{\tau-1}^{(n)}(v^{\tau-1}) = 1 \right\}. \tag{43}$$

where here $v^{\tau-1}$ denotes the first $\tau - 1$ symbols of v^τ.

Once transmission stops, the receiver has observed the channel outputs $V^{\tau_n} \in \mathcal{V}_{\text{stop}}$ and the side-information Y^n. Based on these observations, it has to guess the hypothesis $\hat{\mathcal{H}} = H_0$ or $\hat{\mathcal{H}} = H_1$. To this end, it chooses a subset $\mathcal{A}_n \subset \mathcal{V}_{\text{stop}} \times \mathcal{Y}^n$, which we call the acceptance region, and it decides on $\hat{\mathcal{H}} = H_0$ whenever $(V^{\tau_n}, Y^n) \in \mathcal{A}_n$. Conversely, it decides on $\hat{\mathcal{H}} = H_1$ whenever (V^{τ_n}, Y^n) lies in the complement $\mathcal{R}_n \triangleq (\mathcal{V}_{\text{stop}} \times \mathcal{Y}^n) \backslash \mathcal{A}_n$, which we call the rejection region.

The type-I error probability is then defined as:

$$\alpha_n \triangleq P_{V^{\tau_n} Y^n}(\mathcal{R}_n) = 1 - P_{V^{\tau_n} Y^n}(\mathcal{A}_n), \tag{44}$$

and the type-II error probability as:

$$\beta_n \triangleq P_{V^{\tau_n}} P_Y^n(\mathcal{A}_n). \tag{45}$$

Definition 2: For any $\epsilon \in [0, 1)$ and a given bandwidth mismatch factor $\kappa \in \mathbb{R}_+$, we say that a type-II error exponent $\theta \in \mathbb{R}_+$ is (ϵ, κ)-achievable if there exists a sequence of encoding functions, stopping functions, and acceptance regions $\left\{\{\Phi_k^{(n)}\}_{k\geq 1}, \{e_k^{(n)}\}_{k\geq 1}, \mathcal{A}_n\right\}_{n\geq 1}$, such that the corresponding sequences of type-I and type-II error probabilities satisfy for sufficiently large blocklengths n:

$$\alpha_n \leq \epsilon, \tag{46}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta, \tag{47}$$

and the average transmission duration $\mathbb{E}[\tau_n]$ satisfies

$$\limsup_{n\to\infty} \frac{\mathbb{E}[\tau_n]}{n} \leq \kappa. \tag{48}$$

Given $\kappa \in \mathbb{R}_+$, the optimal exponent $\theta^*_{\text{DMC},\epsilon}(\kappa)$ is the supremum of all (ϵ, κ)-achievable type-II error exponents $\theta \in \mathbb{R}_+$.

B. Optimal Error Exponent

Theorem 2: The optimal type-II exponent over a DMC $(\mathcal{W}, \mathcal{V}, \Gamma_{V|W})$ with variable-length coding and stop feedback is:

$$\theta^*_{\text{DMC},\epsilon}(\kappa) = \max_{\substack{P_{U|X}\colon \\ |\mathcal{U}|\leq|\mathcal{X}|+1 \\ \kappa C \geq (1-\epsilon) I(U;X)}} I(U;Y), \tag{49}$$

where C denotes the capacity of the DMC $(\mathcal{W}, \mathcal{V}, \Gamma_{V|W})$.

Proof: The converse is proved in Section V. The achievability in the following subsection III-C.

Under fixed-length coding, the optimal type-II error exponent is [17], [32]

$$\theta^*_{\text{DMC,FL}}(\kappa) := \max_{\substack{P_{U|X}:\\ |\mathcal{U}|\le|\mathcal{X}|+1\\ \kappa C\ge I(U;X)}} I(U;Y), \tag{50}$$

irrespective of the allowed type-I error probability ϵ, and thus a "strong converse" holds under fixed-length coding. In contrast, our result in Theorem 2 shows that under variable-length coding a "strong converse" does not hold and it characterises the gain in optimal type-II error exponent when a type-I error probability of $\epsilon > 0$ is tolerated.

C. Coding Scheme Achieving the Optimal Exponent

We now prove achievability of the exponent in (50). Choose two different symbols $w_0, w_1 \in \mathcal{W}$ such that the KL-divergence of the output distributions induced by these inputs is positive, i.e., such that

$$D(\Gamma_{w_0}\|\Gamma_{w_1}) > 0, \tag{51}$$

where

$$\Gamma_{w_0}(\cdot) \triangleq \Gamma(\cdot|w_0), \qquad \Gamma_{w_1}(\cdot) \triangleq \Gamma(\cdot|w_1). \tag{52}$$

Further, choose a positive number $\epsilon' \in (0,\epsilon)$ close to ϵ and a function $q\colon \mathbb{Z}^+ \to \mathbb{Z}^+$ that satisfies the following two conditions:

$$\lim_{n\to\infty} q(n) = \infty \tag{53}$$

$$\lim_{n\to\infty} \frac{q(n)}{n} = 0. \tag{54}$$

Define

$$\mu \triangleq \epsilon - \epsilon'. \tag{55}$$

Fix two pmfs $P_{U|X}$ and P_W and a positive rate R so that the following two conditions hold:

$$R = I(U;X) + \mu, \tag{56}$$

$$R < \frac{\kappa}{1-\epsilon'} I(W;V). \tag{57}$$

Define $P_{UX} \triangleq P_{U|X}\cdot P_X$ and $P_{WV} \triangleq P_W \cdot \Gamma_{V|W}$.

Fix now a large blocklength n and generate two codebooks

$$\mathcal{C}_U \triangleq \left\{u^n(m)\colon m \in \{1,\dots,\lfloor 2^{nR}\rfloor\}\right\}, \tag{58}$$

$$\mathcal{C}_W \triangleq \left\{w^{n'}(m)\colon m \in \{0,\dots,\lfloor 2^{nR}\rfloor\}\right\}, \tag{59}$$

where

$$n' \triangleq \frac{n\kappa}{1-\epsilon'}, \tag{60}$$

and where the entries of the two codebooks are picked i.i.d. according to the pmfs P_U and P_W, respectively. Furthermore, choose a subset $\mathcal{S}_n \subseteq \mathcal{T}^{(n)}_{\mu/2}(P_X)$ such that

$$\Pr\left[X^n \in \mathcal{S}_n\right] = \epsilon'. \tag{61}$$

The coding scheme decomposes into two phases.

Phase 1: Consists of the first $q(n)$ channel uses.

Transmitter: Given that it observes $X^n = x^n$, the transmitter sends the $q(n)$ inputs

$$(W_1,\dots,W_{q(n)}) = \begin{cases} w_1^{\otimes q(n)}, & \text{if } X^n \in \mathcal{S}_n,\\ w_0^{\otimes q(n)}, & \text{otherwise,}\end{cases} \tag{62}$$

where for any input symbol $w \in \mathcal{W}$,

$$w^{\otimes j} \triangleq (\underbrace{w,\dots,w}_{j\text{ times}}), \quad j \in \mathbb{Z}^+. \tag{63}$$

Receiver: Upon observing the first $q(n)$ channel outputs $V_1,\dots,V_{q(n)}$, the receiver performs a Neyman-Pearson test to decide on whether the transmitter sent $w_0^{\otimes q(n)}$ or $w_1^{\otimes q(n)}$. This test only depends on the channel outputs but not on the receiver's side-information Y^n. The threshold of the test is set so that the probability of declaring $w_1^{\otimes q(n)}$ when $w_0^{\otimes q(n)}$ was sent, equals $\mu/3$.

If the receiver detects $w_1^{\otimes q(n)}$, then it decides on

$$\hat{\mathcal{H}} = \mathcal{H}_1 \tag{64}$$

and sends the stop feedback $L_{q(n)} = 0$ to the transmitter, which stops transmission.

If the receiver instead detects $w_0^{\otimes q(n)}$, then it waits to make a decision and also does not send the stop feedback. Both the transmitter and the receiver move on to Phase 2. In this second phase, the receiver will ignore all outputs from the first phase.

Phase 2: This second phase consists of n' channel uses.

Tansmitter: It looks for a codeword $u^n(m)$ such that $(u^n(m), x^n) \in \mathcal{T}^{(n)}_{\mu/2}(P_{UX})$. If no such index exists, it sends $w^{n'}(0)$ over the channel. If one or multiple such indices can be found, the transmitter picks m^* uniformly at random among them and sends the corresponding channel codeword $w^{n'}(m^*)$ over the channel.

Receiver: Let $v_2^{n'}$ denote the n' channel outputs observed at the receiver during this second phase. The receiver looks for a unique index $m \in \{0,\dots,\lfloor 2^{nR}\rfloor\}$ such that

$$(w^{n'}(m), v^{n'}) \in \mathcal{T}^{(n')}_{\mu}(P_{WV}). \tag{65}$$

If $m = 0$ or none of the indices satisfy the condition, the receiver declares $\hat{\mathcal{H}} = H_1$. Otherwise, it produces the decoded message $\hat{M} \in \{1,\dots,\lfloor 2^{nR}\rfloor\}$ equal to the unique index m and proceeds with the following hypothesis test: if $\hat{M} = \hat{m}$ and

$$(u^n(\hat{m}), y^n) \in \mathcal{T}^{(n)}_{\mu}(P_{UY}), \tag{66}$$

then the receiver declares $\hat{\mathcal{H}} = \mathcal{H}_0$, otherwise it declares $\hat{\mathcal{H}} = \mathcal{H}_1$.

