



**HAL**  
open science

# Optimal exponents in cascaded hypothesis testing under expected rate constraints

Mustapha Hamad, Michèle Wigger, Mireille Sarkiss

► **To cite this version:**

Mustapha Hamad, Michèle Wigger, Mireille Sarkiss. Optimal exponents in cascaded hypothesis testing under expected rate constraints. ITW 2021: IEEE Information Theory Workshop, Oct 2021, Kanazawa (online), Japan. pp.1-6, 10.1109/ITW48936.2021.9611470 . hal-03349805

**HAL Id: hal-03349805**

**<https://hal.science/hal-03349805>**

Submitted on 20 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimal Exponents in Cascaded Hypothesis Testing under Expected Rate Constraints

Mustapha Hamad  
*LTCI, Telecom Paris, IP Paris*  
 91120 Palaiseau, France  
 mustapha.hamad@telecom-paris.fr

Michèle Wigger  
*LTCI, Telecom Paris, IP Paris*  
 91120 Palaiseau, France  
 michele.wigger@telecom-paris.fr

Mireille Sarkiss  
*SAMOVAR, Telecom SudParis, IP Paris*  
 91011 Evry, France  
 mireille.sarkiss@telecom-sudparis.eu

**Abstract**—Cascaded binary hypothesis testing is studied in this paper with two decision centers at the relay and the receiver. All terminals have their own observations, where we assume that the observations at the transmitter, the relay, and the receiver form a Markov chain in this order. The communication occurs over two hops, from the transmitter to the relay, and from the relay to the receiver. Expected rate constraints are imposed on both communication links. In this work, we characterize the optimal type-II error exponents at the two decision centers under constraints on the allowed type-I error probabilities. Our recent work characterized the optimal type-II error exponents in the special case when the two decision centers have same type-I error constraints and provided an achievability scheme for the general setup. To obtain the exact characterization for the general case, in this paper we provide a new converse proof as well as a new matching achievability scheme. Our results indicate that under unequal type-I error constraints at the relay and the receiver, a tradeoff arises between the maximum type-II error probabilities at these two terminals. Previous results showed that such a tradeoff does not exist under equal type-I error constraints or under general type-I error constraints when a maximum rate constraint is imposed on the communication links.

**Index Terms**—Multi-hop, distributed hypothesis testing, error exponents, expected rate constraints, variable-length coding.

## I. INTRODUCTION

In a very connected world, where Internet of things (IoT) and sensor networks are emerging widely, distributed hypothesis testing have been utilized for improving decisions under communication constraints. A well-known application is the cascaded hypothesis testing where sensors communicate in a serial way forming a multi-hop network. We consider binary hypothesis testing over a two-hop network composed of a sensor, a relay, a receiver, and two decision centers placed at the relay and the receiver. In such a setup, both decision centers try to correctly guess the binary hypothesis  $\mathcal{H} \in \{0, 1\}$  underlying all terminals' observations including their own. Each decision center aims to maximize the accuracy of its decisions, where the error under the alternative hypothesis  $\mathcal{H} = 1$  (called type-II error) is more critical than the error under the null hypothesis  $\mathcal{H} = 0$  (called type-I error). Specifically, both decision centers aim at maximizing the exponential decay (in the number of observed samples) of the type-II error probabilities under constraints on the accepted type-I error probabilities.

While most information-theoretic works on distributed hypothesis testing constrain the *maximum communication rates* between the terminals [1]–[6], some recent works [7]–[10] have considered *expected rate constraints*. Expected rate constraints were first considered in [7], [8] in a single-sensor single-decision center setup, and the maximum error exponents were

exactly characterized for *testing-against independence* when under the alternative hypothesis the observations are distributed according to the product of the distributions under the null hypothesis. The optimal error exponent for this setup [7], [8], is achieved by a simple coding and decision scheme which chooses an event  $\mathcal{S}_n$  of probability close to the permissible type-I error probability  $\epsilon$ . Under this event, the transmitter sends a single bit to the decision center, allowing it to decide directly on the hypothesis  $\mathcal{H} = 1$ . Otherwise, the transmitter and the receiver run the optimal scheme under the maximum rate constraints [1], [2]. The described scheme achieves same type-II error exponent as in [1], [2], but with a reduced communication rate of  $(1 - \epsilon)^{-1}R$ . This gain is achieved by means of *variable-length coding* which allows to send a message of different rate for each sequence observed at the transmitter. Notice that only under an expected rate constraint, variable-length coding can improve performance, but not under maximum rate constraints. Similar conclusions also hold for more complicated setups, as we showed in [9] for the partially-cooperating multi-access network with two sensors and a single decision center, and in [10] for a special case of the two-hop network studied in this paper.

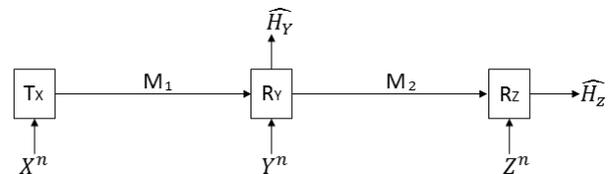


Fig. 1: Cascaded two-hop setup with two decision centers.

We consider the distributed hypothesis testing over the two-hop network in Figure 1, which consists of a transmitter, a relay, and a receiver, and where the observations at the transmitter  $X^n$ , the relay  $Y^n$ , and the receiver  $Z^n$  form a Markov chain  $X^n \rightarrow Y^n \rightarrow Z^n$  under both hypotheses. Under maximum rate-constraints, the optimal type-II error exponents at the relay and the receiver for testing against independence were characterized in [11], [12]. Under expected rate constraints, [10] characterized the optimal type-II error exponents only when the relay and the receiver have same type-I error constraint  $\epsilon > 0$ . The result shows that under equal type-I error probability  $\epsilon > 0$ , maximum type-II error exponents can be simultaneously achieved at both of them. Moreover, the expected rate constraints allow to boost both rates by a factor  $(1 - \epsilon)^{-1}$  as compared to maximum rate-constraints. As in the single-user setup, the optimal exponents are achieved by a simple scheme where the transmitter chooses

an event of probability  $\epsilon$ , and under this event both the transmitter and the relay send a single bit indicating the event to the relay and the receiver, which then decide on  $\mathcal{H} = 1$ , and otherwise the optimal scheme of [11] is run. For the general case, our previous work [10] only provides a set of achievable error exponents but no matching converse.

