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, and 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 ϵ. 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 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.

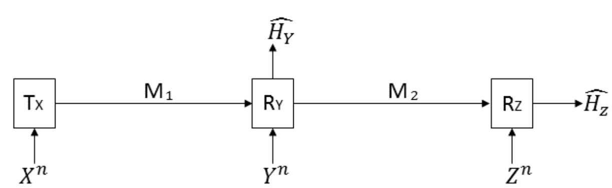

Fig. 1: Cascaded two-hop setup with two decision centers.

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 $X^n \to Y^n \to Z^n$ under both hypothesis. 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 simultaneously be 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 ϵ, 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 $\mathrm{string}(m)$ denote the shortest bit-string representation of a positive integer m, and for any bit-string m we let $\mathrm{len}(\mathsf{m})$ and $\mathrm{dec}(\mathsf{m})$ 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 Fig. 1 under the Markov chain

$$X^n \to Y^n \to Z^n \tag{1}$$

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:

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

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

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 $\mathsf{M}_1 = \phi_1^{(n)}(X^n)$ to R_Y, where the encoding function is of the form $\phi_1^{(n)} : \mathcal{X}^n \to \{0,1\}^\star$ and satisfies the *expected* rate constraint

$$\mathbb{E}\left[\mathrm{len}\left(\mathsf{M}_1\right)\right] \leq nR_1. \tag{3}$$

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 $g_1^{(n)} : \mathcal{Y}^n \times \{0,1\}^\star \to \{0,1\}$:

$$\hat{\mathcal{H}}_Y = g_1^{(n)}\left(\mathsf{M}_1, Y^n\right) \in \{0,1\}. \tag{4}$$

Relay R_Y also computes a bit-string message $\mathsf{M}_2 = \phi_2^{(n)}(Y^n, \mathsf{M}_1)$ using some encoding function $\phi_2^{(n)} : \mathcal{Y}^n \times \{0,1\}^\star \to \{0,1\}^\star$ that satisfies the expected rate constraint

$$\mathbb{E}\left[\mathrm{len}\left(\mathsf{M}_2\right)\right] \leq nR_2. \tag{5}$$

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 $g_2^{(n)} : \mathcal{Z}^n \times \{0,1\}^\star \to \{0,1\}$, it produces the guess:

$$\hat{\mathcal{H}}_Z = g_2^{(n)}\left(\mathsf{M}_2, Z^n\right) \in \{0,1\}. \tag{6}$$

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

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

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

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

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

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

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 (θ_1, θ_2) is called (ϵ_1, ϵ_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 $\forall j \in \{1,2\}$:

$$\mathbb{E}[\mathrm{len}(\mathsf{M}_i)] \leq nR_j, \tag{11}$$

$$\varlimsup_{n\to\infty} \alpha_{j,n} \leq \epsilon_j, \tag{12}$$

$$\varliminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_{j,n}} \geq \theta_j. \tag{13}$$

Definition 2: The closure of the set of all (ϵ_1, ϵ_2)-achievable exponent pairs (θ_1, θ_2) is called the (ϵ_1, ϵ_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:

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

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

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 (θ_1, θ_2) pairs satisfying

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

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

for some conditional pmfs $P_{U_1|X}, P_{U_2|X}, P_{U_3|X}, P_{V_1|Y}, P_{V_2|Y}$ and a number $\sigma \in [1 - (\epsilon_1 + \epsilon_2), 1 - \max\{\epsilon_1, \epsilon_2\}]$ so that

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

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

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 (θ_1, θ_2) pairs satisfying

$$\theta_1 \leq I(U;Y), \tag{17a}$$

$$\theta_2 \leq I(U;Y) + I(V;Z) \tag{17b}$$

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

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

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

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 ϵ_1, ϵ_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 Fig. 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). Fig. 2 illustrates also the gain obtained by the expected rate constraints as opposed to the maximum rate-constraint; in fact, the rectangular region $\mathcal{E}^*_{\text{maxR}}$ shows the maximum exponents region under maximum rate constraints $R_1 = R_2 = 0.5$ for any values of ϵ_1, ϵ_2. (Under maximum rate constraints a strong converse holds, and the exponents region $\mathcal{E}^*_{\text{maxR}}$ does not depend on ϵ_1, ϵ_2.)