In any case it sends the stop-feedback to stop the transmission, $L_{q(n)+n} = 0$.

Analysis: We first analyze the expected transmission duration. Notice that for the described scheme, the transmission duration does not depend on the hypothesis, because it only depends on X^n and the DMC which have same distributions under both hypotheses.

When transmission goes to phase 2, i.e., $L_{q(n)} = 1$, then the transmission duration equals $\tau_n = n' + q(n)$ and when

$L_{q(n)} = 0$, then $\tau_n = q(n)$. Therefore,

$$\mathbb{E}[\tau_n] = q(n) + n' \cdot \Pr\left[L_{q(n)} = 1\right]. \quad (67)$$

To bound $\Pr\left[L_{q(n)} = 1\right]$, we notice that by the way we set the threshold for the Neyman-Pearson test:

$$\Pr\left[L_{q(n)} = 1 \middle| (W_1, \ldots, W_{q(n)}) = w_0^{\otimes q(n)}\right] = 1 - \mu/3. \quad (68)$$

Moreover, by the property of the Neyman-Pearson test, when n (and thus $q(n)$) is sufficiently large, the probability of going to phase 2 after sending $w_1^{\otimes q(n)}$ in phase 1 is bounded as:

$$\begin{aligned} & 2^{-q(n)\left(D(\Gamma_{w_0} \| \Gamma_{w_1}) + \mu\right)} \\ & \leq \Pr\left[L_{q(n)} = 1 \middle| (W_1, \ldots, W_{q(n)}) = w_1^{\otimes q(n)}\right] \\ & \leq 2^{-q(n)\left(D(\Gamma_{w_0} \| \Gamma_{w_1}) - \mu\right)}. \end{aligned} \quad (69)$$

Using that in phase 1 the sequence $w_1^{\otimes q(n)}$ is sent with probability ϵ' and the sequence $w_0^{\otimes q(n)}$ with probability $1-\epsilon'$, we conclude that

$$\begin{aligned} & \Pr\left[L_{q(n)} = 1\right] \\ & = \Pr\left[(W_1, \ldots, W_{q(n)}) = w_0^{\otimes q(n)}\right] \\ & \quad \cdot \Pr\left[L_{q(n)} = 1 \middle| (W_1, \ldots, W_{q(n)}) = w_0^{\otimes q(n)}\right] \\ & \quad + \Pr\left[(W_1, \ldots, W_{q(n)}) = w_1^{\otimes q(n)}\right] \\ & \quad \cdot \Pr\left[L_{q(n)} = 1 \middle| (W_1, \ldots, W_{q(n)}) = w_1^{\otimes q(n)}\right] \quad (70) \\ & \leq (1 - \epsilon') \cdot (1 - \mu/3) + \epsilon' \cdot 2^{-q(n)(D(\Gamma_{w_0} \| \Gamma_{w_1}) - \mu)}. \quad (71) \end{aligned}$$

Since $q(n) \to \infty$ as $n \to \infty$, for sufficiently large n:

$$\Pr\left[L_{q(n)} = 1\right] \leq 1 - \epsilon', \quad (72)$$

and by (67):

$$\mathbb{E}[\tau_n] \leq q(n) + (1 - \epsilon')n'. \quad (73)$$

Dividing both sides of the above inequality by n and letting $n \to \infty$, we obtain by (54)

$$\lim_{n \to \infty} \frac{\mathbb{E}[\tau_n]}{n} \leq \kappa. \quad (74)$$

We analyze the probability of error averaged over the random choice of the codebook. To simplify notation, we introduce a virtual transmitter/receiver pair that always continues to Phase 2 (irrespective of the outcome of the Neyman-Pearson test), and we denote by $\hat{M}_2$ the decoded message produced by this virtual receiver and by $\hat{H}_2$ its guess at the end of Phase 2. Notice that when $L_{q(n)} = 1$, then $\hat{\mathcal{H}}_2 = \hat{\mathcal{H}}$.

Consider first the type-I error probability. When $L_{q(n)} = 0$ then $\hat{\mathcal{H}} = H_1$ with probability 1. Therefore, for sufficiently large values of n:

$$\begin{aligned} & \Pr\left[\hat{\mathcal{H}} = H_1 \middle| \mathcal{H} = H_0\right] \quad (75) \\ & = \Pr\left[L_{q(n)} = 0 \middle| \mathcal{H} = H_0\right] \\ & \quad + \Pr\left[\hat{\mathcal{H}} = H_1, L_{q(n)} = 1 \middle| \mathcal{H} = H_0\right] \quad (76) \\ & = \Pr\left[L_{q(n)} = 0, (W_1, \ldots, W_{q(n)}) = w_0^{\otimes q(n)} \middle| \mathcal{H} = H_0\right] \\ & \quad + \Pr\left[L_{q(n)} = 0, (W_1, \ldots, W_{q(n)}) = w_1^{\otimes q(n)} \middle| \mathcal{H} = H_0\right] \\ & \quad + \Pr\left[\hat{\mathcal{H}} = H_1, L_{q(n)} = 1 \middle| \mathcal{H} = H_0\right] \quad (77) \\ & \leq \Pr\left[L_{q(n)} = 0 \middle| (W_1, \ldots, W_{q(n)}) = w_0^{\otimes q(n)}, \mathcal{H} = H_0\right] \\ & \quad + \Pr\left[(W_1, \ldots, W_{q(n)}) = w_1^{\otimes q(n)} \middle| \mathcal{H} = H_0\right] \\ & \quad + \Pr\left[\hat{\mathcal{H}}_2 = H_1, L_{q(n)} = 1 \middle| \mathcal{H} = H_0\right] \quad (78) \\ & \leq \Pr\left[L_{q(n)} = 0 \middle| (W_1, \ldots, W_{q(n)}) = w_0^{\otimes q(n)}, \mathcal{H} = H_0\right] \\ & \quad + \Pr\left[(W_1, \ldots, W_{q(n)}) = w_1^{\otimes q(n)} \middle| \mathcal{H} = H_0\right] \\ & \quad + \Pr\left[\hat{\mathcal{H}}_2 = H_1 \middle| \mathcal{H} = H_0\right] \quad (79) \\ & \leq \mu/3 + (\mu/3 + \epsilon') + \mu/3 = \epsilon, \quad (80) \end{aligned}$$

where the last inequality holds by the threshold chosen for the Neyman-Pearson test, by the properties of the typical set and the set $\mathcal{S}_n$, and because both the probability of channel decoding error and of wrong hypothesis testing vanish as $n \to \infty$, see for example [18].

Before analyzing the type-II error probability, we notice that $\hat{\mathcal{H}} = H_0$ is only possible when $L_{q(n)} = 1$ and $\hat{M} \neq 0$, in which case $\hat{\mathcal{H}}_2 = \hat{\mathcal{H}}$ and $\hat{M}_2 = \hat{M} \geq 1$. Therefore, for the type-II error probability:

$$\begin{aligned} & \Pr\left[\hat{\mathcal{H}} = H_0 \middle| \mathcal{H} = H_1\right] \\ & = \Pr\left[\hat{\mathcal{H}} = H_0, L_{q(n)} = 1, \hat{M} \neq 0 \middle| \mathcal{H} = H_1\right] \quad (81) \\ & = \Pr\left[\hat{\mathcal{H}}_2 = H_0, L_{q(n)} = 1, \hat{M}_2 \neq 0 \middle| \mathcal{H} = H_1\right] \quad (82) \\ & \leq \Pr\left[\hat{\mathcal{H}}_2 = H_0 \middle| \hat{M}_2 \neq 0, \mathcal{H} = H_1\right] \quad (83) \\ & = \Pr\left[(U^n(\hat{M}_2), Y^n) \in \mathcal{T}_\mu^{(n)}(P_{UY}) \middle| \hat{M}_2 \neq 0, \mathcal{H} = H_1\right]. \quad (84) \end{aligned}$$

Under H_1, the observations Y^n are i.i.d. according to P_Y and independent of $(U^n(\tilde{M}), \tilde{M})$, and thus by a conditional version of Sanov's theorem and continuity of the mutual information measure:

$$\begin{aligned} & \Pr\left[(U^n(\hat{M}_2), Y^n) \in \mathcal{T}_\mu^{(n)}(P_{UY}) \middle| \hat{M}_2 \neq 0, \mathcal{H} = H_1\right] \\ & \quad \leq 2^{-n(I(U;Y) + \delta(\mu))}, \end{aligned} \quad (85)$$

where $\delta(\mu)$ is a function that tends to 0 as $\mu \to 0$. Combining these last two inequalities, one obtains:

$$\Pr\left[\hat{\mathcal{H}} = H_0 \middle| \mathcal{H} = H_1\right] \leq 2^{-n(I(U;Y) + \delta(\mu))}. \quad (86)$$

Taking $n \to \infty$ and $\mu \to 0$, it can be concluded that averaged over the random code construction the desired error exponent is achievable. By standard arguments it then follows that there exist deterministic codebooks achieving the desired exponents. ■

IV. Proof of Converse to Theorem 1

Before proving the converse, we state a standard auxiliary lemma commonly used for hypothesis testing converses.