In this paper, we provide an exact characterization of the optimal error exponents in the general case. We thus recover the main results of [10] as a special case. To obtain our results, we present both a new achievability result as well as a new converse proof.

*Notation:* We follow the notation in [13], [8]. In particular, we use sans serif font for bit-strings: e.g.,  $m$  for a deterministic and  $M$  for a random bit-string. We let  $\text{string}(m)$  denote the shortest bit-string representation of a positive integer  $m$ , and for any bit-string  $m$ , we let  $\text{len}(m)$  and  $\text{dec}(m)$  denote its length and its corresponding positive integer. In addition,  $\mathcal{T}_\mu^{(n)}$  denotes the strongly typical set as defined in [14, Definition 2.8].

## II. SYSTEM MODEL

Consider the distributed hypothesis testing problem in Figure 1 under the Markov chain

$$X^n \rightarrow Y^n \rightarrow Z^n \quad (1)$$

and in the special case of testing against independence, i.e., depending on the binary hypothesis  $\mathcal{H} \in \{0, 1\}$ , the tuple  $(X^n, Y^n, Z^n)$  is distributed as:

$$\text{under } \mathcal{H} = 0 : (X^n, Y^n, Z^n) \sim \text{i.i.d. } P_{XY} \cdot P_{Z|Y}; \quad (2a)$$

$$\text{under } \mathcal{H} = 1 : (X^n, Y^n, Z^n) \sim \text{i.i.d. } P_X \cdot P_Y \cdot P_Z \quad (2b)$$

for given pmfs  $P_{XY}$  and  $P_{Z|Y}$ .

The system consists of a transmitter  $T_X$ , a relay  $R_Y$ , and a receiver  $R_Z$ . The transmitter  $T_X$  observes the source sequence  $X^n$  and sends its bit-string message  $M_1 = \phi_1^{(n)}(X^n)$  to  $R_Y$ , where the encoding function is of the form  $\phi_1^{(n)} : \mathcal{X}^n \rightarrow \{0, 1\}^*$  and satisfies the *expected* rate constraint

$$\mathbb{E}[\text{len}(M_1)] \leq nR_1. \quad (3)$$

The relay  $R_Y$  observes the source sequence  $Y^n$  and with the message  $M_1$  received from  $T_X$ , it produces a guess  $\hat{\mathcal{H}}_Y$  of the hypothesis  $\mathcal{H}$  using a decision function  $g_1^{(n)} : \mathcal{Y}^n \times \{0, 1\}^* \rightarrow \{0, 1\}$ :

$$\hat{\mathcal{H}}_Y = g_1^{(n)}(Y^n, M_1) \in \{0, 1\}. \quad (4)$$

Relay  $R_Y$  also computes a bit-string message  $M_2 = \phi_2^{(n)}(Y^n, M_1)$  using some encoding function  $\phi_2^{(n)} : \mathcal{Y}^n \times \{0, 1\}^* \rightarrow \{0, 1\}^*$  that satisfies the expected rate constraint

$$\mathbb{E}[\text{len}(M_2)] \leq nR_2. \quad (5)$$

Then it sends  $M_2$  to the receiver  $R_Z$ , which guesses hypothesis  $\mathcal{H}$  using its observation  $Z^n$  and the received message  $M_2$ , i.e., using a decision function  $g_2^{(n)} : \mathcal{Z}^n \times \{0, 1\}^* \rightarrow \{0, 1\}$ , it produces the guess:

$$\hat{\mathcal{H}}_Z = g_2^{(n)}(Z^n, M_2) \in \{0, 1\}. \quad (6)$$

The goal is to design encoding and decision functions such that their type-I error probabilities

$$\alpha_{1,n} \triangleq \Pr[\hat{\mathcal{H}}_Y = 1 | \mathcal{H} = 0] \quad (7)$$

$$\alpha_{2,n} \triangleq \Pr[\hat{\mathcal{H}}_Z = 1 | \mathcal{H} = 0] \quad (8)$$

stay below given thresholds  $\epsilon_1 > 0$ ,  $\epsilon_2 > 0$ , and the type-II error probabilities

$$\beta_{1,n} \triangleq \Pr[\hat{\mathcal{H}}_Y = 0 | \mathcal{H} = 1] \quad (9)$$

$$\beta_{2,n} \triangleq \Pr[\hat{\mathcal{H}}_Z = 0 | \mathcal{H} = 1] \quad (10)$$

decay to 0 with largest possible exponential decay.

*Definition 1:* Fix maximum type-I error probabilities  $\epsilon_1, \epsilon_2 \in (0, 1)$  and rates  $R_1, R_2 \geq 0$ . The exponent pair  $(\theta_1, \theta_2)$  is called  $(\epsilon_1, \epsilon_2)$ -*achievable* if there exists a sequence of encoding and decision functions  $\{\phi_1^{(n)}, \phi_2^{(n)}, g_1^{(n)}, g_2^{(n)}\}_{n \geq 1}$  satisfying  $\forall j \in \{1, 2\}$ :

$$\mathbb{E}[\text{len}(M_i)] \leq nR_j, \quad (11)$$

$$\lim_{n \rightarrow \infty} \overline{\alpha}_{j,n} \leq \epsilon_j, \quad (12)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_{j,n}} \geq \theta_j. \quad (13)$$

*Definition 2:* The closure of the set of all  $(\epsilon_1, \epsilon_2)$ -achievable exponent pairs  $(\theta_1, \theta_2)$  is called the  $(\epsilon_1, \epsilon_2)$ -*exponents region* (or *exponents region for short*) and is denoted by  $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ .