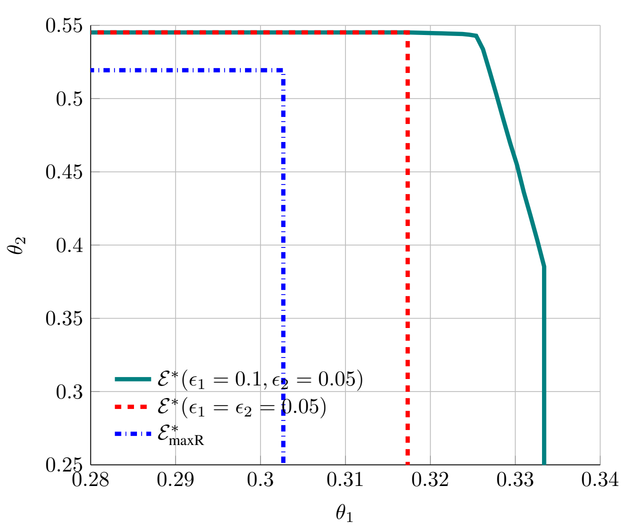

Fig. 2: Error exponents regions under expected and maximum rate constraints when $\epsilon_1 \geq \epsilon_2$.

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

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

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 $\mathcal{D}_3$

$$\mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3 = \mathcal{X}^n \backslash \mathcal{S}_n$$
$$\mathcal{D}_i \cap \mathcal{D}_j = \emptyset, \quad i,j \in \{1,2,3\}, i \neq j \tag{19}$$

such that

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

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

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

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 $\mathsf{M}_1 = \mathsf{M}_2 = [0,0]$, and R_Y and R_Z declare $\hat{\mathcal{H}}_Y = \hat{\mathcal{H}}_Z = 1$.

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

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

and where T_X additionally sends a $[0,1]$-flag at the beginning of M_1 to R_Y, which simply relays this flag $\mathsf{M}_2 = [0,1]$ without adding additional information. Upon observing $\mathsf{M}_2 = [0,1]$, R_Z immediately declares $\hat{\mathcal{H}}_Z = 1$.

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

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

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

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

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

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

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 $\hat{\mathcal{H}}_Y = 1$ even if the computed decision $\hat{\mathcal{H}}_{Y,3}$ 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 $n \to \infty$ and $\mu \downarrow 0$. Details are presented in Appendix A.

V. Converse Proof to Theorem 1

Fix $\theta_1 < \theta^*_{1,\epsilon_1}(R_1)$, $\theta_2 < \theta^*_{2,\epsilon_2}(R_1, R_2)$, a sequence of encoding and decision functions satisfying the type-I and type-II error constraints and a blocklength n. Our proof relies on the following lemma:

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

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

$$P_{X^nY^nZ^n}(x^n,y^n,z^n)\cdot\frac{\mathbb{1}\{x^n\in\mathcal{D}\}}{P_{X^n}(\mathcal{D})}$$
$$\cdot\mathbb{1}\{\phi_1(x^n)=\mathsf{m}_1\}\cdot\mathbb{1}\{\phi_2(y^n,\phi_1(x^n))=\mathsf{m}_2\}. \quad (28)$$

Further, define $U=(\tilde{\mathsf{M}}_1,\tilde{X}^{T-1},T), V=(\tilde{\mathsf{M}}_2,\tilde{Y}^{T-1},T)$, $\tilde{X}=\tilde{X}_T, \tilde{Y}=\tilde{Y}_T$, and $\tilde{Z}=\tilde{Z}_T$, where T is uniform over $\{1,\dots,n\}$ and independent of all other random variables, and notice the Markov chains $U\to\tilde{X}\to\tilde{Y}$ and $V\to\tilde{Y}\to\tilde{Z}$. Then,

$$H(\tilde{M}_1)\geq nI(U;\tilde{X})+\log P_{X^n}(\mathcal{D}), \quad (29)$$
$$H(\tilde{M}_2)\geq nI(V;\tilde{Y})+\log P_{X^n}(\mathcal{D}), \quad (30)$$

Moreover, let $\eta>0$ be arbitrary. If

$$\Pr[\hat{\mathcal{H}}_Z=0|\mathcal{H}=0,X^n=x^n]\geq\eta, \quad \forall x^n\in\mathcal{D}, \quad (31)$$

then

$$-\frac{1}{n}\log\Pr[\hat{\mathcal{H}}_Z=0|\mathcal{H}=1,X^n\in\mathcal{D}]$$
$$\leq I(U;\tilde{Y})+I(V;\tilde{Z})+\varnothing_2(n), \quad (32)$$

and if

$$\Pr[\hat{\mathcal{H}}_Y=0|\mathcal{H}=0,X^n=x^n]\geq\eta, \quad \forall x^n\in\mathcal{D}, \quad (33)$$

then

$$-\frac{1}{n}\log\Pr[\hat{\mathcal{H}}_Y=0|\mathcal{H}=1,X^n\in\mathcal{D}]\leq I(U;\tilde{Y})+\varnothing_3(n), \quad (34)$$

where $\varnothing_2(n),\varnothing_3(n)$ are functions tending to 0 as $n\to\infty$.