Lemma 1: Let Q and P be arbitrary pmfs over a discrete and finite set $\mathcal{Z}$ and $\mathcal{A}$ be a subset of $\mathcal{Z}$. Then,

$$-\log Q(\mathcal{A}) \leq \frac{1}{P(\mathcal{A})}(D(P\|Q)+1). \tag{87}$$

Proof: By the data processing inequality for KL-divergence:

$$D(P\|Q) \geq P(\mathcal{A})\log\frac{P(\mathcal{A})}{Q(\mathcal{A})} + (1-P(\mathcal{A}))\log\frac{(1-P(\mathcal{A}))}{(1-Q(\mathcal{A}))} \tag{88}$$

$$= -H_b(P(\mathcal{A})) - P(\mathcal{A})\log Q(\mathcal{A}) \\ -(1-P(\mathcal{A}))\log(1-Q(\mathcal{A})). \tag{89}$$

Upper bounding $H_b(P(\mathcal{A}))$ by 1 and $(1-P(\mathcal{A}))\log(1-Q(\mathcal{A}))$ by 0, and rearranging terms yields the desired inequality. ■

We now prove the desired converse. Fix an achievable exponent $\theta < \theta^*_\epsilon(R)$ and a sequence of encoding and decision functions so that (12) and (13) are satisfied. Further fix a blocklength $n > 0$ and let M and $\hat{\mathcal{H}}$ be the bit-string message and the guess produced by the chosen encoding and decision functions for this given blocklength. Let then μ, η be small positive numbers and define $\mathcal{B}_n(\eta)$ as a subset of $\mathcal{X}^n \times \mathcal{M}$:

$$\mathcal{B}_n(\eta) \triangleq \\ \left\{(x^n,\mathsf{m})\colon \Pr\left[\hat{\mathcal{H}} = H_0 \middle| X^n = x^n, \mathsf{M}=\mathsf{m}, \mathcal{H}=H_0\right] \geq \eta\right\}. \tag{90}$$

By the constraint on the type-I error probability, (12),

$$1-\epsilon \\ \leq \sum_{(x^n,\mathsf{m})\in\mathcal{B}_n(\eta)} \Pr\left[\hat{\mathcal{H}} = H_0 \middle| X^n = x^n, \mathsf{M}=\mathsf{m}, \mathcal{H}=H_0\right] \\ \cdot P_{X^n\mathsf{M}}(x^n,\mathsf{m}) \\ + \sum_{(x^n,\mathsf{m})\in(\mathcal{X}^n\times\mathcal{M})\setminus\mathcal{B}_n(\eta)} \Pr\left[\hat{\mathcal{H}} = H_0 \middle| X^n = x^n, \mathsf{M}=\mathsf{m}, \mathcal{H}=H_0\right] \\ \cdot P_{X^n\mathsf{M}}(x^n,\mathsf{m}) \tag{91}$$

$$\leq P_{X^n\mathsf{M}}(\mathcal{B}_n(\eta)) + \eta(1-P_{X^n\mathsf{M}}(\mathcal{B}_n(\eta))), \tag{92}$$

and as a consequence:

$$P_{X^n\mathsf{M}}(\mathcal{B}_n(\eta)) \geq \frac{1-\epsilon-\eta}{1-\eta}. \tag{93}$$

We next define the subset $\mathcal{D}_n(\eta)$ of $\mathcal{X}^n \times \mathcal{M}$:

$$\mathcal{D}_n(\eta) \triangleq \mathcal{B}_n(\eta) \cap (\mathcal{T}^{(n)}_\mu(P_X)\times\mathcal{M}) \tag{94}$$

By [21, Lemma 2.12]:

$$P_X^n(\mathcal{T}^{(n)}_\mu(P_X)) \geq 1 - \frac{|\mathcal{X}|}{2\mu n}, \tag{95}$$

which combined with (93) and the general identity $\Pr(A\cap B) \geq \Pr(A)+\Pr(B)-1$ implies:

$$P_{X^n\mathsf{M}}(\mathcal{D}_n(\eta)) \geq \frac{1-\epsilon-\eta}{1-\eta} - \frac{|\mathcal{X}|}{2\mu n} \triangleq \Delta_n. \tag{96}$$

Define finally the random variables $(\tilde{\mathsf{M}}, \tilde{X}^n, \tilde{Y}^n)$ as the restriction of the triple (M, X^n, Y^n) to $(X^n,\mathsf{M}) \in \mathcal{D}_n(\eta)$. The probability distribution of the restricted triple is given by:

$$P_{\tilde{\mathsf{M}}\tilde{X}^n\tilde{Y}^n}(\mathsf{m}, x^n, y^n) \triangleq P^n_{XY}(x^n,y^n)\cdot\frac{\mathbb{1}\{(x^n,\mathsf{m})\in\mathcal{D}_n(\eta)\}}{\Pr(\mathcal{D}_n(\eta))} \tag{97}$$

This implies in particular:

$$P_{\tilde{X}^n}(x^n) \leq P_X^n(x^n)\cdot\Delta_n^{-1}, \tag{98}$$
$$P_{\tilde{Y}^n}(y^n) \leq P_Y^n(y^n)\cdot\Delta_n^{-1}, \tag{99}$$
$$P_{\tilde{\mathsf{M}}}(\mathsf{m}) \leq P_{\mathsf{M}}(\mathsf{m})\cdot\Delta_n^{-1} \tag{100}$$

and

$$D\left(P_{\tilde{X}^n}\|P_X^n\right) \leq \log\Delta_n^{-1}. \tag{101}$$

We are now ready to provide a lower bound on the expected rate and an upper bound on the type-II error exponent with the desired single-letter correspondences in the asymptotic regimes where the blocklength grows to ∞ and the parameters $\mu, \eta \to 0$.

<u>*Lower bound on the expected rate*</u>: Define the random variable $\tilde{L} \triangleq \text{len}(\tilde{\mathsf{M}})$ and notice that by the rate constraint (6):

$$nR \geq \mathbb{E}[L] \tag{102}$$
$$= \mathbb{E}[L|(X^n,\mathsf{M})\in\mathcal{D}_n(\eta)]\cdot P_{X^n\mathsf{M}}(\mathcal{D}_n(\eta)) \\ +\mathbb{E}[L|X^n\notin\mathcal{D}_n(\eta)]\cdot(1-P_{X^n\mathsf{M}}(\mathcal{D}_n(\eta))) \tag{103}$$
$$\geq \mathbb{E}[L|X^n\in\mathcal{D}_n(\eta)]\cdot P_{X^n\mathsf{M}}(\mathcal{D}_n(\eta)) \tag{104}$$
$$= \mathbb{E}\left[\tilde{L}\right]\cdot P_{X^n\mathsf{M}}(\mathcal{D}_n(\eta)) \tag{105}$$
$$\geq \mathbb{E}\left[\tilde{L}\right]\cdot\Delta_n, \tag{106}$$

where (105) holds because $\tilde{\mathsf{M}}$ is obtained by restricting M to the event $(X^n,\mathsf{M})\in\mathcal{D}_n(\eta)$ and $\tilde{L}$ denotes the length of $\tilde{\mathsf{M}}$; and step (106) holds by the definition of Δ_n in (96).

Now, since $\tilde{L}$ is a function of $\tilde{\mathsf{M}}$, we have:

$$H(\tilde{\mathsf{M}}) = H(\tilde{\mathsf{M}},\tilde{L}) \tag{107}$$
$$= H(\tilde{\mathsf{M}}|\tilde{L}) + H(\tilde{L}) \tag{108}$$
$$= \sum_\ell \Pr(\tilde{L}=\ell)H(\tilde{\mathsf{M}}|\tilde{L}=\ell) + H(\tilde{L}) \tag{109}$$
$$\leq \sum_\ell \Pr(\tilde{L}=\ell)\ell + H(\tilde{L}) \tag{110}$$
$$= \mathbb{E}[\tilde{L}] + H(\tilde{L}) \tag{111}$$
$$\leq \frac{nR}{\Delta_n} + H(\tilde{L}) \tag{112}$$
$$\leq \frac{nR}{\Delta_n} + \frac{nR}{\Delta_n}h_{\mathsf{b}}\left(\frac{\Delta_n}{nR}\right) \tag{113}$$
$$= \frac{nR}{\Delta_n}\left(1+h_{\mathsf{b}}\left(\frac{\Delta_n}{nR}\right)\right). \tag{114}$$

Here, (112) follows from (106); and (113) holds because when $\mathbb{E}[\tilde{L}] \leq \frac{nR}{\Delta_n}$, then the entropy of $\tilde{L}$ can be at most that of a Geometric distribution with mean $\frac{nR}{\Delta_n}$, which is $\frac{nR}{\Delta_n}\cdot h_{\mathsf{b}}\left(\frac{\Delta_n}{nR}\right)$.