The maximum exponents that are achievable at each of the two decision centers are also of interest:

$$\theta_{1,\epsilon_1}^*(R_1) := \max\{\theta_1 : (\theta_1, \theta_2) \in \mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2) \text{ for some } \epsilon_2 > 0, \theta_2 \geq 0\} \quad (14)$$

$$\theta_{2,\epsilon_2}^*(R_1, R_2) := \max\{\theta_2 : (\theta_1, \theta_2) \in \mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2) \text{ for some } \epsilon_1 > 0, \theta_1 \geq 0\}. \quad (15)$$

## III. MAIN RESULTS

Our main result provides an exact characterization of the exponents region  $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ .

*Theorem 1:*  $\forall \epsilon_1 + \epsilon_2 \leq 1$ , the exponents region  $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$  is the set of all  $(\theta_1, \theta_2)$  pairs satisfying

$$\theta_1 \leq \min\{I(U_1; Y), I(U_2; Y)\}, \quad (16a)$$

$$\theta_2 \leq \min\{I(U_2; Y) + I(V_2; Z), I(U_3; Y) + I(V_3; Z)\}, \quad (16b)$$

for some conditional pmfs  $P_{U_1|X}, P_{U_2|X}, P_{U_3|X}, P_{V_2|Y}, P_{V_3|Y}$  and a number  $\sigma \in [1 - (\epsilon_1 + \epsilon_2), 1 - \max\{\epsilon_1, \epsilon_2\}]$  so that

$$R_1 \geq (1 - \epsilon_1 - \sigma)I(U_1; X) + \sigma I(U_2; X) + (1 - \epsilon_2 - \sigma)I(U_3; X), \quad (16c)$$

$$R_2 \geq \sigma I(V_2; Y) + (1 - \epsilon_2 - \sigma)I(V_3; Y). \quad (16d)$$

*Proof:* Achievability is proved in Section IV, and the converse is proved in Section V. ■

It can be shown that in the special case  $\epsilon_1 = \epsilon_2$ , in Theorem 1, one can set without loss in optimality  $\sigma = (1 - \epsilon_1) = (1 - \epsilon_2)$ ,  $U_1 = U_3 = X$ ,  $V_3 = Y$ . This recovers the simpler characterization of the exponents region in [10, Theorem 1]. The result is presented in the following corollary, where for readability we exchanged  $U_2$  by  $U$  and  $V_2$  by  $V$ .

*Corollary 1 (Theorem 1 in [10]):* If  $\epsilon_1 = \epsilon_2 = \epsilon$ , then the exponents region  $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$  is the set of all  $(\theta_1, \theta_2)$  pairs satisfying

$$\theta_1 \leq I(U; Y), \quad (17a)$$

$$\theta_2 \leq I(U; Y) + I(V; Z) \quad (17b)$$

for some conditional pmfs  $P_{U|X}, P_{V|Y}$  so that

$$R_1 \geq (1 - \epsilon)I(U; X), \quad (17c)$$

$$R_2 \geq (1 - \epsilon)I(V; Y). \quad (17d)$$

*Proof:* See [10]. ■

We remark the factors  $(1 - \epsilon)$  in the rate constraints (17c) and (17d) compared to the optimal exponents under a maximum rate constraint determined in [12]. Under equal type-I error probabilities  $\epsilon_1 = \epsilon_2 = \epsilon$ , the *expected* rate constraint thus allows to boost the communication rates by a factor  $(1 - \epsilon)^{-1}$  compared to *maximum rate constraints*. Similar boosts can also be observed in the rate constraints (16c) and (16d) under general maximum type-I error probabilities  $\epsilon_1, \epsilon_2$ .

*Example 1:* In this example, we confirm the benefit of variable-length coding compared to fixed-length coding for general permissible type-I error probabilities. Let  $X, S, T$  be independent Bernoulli random variables of parameters  $p_X = 0.5, p_S = 0.9, p_T = 0.8$  and set  $Y = X \oplus S$  and  $Z = Y \oplus T$ . We consider  $\epsilon_1 = 0.1 > \epsilon_2 = 0.05$  and we plot in Figure 2 the optimal error exponents region  $\mathcal{E}^*$  for  $R_1 = R_2 = 0.5$ , which shows a tradeoff between the two exponents at the relay and the receiver. As already mentioned, such a tradeoff does not exist in the case of equal type-I error probabilities  $\epsilon_1 = \epsilon_2 = 0.05$  (obtained by Corollary 1). Figure 2 illustrates also the gain obtained by the expected rate constraints as opposed to the maximum rate-constraints; in fact, the rectangular region  $\mathcal{E}_{\max R}^*$  shows the maximum exponents region under maximum rate constraints  $R_1 = R_2 = 0.5$  for any values of  $\epsilon_1, \epsilon_2$ . (Under maximum rate constraints a strong converse holds, and the exponents region  $\mathcal{E}_{\max R}^*$  does not depend on  $\epsilon_1, \epsilon_2$ .)

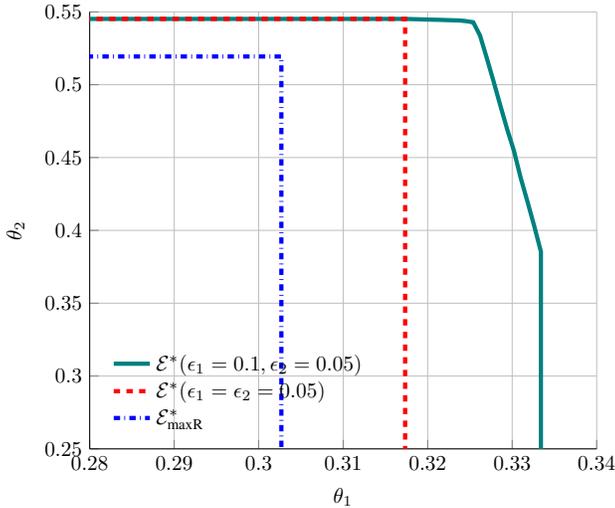


Fig. 2: Error exponents regions under expected and maximum rate constraints when  $\epsilon_1 \geq \epsilon_2$ .