Proof: See Appendix B. ■

We now prove the converse to Theorem 1. Fix a positive $\eta>0$, set $\mu_n=n^{-3}$, and define the sets

$$\mathcal{B}_1(\eta)\triangleq\{x^n\in\mathcal{T}_{\mu_n}^{(n)}(P_X)\colon \Pr[\hat{\mathcal{H}}_Y=0|X^n=x^n,\mathcal{H}=0]\geq\eta\}, \quad (35)$$
$$\mathcal{B}_2(\eta)\triangleq\{x^n\in\mathcal{T}_{\mu_n}^{(n)}(P_X)\colon \Pr[\hat{\mathcal{H}}_Z=0|X^n=x^n,\mathcal{H}=0]\geq\eta\}, \quad (36)$$
$$\mathcal{D}_2(\eta)\triangleq\mathcal{B}_1(\eta)\cap\mathcal{B}_2(\eta), \quad (37)$$
$$\mathcal{D}_1(\eta)\triangleq\mathcal{B}_1(\eta)\backslash\mathcal{D}_2(\eta), \quad (38)$$
$$\mathcal{D}_3(\eta)\triangleq\mathcal{B}_2(\eta)\backslash\mathcal{D}_2(\eta). \quad (39)$$

Further define:

$$\Delta_i\triangleq P_{X^n}(\mathcal{D}_i(\eta)), \quad i\in\{1,2,3\}, \quad (40)$$

and notice that

$$\Delta_1+\Delta_2=P_{X^n}(\mathcal{B}_1(\eta)) \quad \text{and} \quad \Delta_2+\Delta_3=P_{X^n}(\mathcal{B}_2(\eta)), \quad (41)$$

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

$$P_{X^n}(\mathcal{B}_j(\eta))\geq\frac{1-\epsilon_j-\eta}{1-\eta}-\frac{|\mathcal{X}|}{2\mu_n n}, \quad j\in\{1,2\}. \quad (42)$$

In the following, we assume that $\Delta_i>0$ for all $i\in\{1,2,3\}$. Note that degenerate cases can be treated similarly where if $\Delta_i=0$, we do not apply Lemma 1 to the corresponding set $\mathcal{D}_i$. In such case, (58) and (59) will still hold since the terms corresponding to the set $\mathcal{D}_i$ are always equal to 0. This further leads to no restrictions on the choices of U_i and V_i and therefore we can choose $U_i=\tilde{X}$ and $V_i=\tilde{Y}$, without loss of optimality. The latter result indicates that (47) and (48) will still hold without the need to use Lemma 1. We also note that by definition, it is impossible to have $\Delta_1=\Delta_2=0$ or $\Delta_2=\Delta_3=0$.

Now we proceed with our assumption and we apply Lemma 1 to each set $\mathcal{D}_i$, to obtain

$$H(\tilde{M}_{1,i})\geq nI(U_i;\tilde{X}_i)+\log P_{X^n}(\mathcal{D}_i) \qquad i\in\{1,2,3\}, \quad (43)$$
$$H(\tilde{M}_{2,i})\geq nI(V_i;\tilde{Y}_i)+\log P_{X^n}(\mathcal{D}_i) \qquad i\in\{2,3\}, \quad (44)$$

and

$$-\frac{1}{n}\log\Pr[\hat{\mathcal{H}}_Y=0|\mathcal{H}=1,X^n\in\mathcal{D}_i]$$
$$\leq I(U_i;\tilde{Y}_i)+\varnothing_{3,i}(n), \qquad i\in\{1,2\}, \quad (45)$$

$$-\frac{1}{n}\log\Pr[\hat{\mathcal{H}}_Z=0|\mathcal{H}=1,X^n\in\mathcal{D}_i]$$
$$\leq I(U_i;\tilde{Y}_i)+I(V_i;\tilde{Z}_i)+\varnothing_{2,i}(n), \quad i\in\{2,3\}, \quad (46)$$