On the other hand, we can lower bound $H(\tilde{\mathsf{M}})$ in the following way:

$$H(\tilde{\mathsf{M}}) \geq I(\tilde{\mathsf{M}};\tilde{X}^n) \tag{115}$$
$$= H(\tilde{X}^n) - H(\tilde{X}^n|\tilde{\mathsf{M}}) \tag{116}$$

$$= -\sum_{x^n} P_{\tilde{X}^n}(x^n)\log P_{\tilde{X}^n}(x^n) - H(\tilde{X}^n|\tilde{\mathsf{M}}) \tag{117}$$

$$\geq -\sum_{x^n} P_{\tilde{X}^n}(x^n)\log P_{X^n}(x^n) + \log\Delta_n - H(\tilde{X}^n|\tilde{\mathsf{M}}) \tag{118}$$

$$= -\sum_{x^n} P_{\tilde{X}^n}(x^n)\sum_{t=1}^n \log P_X(x_t) + \log\Delta_n - H(\tilde{X}^n|\tilde{\mathsf{M}}) \tag{119}$$

$$= -\sum_{t=1}^n\sum_{x_t} P_{\tilde{X}_t}(x_t)\log P_X(x_t) + \log\Delta_n - H(\tilde{X}^n|\tilde{\mathsf{M}}) \tag{120}$$

$$= \sum_{t=1}^n H(\tilde{X}_t) + \sum_{t=1}^n D(P_{\tilde{X}_t}\|P_X) + \log\Delta_n - H(\tilde{X}^n|\tilde{\mathsf{M}}) \tag{121}$$

$$= \sum_{t=1}^n \Big[H(\tilde{X}_t) - H(\tilde{X}_t|\tilde{M},\tilde{X}^{t-1})\Big] + \sum_{t=1}^n D(P_{\tilde{X}_t}\|P_X) + \log\Delta_n \tag{122}$$

$$= \sum_{t=1}^n I(\tilde{U}_t;\tilde{X}_t) + \sum_{t=1}^n D(P_{\tilde{X}_t}\|P_X) + \log\Delta_n \tag{123}$$

$$= nI(\tilde{U}_T;\tilde{X}_T|T) + \sum_{t=1}^n\sum_{x\in\mathcal{X}} P_{\tilde{X}_T|T=t}(x)\log\frac{P_{\tilde{X}_T|T=t}(x)}{P_X(x)} + \log\Delta_n \tag{124}$$

$$= nI(\tilde{U}_T;\tilde{X}_T|T) + \sum_{t=1}^n\sum_{x\in\mathcal{X}} P_{\tilde{X}_T|T=t}(x)\log\frac{P_{\tilde{X}_T|T=t}(x)}{P_{\tilde{X}_T}(x)} + \sum_{t=1}^n\sum_{x\in\mathcal{X}} P_{\tilde{X}_T|T=t}(x)\log\frac{P_{\tilde{X}_T}(x)}{P_{X_t}(x)} + \log\Delta_n \tag{125}$$

$$= nI(\tilde{U}_T;\tilde{X}_T|T) + nI(\tilde{X}_T;T) + nD(P_{\tilde{X}_T}\|P_{X_T}) + \log\Delta_n \tag{126}$$

$$\geq nI(\tilde{U}_T,T;\tilde{X}_T) + \log\Delta_n \tag{127}$$

$$= nI(\tilde{U};\tilde{X}_T) + \log\Delta_n, \tag{128}$$

where

- (118) holds by (98);
- (119) holds because X^n is i.i.d. under P_X^n;
- (123) holds by defining $\tilde{U}_t \triangleq (\tilde{\mathsf{M}},\tilde{X}^{t-1})$;
- (126) holds because T is chosen uniformly over $\{1,\ldots,n\}$;
- (128) follows by defining $\tilde{U} \triangleq (\tilde{U}_T,T)$.

Combining (114) and (128), we obtain:

$$R \geq \frac{I(\tilde{U};\tilde{X}_T) + \frac{1}{n}\log\Delta_n}{1 + h_{\mathsf{b}}\left(\frac{\Delta_n}{nR}\right)}\cdot\Delta_n, \tag{129}$$

and conclude that in the limit $n\to\infty$ the rate R needs to be lower bounded by the limit of the mutual information $I(\tilde{U};\tilde{X})\frac{1-\eta-\epsilon}{1-\eta}$.

Upper bound on the type-II error exponent: For each string $\mathsf{m}\in\{0,1\}^\star$, define the following set:

$$\mathcal{A}_n(\mathsf{m}) \triangleq \{y^n\colon (\mathsf{m},y^n)\in\mathcal{A}_n\}, \tag{130}$$

By definition of the set $\mathcal{D}_n(\eta)$:

$$P^n_{Y|X}(\mathcal{A}_n(\mathsf{m})|x^n) \geq \eta, \qquad (x^n,\mathsf{m})\in\mathcal{D}_n(\eta). \tag{131}$$

Let now $\{\ell_n\}_{n\geq 1}$ be a sequence satisfying $\lim_{n\to\infty}\ell_n/\sqrt{n}=\infty$ and $\lim_{n\to\infty}\ell_n/n=0$, and define for each $\mathsf{m}\in\mathcal{M}$ the blown up region

$$\hat{\mathcal{A}}_n^{\ell_n}(\mathsf{m}) \triangleq \{\tilde{y}^n\colon \exists y^n\in\mathcal{A}_n(\mathsf{m}) \text{ s.t. } d_{\mathsf{H}}(\tilde{y}^n,y^n)\leq\ell_n\}. \tag{132}$$

By (131) and the blowing-up lemma [22, remark p. 446]:

$$P^n_{Y|X}\left(\hat{\mathcal{A}}_n^{\ell_n}(\mathsf{m})\middle|x^n\right) \geq 1-\frac{\sqrt{n\ln 1/\eta}}{\ell_n} = 1-\lambda_n, \quad (x^n,\mathsf{m})\in\mathcal{D}_n(\eta), \tag{133}$$

where we defined $\lambda_n \triangleq \frac{\sqrt{n\ln 1/\eta}}{\ell_n}$. (Notice that λ_n goes to zero as $n\to\infty$.) Defining the new acceptance region

$$\hat{\mathcal{A}}_n^{\ell_n} \triangleq \bigcup_{m\in\mathcal{M}}\{\mathsf{m}\}\times\hat{\mathcal{A}}_n^{\ell_n}(\mathsf{m}), \tag{134}$$

and taking expectation over (133), we obtain:

$$P_{\tilde{M}\tilde{Y}^n}(\hat{\mathcal{A}}_n^{\ell_n}) = \sum_{(x^n,\mathsf{m})\in\mathcal{D}_n(\eta)} P^n_{Y|X}(\hat{\mathcal{A}}_n^{\ell_n}(\mathsf{m})|x^n)\cdot P_{\tilde{X}^n\tilde{\mathsf{M}}}(x^n,\mathsf{m}) \geq 1-\lambda_n. \tag{135}$$

We next show that the probability of this new acceptance region under the product distribution $P_{\tilde{M}}P_{\tilde{Y}^n}$ is close (in terms of exponential decay rate) to the type-II error probability of our original hypothesis testing problem:

$$P_{\tilde{M}}P_{\tilde{Y}^n}(\hat{\mathcal{A}}_n^{\ell_n}) \leq P_M P_Y^n(\hat{\mathcal{A}}_n^{\ell_n})\cdot\Delta_n^{-2} \tag{136}$$

$$\leq P_M P_Y^n(\mathcal{A}_n)\cdot e^{nh_{\mathsf{b}}(\ell_n/n)}\cdot|\mathcal{Y}|^{\ell_n}\cdot K_n^{\ell_n}\cdot\Delta_n^{-2} \tag{137}$$

$$= \beta_n\cdot e^{nh_{\mathsf{b}}(\ell_n/n)}\cdot|\mathcal{Y}|^{\ell_n}\cdot K_n^{\ell_n}\cdot\Delta_n^{-2}, \tag{138}$$

where we defined $K_n \triangleq \min_{y\colon P_Y(y')>0} P_Y(y)$, and where (136) holds by (98) and (137) by [21, see the Proof of Lemma 5.1]. Define $\delta_n \triangleq -\frac{2}{n}\log\Delta_n + \frac{\ell_n}{n}\log(K_n|\mathcal{Y}|) + h_{\mathsf{b}}(\ell_n/n)$ and notice that $\delta_n\to 0$ as $n\to\infty$. We rewrite (138) as

$$-\frac{1}{n}\log\beta_n \leq -\frac{1}{n}\log P_{\tilde{M}}P_{\tilde{Y}^n}(\hat{\mathcal{A}}_n^{\ell_n}) + \delta_n \tag{139}$$

$$\leq \frac{1}{n(1-\lambda_n)}[D\left(P_{\tilde{\mathsf{M}}\tilde{Y}^n}\|P_{\tilde{\mathsf{M}}}P_{\tilde{Y}^n}\right)+1] + \delta_n \tag{140}$$

$$= \frac{1}{n(1-\lambda_n)}\Big[I(\tilde{\mathsf{M}};\tilde{Y}^n)+1\Big] + \delta_n \tag{141}$$

$$= \frac{1}{n(1-\lambda_n)}\Big[\sum_{t=1}^n I(\tilde{\mathsf{M}};\tilde{Y}_t|\tilde{Y}^{t-1})+1\Big] + \delta_n$$

$$\leq \frac{1}{n(1-\lambda_n)}\Big[\sum_{t=1}^{n} I(\tilde{\mathsf{M}},\tilde{Y}^{t-1};\tilde{Y}_t)+1\Big]+\delta_n \tag{142}$$

$$\leq \frac{1}{n(1-\lambda_n)}\Big[\sum_{t=1}^{n} I(\underbrace{\tilde{\mathsf{M}},\tilde{X}^{t-1}}_{=\tilde{U}_t};\tilde{Y}_t)+1\Big]+\delta_n \tag{143}$$

$$(144)$$

$$= \frac{1}{n(1-\lambda_n)}\Big[\sum_{t=1}^{n} I(\tilde{U}_t;\tilde{Y}_t)+1\Big]+\delta_n \tag{145}$$

$$= \frac{1}{n(1-\lambda_n)}[I(\tilde{U}_T;\tilde{Y}_T|T)+1]+\delta_n \tag{146}$$

$$\leq \frac{1}{1-\lambda_n}[I(\underbrace{\tilde{U}_T,T}_{=\tilde{U}};\tilde{Y}_T)+1]+\delta_n \tag{147}$$

$$\leq \frac{1}{1-\lambda_n}[I(\tilde{U};\tilde{Y}_T)+1]+\delta_n, \tag{148}$$

where

- (140) holds by Lemma 1 and Inequality (135);
- (144) holds by the Markov chain $\tilde{Y}^{t-1}\to(\tilde{\mathsf{M}},\tilde{X}^{t-1})\to\tilde{Y}_t$.