#### IV. GENERAL ACHIEVABILITY SCHEME

We provide a general coding and decision scheme that includes the coding and decision schemes described in [10, Section III]. The idea is to employ three different versions of the basic two-hop scheme [11], depending on the observed sequence  $x^n$ . For each version, we can choose different codebooks, rates, and decision making strategy. To distinguish each case, 2-bit flags are used. Details are as follows.

We first choose a subset  $\mathcal{S}_n \subseteq \mathcal{T}_\mu^{(n)}(P_X)$  of probability

$$\Pr[X^n \in \mathcal{S}_n] = \sigma + \epsilon_1 + \epsilon_2 - 1 - \mu, \quad (18)$$

where  $\mu \in [0, \sigma - (1 - (\epsilon_1 + \epsilon_2))]$ . We partition the remaining subset of  $\mathcal{X}^n$  into three disjoint sets  $\mathcal{D}_1, \mathcal{D}_2$  and  $\mathcal{D}_3$

$$\begin{aligned} \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3 &= \mathcal{X}^n \setminus \mathcal{S}_n \\ \mathcal{D}_i \cap \mathcal{D}_j &= \emptyset, \quad i, j \in \{1, 2, 3\}, i \neq j \end{aligned} \quad (19)$$

such that

$$\Pr[X^n \in \mathcal{D}_1] = 1 - \epsilon_1 - \sigma \quad (20)$$

$$\Pr[X^n \in \mathcal{D}_2] = \sigma + \mu \quad (21)$$

$$\Pr[X^n \in \mathcal{D}_3] = 1 - \epsilon_2 - \sigma. \quad (22)$$

We further split  $R_1 = R_{1,1} + R_{1,2} + R_{1,3}$  and  $R_2 = R_{2,2} + R_{2,3}$  for  $R_{1,1}, R_{1,2}, R_{1,3}, R_{2,2}, R_{2,3} > 0$ .

Whenever  $X^n \in \mathcal{S}_n$ ,  $T_X$  and  $R_Y$  both send the 2-bit flag  $M_1 = M_2 = [0, 0]$ , and  $R_Y$  and  $R_Z$  declare  $\hat{H}_Y = \hat{H}_Z = 1$ .

Whenever  $X^n \in \mathcal{D}_1$ ,  $T_X$  and  $R_Y$  follow the basic single-hop scheme in [1], [2] (which is included in the two-hop scheme [11] as a special case) with a choice of parameters  $\mu, P_{U_1|X}$  satisfying

$$R_{1,1} = (1 - \epsilon_1 - \sigma)(I(U_1; X) + 2\mu), \quad (23)$$

and where  $T_X$  additionally sends a  $[0, 1]$ -flag at the beginning of  $M_1$  to  $R_Y$ , which simply relays this flag  $M_2 = [0, 1]$  without adding additional information. Upon observing  $M_2 = [0, 1]$ ,  $R_Z$  immediately declares  $\hat{H}_Z = 1$ .

Whenever  $X^n \in \mathcal{D}_2$ ,  $T_X, R_Y$ , and  $R_Z$  follow the basic two-hop scheme in [11] but now for a different choice of parameters  $\mu, P_{U_2|X}, P_{V_2|Y}$  satisfying

$$R_{1,2} = (\sigma + \mu)(I(U_2; X) + 2\mu) \quad (24)$$

$$R_{2,2} = (\sigma + \mu)(I(V_2; Y) + 2\mu). \quad (25)$$

Whenever  $X^n \in \mathcal{D}_3$ ,  $T_X, R_Y$ , and  $R_Z$  follow the basic two-hop scheme but now for parameters  $\mu, P_{U_3|X}, P_{V_3|Y}$  satisfying

$$R_{1,3} = (1 - \epsilon_2 - \sigma)(I(U_3; X) + 2\mu) \quad (26)$$

$$R_{2,3} = (1 - \epsilon_2 - \sigma)(I(V_3; Y) + 2\mu), \quad (27)$$

and  $T_X$  and  $R_Y$  add a  $[1, 1]$ -flag to their messages  $M_1$  and  $M_2$  to indicate to  $R_Y$  and  $R_Z$  that  $X^n \in \mathcal{D}_3$ . Here, we note that  $R_Y$ , upon observing the  $[1, 1]$ -flag, declares  $\hat{H}_Y = 1$  even if the computed decision  $\hat{H}_{Y,3}$  following the basic two-hop scheme is different.

In a similar way to [10], it can be shown that this scheme achieves the error exponents in Theorem 1 when  $n \rightarrow \infty$  and  $\mu \downarrow 0$ . Details omitted for brevity and can be found in [15].

#### V. CONVERSE PROOF TO THEOREM 1

Fix an exponent pair in  $\mathcal{E}^*(\epsilon_1, \epsilon_2, R_1, R_2)$  and a sequence (in  $n$ ) of encoding and decision functions  $\{(\phi_1^{(n)}, \phi_2^{(n)}, g_1^{(n)}, g_2^{(n)})\}$  satisfying the constraints on the rate and the error probabilities in (11)–(13). Our proof relies on the following lemma:

*Lemma 1:* Fix a blocklength  $n$  and a set  $\mathcal{D} \subseteq \mathcal{X}^n \times \mathcal{Y}^n$  of positive probability, and let the tuple  $(\tilde{M}_1, \tilde{M}_2, \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n)$  follow the pmf

$$P_{\tilde{M}_1 \tilde{M}_2 \tilde{X}^n \tilde{Y}^n \tilde{Z}^n}(\mathbf{m}_1, \mathbf{m}_2, x^n, y^n, z^n) \triangleq$$

$$P_{X^n Y^n Z^n}(x^n, y^n, z^n) \cdot \frac{\mathbb{1}\{(x^n, y^n) \in \mathcal{D}\}}{P_{X^n Y^n}(\mathcal{D})} \lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} (\Delta_2 + \Delta_3) \geq 1 - \epsilon_2 \quad (46b)$$

$$\cdot \mathbb{1}\{\phi_1(x^n) = \mathbf{m}_1\} \cdot \mathbb{1}\{\phi_2(y^n, \phi_1(x^n)) = \mathbf{m}_2\}. \quad (28) \quad \lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \Delta_2 \geq 1 - \epsilon_1 - \epsilon_2 \quad (46c)$$