where for each i the functions $\varnothing_{2,i}(n),\varnothing_{3,i}(n)\to 0$ as $n\to\infty$ 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 $\mathcal{D}_i$. By the total law of probability, we can then obtain

$$-\frac{1}{n}\log\beta_{1,n}\leq\min\{I(U_1;\tilde{Y});I(U_2;\tilde{Y})\}+\varnothing_3(n), \quad (47)$$
$$-\frac{1}{n}\log\beta_{2,n}\leq\min\{I(U_2;\tilde{Y})+I(V_2;\tilde{Z});I(U_3;\tilde{Y})+I(V_3;\tilde{Z})\}$$
$$+\varnothing_2(n), \quad (48)$$

where $\varnothing_2(n)$ and $\varnothing_3(n)$ are functions tending to 0 as $n\to\infty$.

Further define the following random variables for $j\in\{1,2\}$ and $i\in\{1,2,3\}$

$$\tilde{L}_{j,i}\triangleq\operatorname{len}(\tilde{\mathsf{M}}_{j,i}), \quad (49)$$

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

$$nR_j\geq\mathbb{E}[L_j] \quad (50)$$
$$\geq\sum_{i\in\{1,2,3\}}\mathbb{E}[\tilde{L}_{j,i}]\Delta_i. \quad (51)$$

Moreover,

$$H(\tilde{\mathsf{M}}_{j,i})=H(\tilde{\mathsf{M}}_{j,i},\tilde{L}_{j,i}) \quad (52)$$
$$=\sum_{l_i}\Pr[\tilde{L}_{j,i}=l_i]H(\tilde{\mathsf{M}}_{j,i}|\tilde{L}_{j,i}=l_i)+H(\tilde{L}_{j,i}) \quad (53)$$
$$\leq\sum_{l_i}\Pr[\tilde{L}_{j,i}=l_i]l_i+H(\tilde{L}_{j,i}) \quad (54)$$
$$=\mathbb{E}[\tilde{L}_{j,i}]+H(\tilde{L}_{j,i}), \quad (55)$$

which combined with (51) establishes

$$\sum_{i\in\{1,2,3\}}\Delta_iH(\tilde{\mathsf{M}}_{1,i})\leq\sum_{i\in\{1,2,3\}}\Delta_i\mathbb{E}[\tilde{L}_{1,i}]+\Delta_iH(\tilde{L}_{1,i}) \quad (56)$$
$$\leq nR_1+nR_1\sum_{i\in\{1,2,3\}}h_b\left(\mathbb{E}[\tilde{L}_{1,i}]^{-1}\right) \quad (57)$$
$$=nR_1\left(1+\sum_{i\in\{1,2,3\}}h_b\left(\frac{\Delta_i}{nR_1}\right)\right), \quad (58)$$

where (57) holds because a Geometric distribution maximizes entropy of random variables over the positive integers under an expectation constraint [15, Theorem 12.1.1].

In a similar way we obtain

$$\sum_{i\in\{2,3\}} \Delta_i H(\tilde{\mathsf{M}}_{2,i}) \leq nR_2 \left(1+\sum_{i\in\{2,3\}} h_b\left(\frac{\Delta_i}{nR_2}\right)\right). \quad (59)$$

The desired converse then follows by combining (58) and (59) with (43) and (44), noting (41) and (42), considering also (47) and (48), and letting $n\to\infty$ and $\eta\downarrow 0$. Notice in particular that since $\mathcal{D}_1,\mathcal{D}_2,\mathcal{D}_3$ are all subsets of $\mathcal{T}^{(n)}_{\mu_n}(P_X)$, for each $i\in\{1,2,3\}$ we have $|P_{\tilde{X}_i}-P_X|\leq\mu_n$, which tends to 0 as $n\to\infty$, and given $\tilde{X}_i$ the random variables $(\tilde{Y}_i,\tilde{Z}_i)$ are obtained by the conditional law $P_{YZ|X}$.