The alphabet of $\tilde{U}$ grows exponentially in n. However, by Charathodory's theorem, for each blocklength n there exists a random variable U_n over an alphabet of size $|\mathcal{X}|+1$ and so that the Markov chain $U_n\to\tilde{X}_T\to\tilde{Y}_T$ and the equalities $I(U_n;\tilde{X}_T)=I(\tilde{U};\tilde{X}_T)$ and $I(U_n;\tilde{Y}_T)=I(\tilde{U};\tilde{Y}_T)$ are satisfied. We can thus replace in (129) and (148) the random variable $\tilde{U}$ by this new random variable U_n.

The proof is then concluded by taking $n\to\infty$ and then $\mu,\eta\to 0$. In fact, recall that $\tilde{X}^n\in\mathcal{T}^{(n)}_{\mu/2}(P_X)$ and $\tilde{Y}^n$ is obtained by passing $\tilde{X}^n$ through the memoryless channel $P_{Y|X}$, which implies that as $n\to\infty$ and $\mu\to 0$ the distribution of $P_{\tilde{X}_T\tilde{Y}_T}$ tends to P_{XY}. By standard continuity considerations, by the modified bounds (129) and (148) with $\tilde{U}$ replaced by U_n, and because all random variables have fixed and finite alphabet sizes, we can then conclude that

$$\overline{\lim_{n\to\infty}} -\frac{1}{n}\log\beta_n \leq I(U;Y) \tag{149}$$

for a random variable U satisfying

$$R\geq I(U;X)(1-\epsilon) \tag{150}$$

and the Markov chain $U\to X\to Y$ and $(X,Y)\sim P_{XY}$. This concludes the proof of the converse.

V. Proof of Converse to Theorem 2

Fix an achievable exponent $\theta<\theta^*_{\text{DMC},\epsilon}(\kappa)$ and a sequence of encoding functions $\{\Phi_1^{(n)},\Phi_2^{(n)},\ldots\}_{n\geq 1}$, stopping functions $\{e_1^{(n)},e_2^{(n)},\ldots\}_{n\geq 1}$, and acceptance/rejection regions $\{\mathcal{A}_n,\mathcal{R}_n\}_{n\geq 1}$ so that (46)–(48) are satisfied. Further fix a large blocklength n, and let $\tau_n, W^{\tau_n}, V^{\tau_n}$ be the stopping time, channel inputs and outputs as implied by these encoding and stopping functions. Let μ,η be small positive real numbers and define

$$\sigma\triangleq\ln(n)\cdot n \tag{151}$$

and a new acceptance region $\mathcal{A}_n^{\text{new}}\subseteq\mathcal{A}_n$ which only contains output sequences v^τ of length not exceeding σ:

$$\mathcal{A}_n^{\text{new}}\triangleq\{(v^\tau,y^n)\in\mathcal{V}^\star\times\mathcal{Y}^n\colon (v^\tau,y^n)\in\mathcal{A}_n \text{ and } \tau\leq\sigma\}. \tag{152}$$

Define also the set

$$\mathcal{D}_n(\eta)\triangleq\Big\{(x^n,w^\sigma)\colon \Pr\Big[(V^{\tau_n},Y^n)\in\mathcal{A}_n^{\text{new}}|\mathcal{H}=H_0,X^n=x^n,W'^\sigma=w^\sigma\Big]\geq\eta\Big\} \cap \Big(\mathcal{T}_\mu^{(n)}(P_X)\times\mathcal{W}^\sigma\Big). \tag{153}$$

Notice that the set $\mathcal{D}_n(\eta)$ is defined in terms of the random variable W'^σ but not W^σ because the actual transmission duration might be shorter than σ, i.e. $\tau_n<\sigma$ is possible.

By standard arguments, we have

$$1-\epsilon\leq P_{V^{\tau_n}Y^n}(\mathcal{A}_n) \tag{154}$$

$$=\Pr[\tau_n\leq\sigma]\cdot P_{V^{\tau_n}Y^n}(\mathcal{A}_n|\tau_n\leq\sigma) + \Pr[\tau_n\geq\sigma]P_{V^{\tau_n}Y^n}(\mathcal{A}_n|\tau_n\geq\sigma) \tag{155}$$

$$\leq P_{V^{\tau_n}Y^n}(\mathcal{A}_n^{\text{new}})+\frac{\mathbb{E}[\tau_n]}{\sigma} \tag{156}$$

$$\leq P_{V^{\tau_n}Y^n}(\mathcal{A}_n^{\text{new}})+\frac{\kappa+\eta}{\ln(n)} \tag{157}$$

$$=\sum_{x^n,w^\sigma}P_{X^nW'^m}(x^n,w^\sigma)\cdot\sum_{(v^\tau,y^n)\in\mathcal{A}_n^{\text{new}}}P_{V^{\tau_n}Y^n|X^nW'^\sigma}(v^\tau,y^n|x^n,w^\sigma)+\frac{\kappa+\eta}{\ln(n)} \tag{158}$$

$$=\sum_{(x^n,w^\sigma)\in\mathcal{D}_n(\eta)}P_{X^nW'^m}(x^n,w^\sigma)\cdot\sum_{(v^\tau,y^n)\in\mathcal{A}_n^{\text{new}}}P_{V^{\tau_n}Y^n|X^nW'^\sigma}(v^\tau,y^n|x^n,w^\sigma) + \sum_{(x^n,w^\sigma)\in\mathcal{D}_n^c(\eta)}P_{X^nW'^\sigma}(x^n,w^\sigma)\cdot\sum_{(v^\tau,y^n)\in\mathcal{A}_n^{\text{new}}}P_{V^{\tau}Y^n|X^nW'^\sigma}(v^\tau,y^n|x^n,w^\sigma)+\frac{\kappa+\eta}{\ln(n)} \tag{159}$$

$$\leq P_{X^nW'^\sigma}(\mathcal{D}_n(\eta))+(1-P_{X^nW'^\sigma}(\mathcal{D}_n(\eta)))\cdot\eta+\frac{\kappa+\eta}{\ln(n)}, \tag{160}$$

where

- (156) follows from the definition of the new acceptance region $\mathcal{A}_n^{\text{new}}$ in (152) and from Markov's inequality;
- (157) follows from (48) and the definition $\sigma=\ln(n)\cdot n$.

This implies

$$P_{X^nW'^\sigma}(\mathcal{D}_n(\eta)) \geq \frac{1-\epsilon-\eta-\frac{\kappa+\eta}{\ln(n)}}{1-\eta} - \frac{|\mathcal{X}|}{2\mu n} \triangleq \Delta_n. \tag{161}$$

Define then the random tuple $(\tilde{X}^n, \tilde{Y}^n, \tilde{\tau}_n, \tilde{W'}^{\sigma}, \tilde{W}^{\tilde{\tau}_n}, \tilde{V}^{\tilde{\tau}_n})$ as the restriction of the tuple $(X^n, Y^n, \tau_n, W'^{\sigma}, W^{\tau_n}, V^{\tau_n})$ to $(X^n, W'^{\sigma}) \in \mathcal{D}_n(\eta)$. (Here we consider both sequences W'^{σ} and $W^{\tilde{\tau}_n}$ but the restriction is only on sequences W'^{σ}.) The restricted pmf is given by

$$P_{\tilde{X}^n\tilde{Y}^n\tilde{\tau}_n\tilde{W}'^\sigma\tilde{W}^{\tilde{\tau}_n}\tilde{V}^{\tilde{\tau}_n}}(x^n, y^n, \tau, w^\sigma, w^\tau, v^\tau) \tag{162}$$

$$\triangleq \frac{P_{X^nW'^\sigma}(x^n, w^\sigma)}{P_{X^nW'^\sigma}(\mathcal{D}_n(\eta))} \cdot \mathbb{1}\{(x^n, w^\sigma) \in \mathcal{D}_n(\eta)\}$$
$$\cdot P^n_{Y|X}(y^n|x^n) \cdot P_{\tau_nW^{\tau_n}V^{\tau_n}|W'^\sigma X^n}(\tau, w^\tau, v^\tau|w^\sigma, x^n), \tag{163}$$

and satisfies

$$P_{\tilde{X}^n\tilde{W}^{\tilde{\tau}_n}}(x^n, w^\tau) \leq P_{X^nW^{\tau_n}}(x^n, w^\tau) \cdot \Delta_n^{-1}, \tag{164}$$
$$P_{\tilde{Y}^n}(y^n) \leq P^n_Y(y^n) \cdot \Delta_n^{-1}, \tag{165}$$
$$P_{\tilde{V}^{\tilde{\tau}_n}}(v^\tau) \leq P_{V^{\tau_n}}(v^\tau) \cdot \Delta_n^{-1}. \tag{166}$$