Further, define  $U \triangleq (\tilde{M}_1, \tilde{X}^{T-1}, \tilde{Y}^{T-1}, T)$ ,  $V \triangleq (\tilde{M}_2, \tilde{X}^{T-1}, \tilde{Y}^{T-1}, T)$ ,  $\tilde{X} \triangleq \tilde{X}_T$ ,  $\tilde{Y} \triangleq \tilde{Y}_T$ , and  $\tilde{Z} \triangleq \tilde{Z}_T$ , where  $T$  is uniform over  $\{1, \dots, n\}$  and independent of all other random variables. Notice the Markov chain  $V \rightarrow \tilde{Y} \rightarrow \tilde{Z}$ . The following (in)equalities hold:

$$H(\tilde{M}_1) \geq nI(U; \tilde{X}) + \log P_{X^n Y^n}(\mathcal{D}), \quad (29)$$

$$H(\tilde{M}_2) \geq nI(V; \tilde{Y}) + \log P_{X^n Y^n}(\mathcal{D}), \quad (30)$$

$$I(U; \tilde{Y} | \tilde{X}) = \phi_1(n), \quad (31)$$

where  $\phi_1(n)$  is a function that tends to 0 as  $n \rightarrow \infty$ .

Let  $\eta > 0$  be arbitrary. If

$$\Pr[\hat{\mathcal{H}}_Z = 0 | \mathcal{H} = 0, X^n = x^n, Y^n = y^n] \geq \eta, \quad \forall (x^n, y^n) \in \mathcal{D}, \quad (32)$$

then

$$-\frac{1}{n} \log \beta_{2,n} \leq I(U; \tilde{Y}) + I(V; \tilde{Z}) + \phi_2(n), \quad (33)$$

and if

$$\Pr[\hat{\mathcal{H}}_Y = 0 | \mathcal{H} = 0, X^n = x^n, Y^n = y^n] \geq \eta, \quad \forall (x^n, y^n) \in \mathcal{D}, \quad (34)$$

then

$$-\frac{1}{n} \log \beta_{1,n} \leq I(U; \tilde{Y}) + \phi_3(n), \quad (35)$$

where  $\phi_2(n), \phi_3(n)$  are functions that tend to 0 as  $n \rightarrow \infty$ .

*Proof:* See Appendix A.  $\blacksquare$

We now prove the converse to Theorem 1. Fix a positive  $\eta > 0$ . Denote for each blocklength  $n$ , the set of strongly jointly typical sequences in  $\mathcal{X}^n \times \mathcal{Y}^n$  by  $\mathcal{T}_{\mu_n}^{(n)}(P_{XY})$ . Set  $\mu_n = n^{-1/3}$  and define the sets

$$\mathcal{B}_1(\eta) \triangleq \{(x^n, y^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{XY}) : \Pr[\hat{\mathcal{H}}_Y = 0 | X^n = x^n, Y^n = y^n, \mathcal{H} = 0] \geq \eta\}, \quad (36)$$

$$\mathcal{B}_2(\eta) \triangleq \{(x^n, y^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{XY}) : \Pr[\hat{\mathcal{H}}_Z = 0 | X^n = x^n, Y^n = y^n, \mathcal{H} = 0] \geq \eta\}, \quad (37)$$

$$\mathcal{D}_2(\eta) \triangleq \mathcal{B}_1(\eta) \cap \mathcal{B}_2(\eta), \quad (38)$$

$$\mathcal{D}_1(\eta) \triangleq \mathcal{B}_1(\eta) \setminus \mathcal{D}_2(\eta), \quad (39)$$

$$\mathcal{D}_3(\eta) \triangleq \mathcal{B}_2(\eta) \setminus \mathcal{D}_2(\eta). \quad (40)$$

Further define for each  $n$  the probabilities

$$\Delta_i \triangleq P_{X^n Y^n}(\mathcal{D}_i(\eta)), \quad i \in \{1, 2, 3\}, \quad (41)$$

and notice that by the laws of probability

$$\Delta_1 + \Delta_2 = P_{X^n Y^n}(\mathcal{B}_1(\eta)) \quad (42)$$

$$\Delta_2 + \Delta_3 = P_{X^n Y^n}(\mathcal{B}_2(\eta)) \quad (43)$$

$$\Delta_2 \geq P_{X^n Y^n}(\mathcal{B}_1(\eta)) + P_{X^n Y^n}(\mathcal{B}_2(\eta)) - 1. \quad (44)$$

By [14, Remark to Lemma 2.12] and the type-I error probability constraints in (12):

$$P_{X^n Y^n}(\mathcal{B}_j(\eta)) \geq \frac{1 - \epsilon_j - \eta}{1 - \eta} - \frac{|\mathcal{X}||\mathcal{Y}|}{4\mu_n^2 n}, \quad j \in \{1, 2\}, \quad (45)$$

and we thus conclude that in the limit  $n \rightarrow \infty$  and  $\eta \downarrow 0$ :

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} (\Delta_1 + \Delta_2) \geq 1 - \epsilon_1 \quad (46a)$$

$$\lim_{\eta \downarrow 0} \lim_{n \rightarrow \infty} \sum_{i \in \{1, 2, 3\}} \Delta_i \leq 1. \quad (46d)$$