Communication constraint: Similarly to (106), we obtain:

$$\mathbb{E}[\tau_n] \geq \mathbb{E}[\tilde{\tau}_n] \cdot \Delta_n, \tag{167}$$

Since the original transmission durations $\{\tau_n\}_{n=1}^{\infty}$ have to satisfy (48), for arbitrary $\eta > 0$ and all sufficiently large blocklengths n:

$$\mathbb{E}[\tilde{\tau}_n] \leq \mathbb{E}[\tau_n]\Delta_n^{-1} \leq n(\kappa+\eta)\Delta_n^{-1}, \tag{168}$$

Following the same steps as in (116)–(128) but where $\tilde{\mathsf{M}}$ is replaced by $\tilde{V}^{\tilde{\tau}_n}$, we obtain:

$$I(\tilde{V}^{\tilde{\tau}_n}; \tilde{X}^n) \geq nI(\tilde{U}; \tilde{X}_T) + \log \Delta_n, \tag{169}$$

where here $\tilde{U}$ is defined as $(\tilde{V}^{\tilde{\tau}_n}, \tilde{X}^{T-1}, T)$ for T uniformly distributed over $\{1, \ldots, n\}$ independent of $(\tilde{V}^{\tilde{\tau}_n}, \tilde{X}^n, \tilde{Y}^n)$.

In the following, we upper bound $I(\tilde{V}^{\tilde{\tau}_n}; \tilde{X}^n)$ by n times the capacity C of the DMC $\Gamma_{V|W}$ plus some additive terms that vanish in the asymptotic regimes $n \to \infty$ and $\eta, \mu \to 0$. Define for $i = 1, 2, \ldots$ the random variables $\tilde{L}_i \triangleq \mathbb{1}\{\tilde{\tau}_n \geq i\}$ and

$$\hat{V}_i \triangleq \begin{cases} \tilde{V}_i, & \text{if } \tilde{\tau}_n \geq i, \\ 0, & \text{if } \tilde{\tau}_n < i. \end{cases} \tag{170}$$

Notice that we can write $I(\tilde{V}^{\tilde{\tau}_n}; \tilde{X}^n)$ as:

$$I(\tilde{V}^{\tilde{\tau}_n}; \tilde{X}^n) = I(\tilde{L}^\infty, \hat{V}^\infty; \tilde{X}^n) \tag{171}$$
$$= \sum_{i=1}^{\infty} I(\tilde{L}_i, \hat{V}_i; \tilde{X}^n|\tilde{L}^{i-1}, \hat{V}^{i-1}) \tag{172}$$
$$= \sum_{i=1}^{\infty} I(\tilde{L}_i; \tilde{X}^n|\tilde{L}^{i-1}, \hat{V}^{i-1}) + \sum_{i=1}^{\infty} I(\hat{V}_i; \tilde{X}^n|\tilde{L}^{i}, \hat{V}^{i-1}) \tag{173}$$
$$= \sum_{i=1}^{\infty} I(\tilde{L}_i; \tilde{X}^n|\tilde{L}^{i-1}, \hat{V}^{i-1}) + \sum_{i=1}^{\infty} I(\tilde{V}_i; \tilde{X}^n|\tilde{L}_i = 1, \tilde{L}^{i-1}, \tilde{V}^{i-1}) \cdot \Pr[\tilde{L}_i = 1] \tag{174}$$
$$\leq \sum_{i=1}^{\infty} H(\tilde{L}_i|\tilde{L}^{i-1}) + \sum_{i=1}^{\infty} H(\tilde{V}_i) \cdot \Pr[\tilde{L}_i = 1] - \sum_{i=1}^{\infty} H(\tilde{V}_i|\tilde{L}_i = 1, \tilde{L}^{i-1}, \tilde{W}_i, \tilde{V}^{i-1}, \tilde{X}^n) \cdot \Pr[\tilde{L}_i = 1] \tag{175}$$
$$= H(\tilde{L}^\infty) + \sum_{i=1}^{\infty} \Big(H(\tilde{V}_i) - H(\tilde{V}_i|\tilde{W}_i)\Big) \cdot \Pr[\tilde{L}_i = 1] \tag{176}$$
$$= H(\tilde{L}^\infty) + \sum_{i=1}^{\infty} I(\tilde{V}_i; \tilde{W}_i) \cdot \Pr[\tilde{L}_i = 1] \tag{177}$$
$$\leq H(\tilde{L}^\infty) + C \cdot \sum_{i=1}^{\infty} \Pr[\tilde{L}_i = 1] \tag{178}$$
$$\leq H(\tilde{\tau}_n) + C \cdot \sum_{i=1}^{\infty} \Pr[\tilde{\tau}_n \geq i] \tag{179}$$
$$= H(\tilde{\tau}_n) + C \cdot \mathbb{E}[\tilde{\tau}_n] \tag{180}$$
$$\leq \frac{n(\kappa+\eta)}{\Delta_n} \cdot h_{\text{b}}\left(\frac{\Delta_n}{n(\kappa+\eta)}\right) + nC(\kappa+\eta)\Delta_n^{-1}, \tag{181}$$

where

- (171) holds because there is a bijective function from $(\tilde{L}^\infty, \hat{V}^\infty)$ to $\tilde{V}^{\tilde{\tau}_n}$;
- (174) holds because when $\tilde{L}_i = 0$ then $\hat{V}_i$ is deterministic and when $\tilde{L}_i = 1$ then $\hat{V}_i = \tilde{V}_i$;
- (177) holds because when $\tilde{L}_i = 1$ the Markov chain $\tilde{V}_i \to \tilde{W}_i \to (\tilde{L}^{i-1}, \tilde{V}^{i-1}, \tilde{X}^n)$ holds;
- (178) holds because $P_{\tilde{V}_i|\tilde{W}_i} = \Gamma_{V_i|W_i}$ and thus the mutual information term $I(\tilde{V}_i; \tilde{W}_i)$ is upper bounded by the capacity C of the channel;
- (179) holds because there exists a bijective function from $\tilde{\tau}_n$ to $\tilde{L}^\infty$ and by the definition of $\tilde{L}_i$; and
- (181) holds only for sufficiently large values of n, by (168) and because when $\mathbb{E}[\tilde{\tau}_n] \leq \frac{n(\kappa+\eta)}{\Delta_n}$, then the entropy of $\tilde{\tau}_n$ can be at most that of a Geometric distribution with mean $\frac{n(\kappa+\eta)}{\Delta_n}$, which is $\frac{n(\kappa+\eta)}{\Delta_n} \cdot h_{\text{b}}\left(\frac{\Delta_n}{n(\kappa+\eta)}\right)$.

Combining (128) and (181), we conclude that for all sufficiently large values of n:

$$(\kappa+\eta)C \geq I(\tilde{U}; \tilde{X}_T) \cdot \Delta_n + \frac{\Delta_n}{n} \log \Delta_n - (\kappa+\eta) \cdot h_{\text{b}}\left(\frac{\Delta_n}{n(\kappa+\eta)}\right), \tag{182}$$

and in particular, $(\kappa+\eta)\frac{1-\eta}{1-\eta-\epsilon}C$ upper bounds the limit of the mutual information $I(\tilde{U}; \tilde{X}_T)$ as $n \to \infty$.

Upper bounding the type-II error exponent:
By definition,

$$P_{\tilde{V}^{\tilde{\tau}_n}\tilde{Y}^n|\tilde{X}^n\tilde{W}'^\sigma}(\mathcal{A}_n^{\text{new}}|x^n, w^\sigma) \geq \eta, \quad \forall (x^n, w^\sigma) \in \mathcal{D}_n(\eta).$$

$$(183)$$

We now expand the region $\mathcal{A}_n^{\text{new}}$ to a subset of $\mathcal{V}^\sigma \times \mathcal{Y}^n$, i.e., we expand all channel output sequences to be of same length σ:

$$\mathcal{A}_n^{\text{exp}} \triangleq \big\{(v^\sigma, y^n) \in \mathcal{V}^\sigma \times \mathcal{Y}^n \colon \exists (\tilde{v}^\tau, y^n) \in \mathcal{A}_n \text{ and } \bar{v}^{\sigma-\tau} \colon v^\sigma = (\tilde{v}^\tau, \bar{v}^{\sigma-\tau})\big\}. \quad (184)$$