Fix a blocklength  $n$  where  $\Delta_i > 0$  for all  $i \in \{1, 2, 3\}$ . Degenerate cases can be treated similarly. We proceed by applying Lemma 1 to each set  $\mathcal{D}_i$ , to conclude that

$$H(\tilde{M}_{1,i}) \geq nI(U_i; \tilde{X}_i) + \log P_{X^n Y^n}(\mathcal{D}_i), \quad i \in \{1, 2, 3\}, \quad (47)$$

$$H(\tilde{M}_{2,i}) \geq nI(V_i; \tilde{Y}_i) + \log P_{X^n Y^n}(\mathcal{D}_i), \quad i \in \{2, 3\}, \quad (48)$$

$$I(U_i; \tilde{Y} | \tilde{X}) = \phi_{1,i}(n), \quad (49)$$

and

$$-\frac{1}{n} \log \beta_{1,n} \leq \min\{I(U_1; \tilde{Y}); I(U_2; \tilde{Y})\} + \phi_2(n), \quad (50)$$

$$-\frac{1}{n} \log \beta_{2,n} \leq \min\{I(U_2; \tilde{Y}) + I(V_2; \tilde{Z}); I(U_3; \tilde{Y}) + I(V_3; \tilde{Z})\} + \phi_3(n), \quad (51)$$

where for each  $i$ , the functions  $\phi_{1,i}(n), \phi_2(n), \phi_3(n) \rightarrow 0$  as  $n \rightarrow \infty$  and the random variables  $U_i, V_i, \tilde{X}_i, \tilde{Y}_i, \tilde{Z}_i, \tilde{M}_{1,i}, \tilde{M}_{2,i}$  are defined as in the lemma applied to the subset  $\mathcal{D}_i$ . Further define the following random variables for  $j \in \{1, 2\}$  and  $i \in \{1, 2, 3\}$

$$\tilde{L}_{j,i} \triangleq \text{len}(\tilde{M}_{j,i}). \quad (52)$$

By the rate constraints (3) and (5), and the definition of the random variables  $\tilde{M}_{j,i}$ , we obtain by the total law of expectations:

$$nR_1 \geq \mathbb{E}[L_1] \quad (53)$$

$$\geq \sum_{i \in \{1, 2, 3\}} \mathbb{E}[\tilde{L}_{1,i}] \Delta_i. \quad (54)$$

Moreover,

$$H(\tilde{M}_{1,i}) = H(\tilde{M}_{1,i}, \tilde{L}_{1,i}) \quad (55)$$

$$= \sum_{l_i} \Pr[\tilde{L}_{1,i} = l_i] H(\tilde{M}_{1,i} | \tilde{L}_{1,i} = l_i) + H(\tilde{L}_{1,i}) \quad (56)$$

$$\leq \sum_{l_i} \Pr[\tilde{L}_{1,i} = l_i] l_i + H(\tilde{L}_{1,i}) \quad (57)$$

$$= \mathbb{E}[\tilde{L}_{1,i}] + H(\tilde{L}_{1,i}), \quad (58)$$

which combined with (54) establishes

$$\sum_{i \in \{1, 2, 3\}} \Delta_i H(\tilde{M}_{1,i}) \leq \sum_{i \in \{1, 2, 3\}} \Delta_i \mathbb{E}[\tilde{L}_{1,i}] + \Delta_i H(\tilde{L}_{1,i}) \quad (59)$$

$$\leq nR_1 \left( 1 + \sum_{i \in \{1, 2, 3\}} h_b \left( \frac{\Delta_i}{nR_1} \right) \right), \quad (60)$$

where (60) holds by (54) and because the entropy of the discrete and positive random variable  $\tilde{L}_{1,i}$  of mean  $\mathbb{E}[\tilde{L}_{1,i}] \leq \frac{nR_1}{\Delta_i}$  is bounded by  $\frac{nR_1}{\Delta_i} \cdot h_b \left( \frac{\Delta_i}{nR_1} \right)$ , see [16, Theorem 12.1.1].

In a similar way we obtain

$$\sum_{i \in \{2, 3\}} \Delta_i H(\tilde{M}_{2,i}) \leq nR_2 \left( 1 + \sum_{i \in \{2, 3\}} h_b \left( \frac{\Delta_i}{nR_2} \right) \right). \quad (61)$$

As illustrated with more details in [15], combining (60) and (61) with (47) and (48), noting (42) and (45), and considering

also (50) and (51), we can show by Carathéodory's theorem [13, Appendix C] and by invoking Bolzano-Weierstrass theorem that there exist limiting pmfs  $P_{U_i}^*, P_{V_i}^*$  (as  $n \rightarrow \infty$  and  $\eta \downarrow 0$ ) of the sequences  $P_{U_i|\tilde{X}\tilde{Y}} = P_{U_i|\tilde{X}}P_{\tilde{X}\tilde{Y}}$  for  $i \in \{1, 2, 3\}$ , and  $P_{V_i|\tilde{Y}\tilde{Z}} = P_{V_i|\tilde{Y}}P_{\tilde{Y}\tilde{Z}}$  for  $i \in \{2, 3\}$  satisfying the Markov conditions  $U_i \rightarrow \tilde{X} \rightarrow \tilde{Y}$ ,  $V_i \rightarrow \tilde{Y} \rightarrow \tilde{Z}$ , and where the following conditions hold:

$$R_1 \geq \sigma_1 \cdot IP_{U_1}^*(U_1; X) + \sigma_2 \cdot IP_{U_2}^*(U_2; X) + \sigma_3 \cdot IP_{U_3}^*(U_3; X), \quad (62)$$

$$R_2 \geq \sigma_2 \cdot IP_{V_2}^*(V_2; Y) + \sigma_3 \cdot IP_{V_3}^*(V_3; Y), \quad (63)$$

$$\theta_1 \leq \min\{IP_{U_1}^*(U_1; Y), IP_{U_2}^*(U_2; Y)\}, \quad (64)$$

$$\theta_2 \leq \min\{IP_{V_2}^*(U_2; Y) + IP_{V_2}^*(V_2; Z), IP_{U_3}^*(U_3; Y) + IP_{V_3}^*(V_3; Z)\}, \quad (65)$$

where  $\sigma_1, \sigma_2, \sigma_3 > 0$  are numbers satisfying  $\sigma_1 + \sigma_2 + \sigma_3 \leq 1$ ,  $\sigma_1 + \sigma_2 \geq 1 - \epsilon_1$ ,  $\sigma_2 \geq 1 - \epsilon_1 - \epsilon_2$ , and  $\sigma_2 + \sigma_3 \geq 1 - \epsilon_2$ . Notice further that since for any  $k$  the pair  $(\tilde{X}^{n_k}, \tilde{Y}^{n_k})$  lies in the jointly typical set  $\mathcal{T}_{\mu_{n_k}}^{(n_k)}(P_{\tilde{X}\tilde{Y}})$ , we have  $|P_{\tilde{X}\tilde{Y}} - P_{\tilde{X}\tilde{Y}}| \leq \mu_{n_k}$  and thus the limiting pmfs satisfy  $P_{\tilde{X}\tilde{Y}}^* = P_{\tilde{X}\tilde{Y}}$ . Moreover, since for each  $n_k$  the random variable  $\tilde{Z}$  is drawn according to  $P_{\tilde{Z}|\tilde{Y}}$  given  $\tilde{Y}$ , irrespective of  $\tilde{X}$ , the limiting pmf also satisfies  $P_{\tilde{Z}|\tilde{X}\tilde{Y}}^* = P_{\tilde{Z}|\tilde{Y}}$ . Finally, one can argue that without loss of generality one can restrict to  $\sigma_2 \leq \min\{1 - \epsilon_1, 1 - \epsilon_2\}$  and  $\sigma_1 + \sigma_2 = 1 - \epsilon_1, \sigma_2 + \sigma_3 = 1 - \epsilon_2$ .

#### ACKNOWLEDGMENT

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

#### APPENDIX A PROOF OF LEMMA 1

Note first that by (28):

$$D(P_{\tilde{X}^n\tilde{Y}^n} \| P_{\tilde{X}\tilde{Y}}^n) \leq \log \Delta_n^{-1}, \quad (66)$$

where we defined  $\Delta_n \triangleq P_{\tilde{X}^n\tilde{Y}^n}(\mathcal{D})$ .

Further define  $\tilde{V}_t \triangleq (\tilde{M}_2, \tilde{X}^{t-1}, \tilde{Y}^{t-1})$  and  $\tilde{U}_t \triangleq (\tilde{M}_1, \tilde{X}^{t-1}, \tilde{Y}^{t-1})$  and notice:

$$H(\tilde{M}_1) \geq I(\tilde{M}_1; \tilde{X}^n\tilde{Y}^n) + D(P_{\tilde{X}^n\tilde{Y}^n} \| P_{\tilde{X}\tilde{Y}}^n) + \log \Delta_n \quad (67)$$

$$\geq n[H(\tilde{X}_T\tilde{Y}_T) + D(P_{\tilde{X}_T\tilde{Y}_T} \| P_{\tilde{X}\tilde{Y}})]$$

$$- \sum_{t=1}^n H(\tilde{X}_t\tilde{Y}_t | \tilde{U}_t) + \log \Delta_n \quad (68)$$

$$\geq n[H(\tilde{X}_T\tilde{Y}_T) - H(\tilde{X}_T\tilde{Y}_T | \tilde{U}_T, T)] + \log \Delta_n \quad (69)$$

$$\geq n \left[ I(\tilde{X}; U) + \frac{1}{n} \log \Delta_n \right]. \quad (70)$$

Here, (67) holds by (66); (68) holds by the super-additivity property in [17, Proposition 1], by the chain rule, and by the definition of  $\tilde{U}_t$ ; (69) by defining  $T$  uniform over  $\{1, \dots, n\}$  independent of all other random variables; and (70) by the definitions of  $U, \tilde{X}, \tilde{Y}$  in the lemma.

We can lower bound the entropy of  $\tilde{M}_2$  in a similar way:

$$H(\tilde{M}_2) \geq n \left[ I(\tilde{Y}; V) + \frac{1}{n} \log \Delta_n \right]. \quad (71)$$

We next upper bound the error exponent at the receiver. Define

$$\mathcal{A}_{Z,n}(\mathfrak{m}_2) \triangleq \{z^n : g_2(z^n, \mathfrak{m}_2) = 0\}, \quad (72)$$

and its Hamming neighborhood:

$$\hat{\mathcal{A}}_{Z,n}^{\ell_n}(\mathfrak{m}_2) \triangleq \{\tilde{z}^n : \exists z^n \in \mathcal{A}_{Z,n}(\mathfrak{m}_2) \text{ s.t. } d_H(z^n, \tilde{z}^n) \leq \ell_n\} \quad (73)$$

for some real number  $\ell_n$  satisfying  $\lim_{n \rightarrow \infty} \ell_n/n = 0$  and  $\lim_{n \rightarrow \infty} \ell_n/\sqrt{n} = \infty$ . Since by Condition (32),

$$P_{\tilde{Z}^n|\tilde{X}^n\tilde{Y}^n}(\mathcal{A}_{Z,n}(\mathfrak{m}_2)|x^n, y^n) \geq \eta, \quad \forall (x^n, y^n) \in \mathcal{D}, \quad (74)$$

the blowing-up lemma [18] yields

$$P_{\tilde{Z}^n|\tilde{X}^n\tilde{Y}^n}(\hat{\mathcal{A}}_{Z,n}^{\ell_n}(\mathfrak{m}_2)|x^n, y^n) \geq 1 - \zeta_n, \quad \forall (x^n, y^n) \in \mathcal{D}, \quad (75)$$

for a real number  $\zeta_n > 0$  such that  $\lim_{n \rightarrow \infty} \zeta_n = 0$ .