Similarly, let $\tilde{V}'^\sigma = (\tilde{V}'_1, \ldots, \tilde{V}'_\sigma)$ be outputs of the DMC $\Gamma_{V|W}$ for inputs $\tilde{W}'^\sigma$, and in particular $V'_k = \tilde{V}_k$ with probability 1 when $k \le \tilde{\tau}_n$. Then,

$$P_{\tilde{V}'^\sigma \tilde{Y}^n | \tilde{X}^n \tilde{W}'^\sigma}(\mathcal{A}_n^{\text{exp}} | x^n, w^\sigma) = P_{\tilde{V}^{\tilde{\tau}_n} \tilde{Y}^n | \tilde{X}^n \tilde{W}'^\sigma}(\mathcal{A}_n^{\text{new}} | x^n, w^\sigma) \quad (185)$$

$$\ge \eta. \quad (186)$$

By the blowing-up lemma [22, Remark on p. 446],

$$P_{\tilde{V}'^\sigma \tilde{Y}^n | \tilde{X}^n \tilde{W}'^m}(\hat{\mathcal{A}}_n^{\text{exp},\ell_n} | x^n, w^\sigma) \ge 1 - \frac{\sqrt{(n+\sigma)\ln(1/\eta)}}{\ell_n} \quad (187)$$

$$= 1 - \nu_n, \quad (188)$$

where we defined $\nu_n \triangleq \frac{\sqrt{(n+\sigma)\ln(1/\eta)}}{\ell_n}$ and the blown up region

$$\hat{\mathcal{A}}_n^{\text{exp},\ell_n} \triangleq \{(\tilde{v}^\sigma, \tilde{y}^n) \colon \exists (v^\sigma, y^n) \in \mathcal{A}_n^{\text{exp}} \text{ s.t. } d_{\text{H}}(\tilde{v}^\sigma, v^\sigma) + d_{\text{H}}(\tilde{y}^n, y^n) \le \ell_n\}. \quad (189)$$

Averaging over the sequences $(x^n, w^\sigma) \in \mathcal{D}_n$ we obtain:

$$P_{\tilde{V}'^\sigma \tilde{Y}^n}(\hat{\mathcal{A}}_n^{\text{exp},\ell_n}) \ge 1 - \nu_n. \quad (190)$$

Since $\hat{\mathcal{A}}_n^{\text{exp},\ell_n}$ is the expanded region of $\hat{\mathcal{A}}_n^{\text{new},\ell_n}$, we further have:

$$P_{\tilde{V}^{\tilde{\tau}_n} \tilde{Y}^n}(\hat{\mathcal{A}}_n^{\text{new},\ell_n}) = P_{\tilde{V}'^m \tilde{Y}^n}(\hat{\mathcal{A}}_n^{\text{exp},\ell_n}) \quad (191)$$

$$\ge 1 - \nu_n. \quad (192)$$

Notice next:

$$P_{\tilde{V}^{\tilde{\tau}_n}} P_{\tilde{Y}^n}(\hat{\mathcal{A}}_n^{\text{new},\ell_n}) \le P_{V^{\tau_n}} P_Y^n(\hat{\mathcal{A}}_n^{\text{new},\ell_n}) \cdot \Delta_n^{-2} \quad (193)$$

$$\le P_{V^{\tau_n}} P_Y^n(\mathcal{A}_n^{\text{new}}) \cdot K_n^{\ell_n} \cdot \Delta_n^{-2} \quad (194)$$

$$\le \beta_n \cdot K_n^{\ell_n} \cdot \Delta_n^{-2}, \quad (195)$$

where

$$K_n \triangleq \frac{ne}{\ell_n} pq |\mathcal{Y}||\mathcal{V}| \quad (196)$$

and

$$p \triangleq \max_{y,y' \colon P_Y(y')>0} \frac{P_Y(y)}{P_Y(y')} \quad (197)$$

$$q \triangleq \max_{w,v,v' \colon \Gamma_{V|W}(v'|w)>0} \frac{\Gamma_{V|W}(v|w)}{\Gamma_{V|W}(v'|w)}. \quad (198)$$

Here, (193) holds by (165)–(166); step (194) holds by [21, Proof of Lemma 5.1], and step (195) because the original acceptance region includes the new region: $\mathcal{A}_n \supseteq \mathcal{A}_n^{\text{new}}$.

We use (195) to bound the type-II error exponent of the original test:

$$-\frac{1}{n}\log\beta_n \le -\frac{1}{n}\log P_{\tilde{V}^{\tilde{\tau}_n}} P_{\tilde{Y}^n}(\hat{\mathcal{A}}_n^{\text{new},\ell_n}) - \frac{2}{n}\log\Delta_n + \frac{\ell_n}{n}\log K_n \quad (199)$$

$$\le \frac{1}{n(1-\nu_n)}\left(D(P_{\tilde{V}^{\tilde{\tau}_n}\tilde{Y}^n} \| P_{\tilde{V}^{\tilde{\tau}_n}} P_{\tilde{Y}^n}) + 1\right) - \frac{2}{n}\log\Delta_n + \frac{\ell_n}{n}\log K_n, \quad (200)$$

where the second inequality holds by Lemma 1 stated at the beginning of Appendix IV and by Inequality (190).

We continue to single-letterize the divergence term:

$$\frac{1}{n} D(P_{\tilde{V}^{\tilde{\tau}_n}\tilde{Y}^n} \| P_{\tilde{V}^{\tilde{\tau}_n}} P_{\tilde{Y}^n}) = \frac{1}{n} I(\tilde{V}^{\tilde{\tau}_n}; \tilde{Y}^n) \quad (201)$$

$$= \frac{1}{n}\sum_{t=1}^n I(\tilde{V}^{\tilde{\tau}_n}; \tilde{Y}_t | \tilde{Y}^{t-1}) \quad (202)$$

$$\le \frac{1}{n}\sum_{t=1}^n I(\tilde{V}^{\tilde{\tau}_n}, \tilde{Y}^{t-1}; \tilde{Y}_t) \quad (203)$$

$$\le \frac{1}{n}\sum_{t=1}^n I(\tilde{V}^{\tilde{\tau}_n}, \tilde{X}^{t-1}; \tilde{Y}_t) \quad (204)$$

$$= \frac{1}{n}\sum_{t=1}^n I(\tilde{U}_t; \tilde{Y}_t) \quad (205)$$

$$= I(\tilde{U}_T; \tilde{Y}_T | T) \quad (206)$$

$$\le I(\tilde{U}_T, T; \tilde{Y}_T) \quad (207)$$

$$= I(\tilde{U}; \tilde{Y}_T), \quad (208)$$

where (204) holds by the Markov chain $\tilde{Y}^{t-1} \to (\tilde{V}^{\tilde{\tau}_n}, \tilde{X}^{t-1}) \to \tilde{Y}_t$.

Combining (200) with (148), we obtain:

$$-\frac{1}{n}\log\beta_n \le \frac{1}{1-\nu_n}\left(I(\tilde{U}; \tilde{Y}_T) + \frac{1}{n}\right) - \frac{2}{n}\log\Delta_n + \frac{\ell_n}{n}\log K_n. \quad (209)$$

When $n \to \infty$, then $\nu_n \to 0$, $\frac{2}{n}\log\Delta_n \to 0$, and $\frac{\ell_n}{n}\log K_n \to 0$. So, the asymptotic type-II error exponent is upper bounded by the limit of $I(\tilde{U}; \tilde{Y}_T)$ as $n \to \infty$.

We analyze this limit by first noticing that Charathodory's theorem implies that for each blocklength n there exists a random variable U_n over an alphabet of size $|\mathcal{X}|+1$ satisfying the Markov chain $U_n \to \tilde{X}_T \to \tilde{Y}_T$ and the equalities $I(U_n; \tilde{X}_T) = I(\tilde{U}_T; \tilde{X}_T)$ and $I(U_n; \tilde{Y}_T) = I(\tilde{U}_T; \tilde{Y}_T)$. We can thus replace in (129) and (148) the random variable $\tilde{U}$ by this new random variable U_n.

The proof is concluded by taking $n \to \infty$ and then $\mu, \eta \to 0$. In fact, recall that $\tilde{X}^n \in \mathcal{T}_\mu^{(n)}(P_X)$ and $\tilde{Y}^n$ is obtained by passing $\tilde{X}^n$ through the channel $P_{Y|X}$, which implies that as $n \to \infty$ and $\mu \to 0$ the distribution of $P_{\tilde{X}_T\tilde{Y}_T}$ tends to P_{XY}. By standard continuity considerations, by the modified bounds (182) and (209) with $\tilde{U}$ replaced by U_n, and because all random variables have fixed and finite alphabet sizes, we can then conclude that

$$\varlimsup_{n\to\infty} -\frac{1}{n}\log\beta_n \le I(U;Y) \quad (210)$$

for a random variable U satisfying

$$\frac{\kappa \cdot C}{1-\epsilon} \geq I(U;X) \tag{211}$$

and the Markov chain $U \to X \to Y$ and $(X,Y) \sim P_{XY}$.

VI. Concluding Remarks

We established the optimal type-II error exponent of a distributed testing against independence problem under a constraint on the probability of type-I error and on the expected communication rate. This result can be seen as a variable-length coding version of the well-known result by Ahlswede and Csiszár [1] which holds under a maximum rate-constraint. Interestingly, when the type-I error probability is constrained to be at most $\epsilon \in (0,1)$, then the optimal type-II error exponent under an expected rate constraint R coincides with the optimal type-II error exponent under a maximum rate constraint $R/(1-\epsilon)$. Thus, unlike in the scenario with a maximum rate constraint, here a strong converse does not hold, because the optimal type-II error exponent depends on the allowed type-I error probability ϵ. This latter observation is not surprising in view of similar results that have previously been obtained on variable-length coding for compression systems with positive error probabilities [27]–[29]; in fact also these results show that the minimum compression rate is decreased as a function of the allowed error probability. The contribution of this paper thus rather lies in the converse proofs of the results.