Define

$$\mathcal{A}_{Z,n} \triangleq \bigcup_{\mathfrak{m}_2 \in \mathcal{M}_2} \{\mathfrak{m}_2\} \times \mathcal{A}_{Z,n}(\mathfrak{m}_2), \quad (76)$$

$$\hat{\mathcal{A}}_{Z,n}^{\ell_n} \triangleq \bigcup_{\mathfrak{m}_2 \in \mathcal{M}_2} \{\mathfrak{m}_2\} \times \hat{\mathcal{A}}_{Z,n}^{\ell_n}(\mathfrak{m}_2), \quad (77)$$

and notice that by (75) and (77),

$$P_{\tilde{M}_2\tilde{Z}^n}(\hat{\mathcal{A}}_{Z,n}^{\ell_n}) \geq (1 - \zeta_n). \quad (78)$$

Defining

$$Q_{\tilde{M}_2}(\mathfrak{m}_2) \triangleq \sum_{y^n, \mathfrak{m}_1} P_{\tilde{M}_1}(\mathfrak{m}_1) P_{\tilde{Y}^n}(y^n) \cdot \mathbb{1}\{\phi_2(y^n, \mathfrak{m}_1) = \mathfrak{m}_2\}, \quad (79)$$

we can write

$$Q_{\tilde{M}_2} P_{\tilde{Z}^n}(\hat{\mathcal{A}}_{Z,n}^{\ell_n}) \leq Q_{\mathcal{M}_2} P_{\tilde{Z}^n}(\hat{\mathcal{A}}_{Z,n}^{\ell_n}) \Delta_n^{-3} \quad (80)$$

$$\leq Q_{\mathcal{M}_2} P_{\tilde{Z}^n}(\mathcal{A}_{Z,n}) \cdot e^{n\delta_n} \quad (81)$$

where (81) holds by [14, Proof of Lemma 5.1],  $\delta_n \triangleq h_b(\ell_n/n) + \frac{\ell_n}{n} \log(|\mathcal{Z}| \cdot k_n) - \frac{3}{n} \log \Delta_n$ ,  $\beta_{2,n} = Q_{\mathcal{M}_2} P_{\tilde{Z}^n}(\mathcal{A}_{Z,n})$ ,  $k_n \triangleq \min_{z, z': P_{\tilde{Z}}(z) > 0} \frac{P_{\tilde{Z}}(z)}{P_{\tilde{Z}}(z')}$ .

By standard inequalities (see [8, Lemma 1]), we obtain the following expression:

$$-\frac{1}{n} \log \beta_{2,n} \leq \frac{1}{n(1 - \zeta_n)} D(P_{\tilde{M}_2\tilde{Z}^n} \| Q_{\tilde{M}_2} P_{\tilde{Z}^n}) + \delta_n \quad (82)$$

where  $\delta_n$  tends to 0 as  $n \rightarrow \infty$ , and we can upper bound the divergence term as

$$D(P_{\tilde{M}_2\tilde{Z}^n} \| Q_{\tilde{M}_2} P_{\tilde{Z}^n}) = I(\tilde{M}_2; \tilde{Z}^n) + D(P_{\tilde{M}_2} \| Q_{\tilde{M}_2}) \quad (83)$$

$$\leq I(\tilde{M}_2; \tilde{Z}^n) + D(P_{\tilde{Y}^n\tilde{M}_1} \| P_{\tilde{Y}^n} P_{\tilde{M}_1}) \quad (84)$$

$$\leq \sum_{t=1}^n I(\tilde{M}_2\tilde{Y}^{t-1}; \tilde{Z}_t) + I(\tilde{M}_1\tilde{X}^{t-1}; \tilde{Y}_t) \quad (85)$$

$$= n[I(\tilde{V}_T; \tilde{Z}_T | T) + I(\tilde{U}_T; \tilde{Y}_T | T)] \quad (86)$$

$$\leq n[I(V; \tilde{Z}) + I(U; \tilde{Y})]. \quad (87)$$

Here (84) is obtained by data processing inequality for relative entropy; (85) by the chain rule and by the Markov chains  $\tilde{Z}^{t-1} \rightarrow \tilde{Y}^{t-1} \rightarrow \tilde{Z}_t$  and  $\tilde{Y}^{t-1} \rightarrow \tilde{X}^{t-1} \rightarrow \tilde{Y}_t$ ; and (86)–(87) by definitions of  $\tilde{U}_t, \tilde{V}_t, U, V, \tilde{Y}, \tilde{Z}$ .

Following similar steps, one can prove also the desired upper bound for  $\beta_{1,n}$  if (34) is satisfied. A full detailed proof is presented in [15].

## REFERENCES

- [1] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.
- [2] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, pp. 759–772, Nov. 1987.
- [3] H. Shimokawa, T. Han, and S. I. Amari, "Error bound for hypothesis testing with data compression," in *Proc. ISIT*, p. 114, Jul. 1994.
- [4] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, pp. 6282–6303, Oct. 2012.
- [5] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detection," in *Proc. ISIT*, Jun. 2018.
- [6] W. Zhao and L. Lai, "Distributed testing with cascaded encoders," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7339–7348, 2018.
- [7] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," in *18th Int. Symp. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, 2020.
- [8] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 681–694, 2020.
- [9] M. Hamad, M. Wigger, and M. Sarkiss, "Cooperative multi-sensor detection under variable-length coding." [Online]. Available: <https://arxiv.org/pdf/2010.09616.pdf>, 2020.
- [10] M. Hamad, M. Wigger, and M. Sarkiss, "Two-hop network with multiple decision centers under expected-rate constraints." [Online]. <https://arxiv.org/pdf/2105.05917.pdf>, 2021.
- [11] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Trans. Inf. Theory*, vol. 65, pp. 4411–4433, Jul. 2019.
- [12] 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.
- [13] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [14] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
- [15] M. Hamad, M. Wigger, and M. Sarkiss, "Optimal exponents in cascaded hypothesis testing under expected rate constraints." [Online]. <https://arxiv.org/pdf/2106.12299.pdf>, 2021.
- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed.* Wiley, 2006.
- [17] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 689–703, 2019.
- [18] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.