We also considered testing against independence over a DMC with variable-length coding and stop feedback. As we show, the optimal type-II error exponent depends on the DMC transition law only through its capacity. More specifically, under a type-I error probability constraint $\epsilon \in (0,1)$, the optimal type-II error exponent with variable-length coding over a DMC with capacity C coincides with the optimal type-II error exponent under fixed-length coding over a DMC with capacity $C/(1-\epsilon)$. Thus, a strong converse result does not hold for this setup, neither.

The paper considered setups where the marginal distributions are the same under both hypotheses. The presented results hold also when this assumption is relaxed, the important assumption is the independence of the sources under the alternative hypothesis H_1. An interesting future direction is to investigate whether also this assumption can be relaxed and similar results apply for testing against conditional independence. Another interesting line of future research is to study the optimal exponents under variable-length coding when both the type-I and the type-II error probability are required to decay exponentially fast, as considered in [2], [24]. In such a setup the sensor can "give up" (i.e., send a dummy $[0]$ bit-string message) only on a set of source sequences that has a probability of error that decreases at least exponentially fast in the blocklength. However, modifying a coding scheme only with such a small probability has no effect on the expected rate compared to the original scheme, and thus will not yield the desired improvement in exponents compared to the corresponding fixed-length scheme. A more promising approach is to assign different (non-negligible) lengths to different codeword sequences as also considerd in [24], [33]. It would be interesting to see whether such a variable-length coding approach can result in improved exponents. Yet another interesting extension is to analyze the benefits of variable-length coding in the finite block-length regime and to try to improve over Watanabe's error probabilities [25] that are based on fixed-length coding. Like in the two-exponents scenario, a promising approach seems to be to assign different codeword lengths to different source sequences.

Acknowledgements

S. Salehkalaibar and M. Wigger acknowledge funding support from the ERC under grant agreement 715111.

References

[1] R. Ahlswede and I. Csiszàr, "Hypothesis testing with communication constraints," *IEEE Trans. on Info. Theory*, vol. 32, pp. 533–542, Jul. 1986.

[2] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. on Info. Theory*, vol. 33, no. 6, pp. 759–772, Nov. 1987.

[3] H. Shimokawa, T. Han, and S. I. Amari, "Error bound for hypothesis testing with data compression," in *Proc. IEEE Int. Symp. on Info. Theory*, Jul. 1994, p. 114.

[4] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. on Info. Theory*, vol. 58, no. 10, pp. 6282–6303, Oct. 2012.

[5] C. Tian and J. Chen, "Successive refinement for hypothesis testing and lossless one-helper problem," *IEEE Trans. on Info. Theory*, vol. 54, no. 10, pp. 4666–4681, Oct. 2008.

[6] N. Weinberger and Y. Kochman, "On the reliability function of distributed hypothesis testing under optimal detection," *IEEE Trans. on Info. Theory*, vol. 65, no. 8, pp. 4940–4965, Aug. 2019.

[7] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *Proc. 52nd Allerton Conf. Comm, Cont. and Comp.*, Monticello, IL, USA, Oct. 2014, pp. 1246–1251.

[8] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing against independence," in *Proc. IEEE Int. Symp. on Info. Theory*, Istanbul, Turkey, Jun. 2013, pp. 2840–2844.

[9] Y. Chen, R. S. Blum, B. M. Sadler, and J. Zhang, "Testing the structure of a gaussian graphical model with reduced transmissions in a distributed setting," *IEEE Trans. on Sig. Proc.*, vol. 67, no. 20, pp. 5391–5401, Oct. 2019.

[10] S. Zhang, P. Khanduri, and P. K. Varshney, "Distributed sequential detection: dependent observations and imperfect communication," *IEEE Trans. on Sig. Proc.*, vol. 68, pp. 830–842, Nov. 2019.

[11] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Trans. on Info. Theory*, vol. 65, no. 7, pp. 4411–4433, July 2019.

[12] S. Salehkalaibar, M. Wigger, and R. Timo, "On hypothesis testing against independence with multiple decision centers," *IEEE Trans. on Communications*, vol. 66, no. 6, pp. 2409–2420, Jan. 2018.

[13] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detection," in *Proc. IEEE Int. Symp. on Info. Theory*, Jun. 2018.

[14] G. Katz, P. Piantanida, and M. Debbah, "Distributed binary detection with lossy data compression," *IEEE Trans. on Info. Theory*, vol. 63, no. 8, pp. 5207–5227, Aug. 2017.

[15] 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.

[16] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing: cooperation and concurrent detection," *IEEE Trans. on Info. Theory*, 2020.

[17] S. Sreekuma and D. Gündüz, "Distributed hypothesis testing over discrete memoryless channels," *IEEE Trans. on Info. Theory*, vol. 66, no. 4, pp. 2044–2066, Apr. 2020.

[18] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *IEEE Trans. on Info. Theory*, vol. 66, no. 7, July 2020.

[19] Y. Ugur, I. E. Aguerri, and A. Zaidi, “Vector Gaussian CEO problem under logarithmic loss and applications,” *IEEE Trans. on Info. Theory*, vol. 66, no. 7, July 2018.
[20] E. Haim and Y. Kochman, “Binary distributed hypothesis testing via korner-marton coding,” in *Proc. IEEE Information Theory Workshop (ITW)*, 2016.
[21] I. Csiszar and J. Korner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
[22] K. Marton, “A simple proof of the blowing-up lemma,” *IEEE Trans. on Info. Theory*, vol. 32, no. 3, pp. 445–446, May. 1986.
[23] J. Liu, R. van Handel, and S. Verdu, “Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity.” in *Proc. IEEE Int. Symp. on Info. Theory*, Aachen, Germany, Jun. 2017, pp. 943–947.
[24] N. Weinberger, Y. Kochman, and M. Wigger, “Exponent trade-off for hypothesis testing over noisy channels,” in *Proc. IEEE Int. Symp. on Info. Theory*, Paris, France, Jul. 2019, pp. 1852–1856.
[25] S. Watanabe, “Neyman-pearson test for zero-rate multiterminal hypothesis testing,” *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4923–4939, July 2018.
[26] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Feedback in the non-asymptotic regime,” *IEEE Trans. on Info. Theory*, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.
[27] V. Kostina, Y. Polyanskiy, and S. Verdu, “Variable-length compression allowing errors,” *IEEE Trans. on Info. Theory*, vol. 61, no. 8, pp. 4316–4330, 2015.
[28] H. Koga and H. Yamamoto, “Asymptotic properties on codeword lengthsof an optimal FV code for general sources,” *IEEE Trans. on Info. Theory*, vol. 51, no. 4, pp. 1546–1555, 2005.
[29] A. Kimura and T. Uyematsu, “Weak variable-length Slepian-Wolf coding with linked encoders for mixed sources,” *IEEE Trans. on Info. Theory*, vol. 50, no. 1, pp. 183–193, 2004.
[30] S. Salehkalaibar and M. Wigger, “Distributed hypothesis testing over a noisy channel,” in *Int. Zurich Seminar (IZS)*, Zurich, Switzerland, Feb. 2018, pp. 25–29.
[31] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
[32] S. Sreekumar and D. Gündüz, “Strong converse for testing against independence over a noisy channel.” [Online]. Available: https://arxiv.org/pdf/2004.00775.pdf
[33] T. Weissman and N. Merhav, “Tradeoffs between the excess-code-length exponent and the excess-distortion exponent in lossy source coding,” *IEEE Trans. on Info. Theory*, vol. 48, no. 2, pp. 396–415, 2002.

Sadaf Salehkalaibar (S’10–M’14) received the B.Sc., M.Sc. and Ph.D. degrees in Electrical Engineering from Sharif University of Technology, Tehran, Iran in 2008, 2010 and 2014, respectively. She was a postdoctoral fellow at Telecom ParisTech, Paris, France in 2015 and 2017. She is currently an assistant professor at Electrical and Computer Engineering Department of University of Tehran, Tehran, Iran. Her special fields of interest include network information theory, hypothesis testing and fundamental limits of secure communication with emphasis on information-theoretic security.

Michèle Wigger (S’05–M’09–SM’14) received the M.Sc. degree in electrical engineering, with distinction, and the Ph.D. degree in electrical engineering both from ETH Zurich in 2003 and 2008, respectively. In 2009, she was first a post-doctoral fellow at the University of California, San Diego, USA, and then joined Telecom Paris, France, where she is currently a full professor. Dr. Wigger has held visiting professor appointments at the Technion–Israel Institute of Technology and ETH Zurich. Dr. Wigger has previously served as an Associate Editor of the IEEE Communication Letters and as an Associate Editor for Shannon Theory for the IEEE Transactions on Information Theory. During 2016–2019 she also served on the Board of Governors of the IEEE Information Theory Society. Dr. Wigger’s research interests are in multi-terminal information theory, in particular in distributed source coding and in capacities of networks with states, feedback, user cooperation, or caching.