ACKNOWLEDGMENT

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

APPENDIX A
ANALYSIS OF THE CODING SCHEME IN SECTION IV FOR $\epsilon_1>\epsilon_2$

Let $\tilde{\mathcal{H}}_{Y,1}$ denote the hypothesis guessed by R_Y for the basic single-hop scheme with the first parameter choices $\mu, P_{U_1|X}$. Then let $\tilde{\mathcal{H}}_{Y,2}$ and $\tilde{\mathcal{H}}_{Z,2}$ denote the hypotheses guessed by R_Y and R_Z for the basic two-hop scheme with the different parameter choices $\mu, P_{U_2|X}, P_{V_2|Y}$. Similarly, let $\tilde{\mathcal{H}}_{Y,3}$ and $\tilde{\mathcal{H}}_{Z,3}$ be the hypotheses produced by R_Y and R_Z for the basic two-hop scheme with the parameter choices $\mu, P_{U_3|X}, P_{V_3|Y}$. We then obtain for the type-I error probabilities:

$$\begin{aligned}
\alpha_{1,n} &= \Pr[\hat{\mathcal{H}}_Y=1, X^n\in(\mathcal{S}_n\cup\mathcal{D}_3)|\mathcal{H}=0]\\
&\quad+\Pr[\hat{\mathcal{H}}_Y=1, X^n\in\mathcal{D}_1|\mathcal{H}=0]\\
&\quad+\Pr[\hat{\mathcal{H}}_Y=1, X^n\in\mathcal{D}_2|\mathcal{H}=0] &(60)\\
&=\Pr[X^n\in(\mathcal{S}_n\cup\mathcal{D}_3)|\mathcal{H}=0]\\
&\quad+\Pr[\tilde{\mathcal{H}}_{Y,1}=1, X^n\in\mathcal{D}_1|\mathcal{H}=0]\\
&\quad+\Pr[\tilde{\mathcal{H}}_{Y,2}=1, X^n\in\mathcal{D}_2|\mathcal{H}=0] &(61)\\
&\leq \epsilon_1-\mu+\Pr[\tilde{\mathcal{H}}_{Y,1}=1|\mathcal{H}=0]\\
&\quad+\Pr[\tilde{\mathcal{H}}_{Y,2}=1|\mathcal{H}=0] &(62)
\end{aligned}$$

and

$$\begin{aligned}
\alpha_{2,n} &= \Pr[\hat{\mathcal{H}}_Z=1, X^n\in(\mathcal{S}_n\cup\mathcal{D}_1)|\mathcal{H}=0]\\
&\quad+\Pr[\hat{\mathcal{H}}_Z=1, X^n\in\mathcal{D}_2|\mathcal{H}=0]\\
&\quad+\Pr[\hat{\mathcal{H}}_Z=1, X^n\in\mathcal{D}_3|\mathcal{H}=0] &(63)\\
&=\Pr[X^n\in(\mathcal{S}_n\cup\mathcal{D}_1)|\mathcal{H}=0]\\
&\quad+\Pr[\tilde{\mathcal{H}}_{Z,2}=1, X^n\in\mathcal{D}_2|\mathcal{H}=0]\\
&\quad+\Pr[\tilde{\mathcal{H}}_{Z,3}=1, X^n\in\mathcal{D}_3|\mathcal{H}=0] &(64)\\
&\leq \epsilon_2-\mu+\Pr[\tilde{\mathcal{H}}_{Z,2}=1|\mathcal{H}=0]\\
&\quad+\Pr[\tilde{\mathcal{H}}_{Z,3}=1|\mathcal{H}=0]. &(65)
\end{aligned}$$

Since by [8], [11], $\forall i\in\{1,2,3\}$ and $\forall j\in\{2,3\}$, $\Pr[\tilde{\mathcal{H}}_{Y,i}=1|\mathcal{H}=0]$, and $\Pr[\tilde{\mathcal{H}}_{Z,j}=1|\mathcal{H}=0]\downarrow 0$ as $n\to\infty$, we conclude that for the general scheme in Section IV $\overline{\lim}_{n\to\infty}\alpha_{1,n}\leq\epsilon_1$ and $\overline{\lim}_{n\to\infty}\alpha_{2,n}\leq\epsilon_2$.

For the type-II error probabilities we obtain

$$\begin{aligned}
\beta_{1,n} &= \Pr[\tilde{\mathcal{H}}_{Y,1}=0, X^n\in\mathcal{D}_1|\mathcal{H}=1]\\
&\quad+\Pr[\tilde{\mathcal{H}}_{Y,2}=0, X^n\in\mathcal{D}_2|\mathcal{H}=1] &(66)\\
&\leq \Pr[\tilde{\mathcal{H}}_{Y,1}=0|\mathcal{H}=1]+\Pr[\tilde{\mathcal{H}}_{Y,2}=0|\mathcal{H}=1] &(67)\\
&\leq 2^{-n(I(U_1;Y)+\delta(\mu))}+2^{-n(I(U_2;Y)+\delta(\mu))}, &(68)
\end{aligned}$$

and

$$\begin{aligned}
\beta_{2,n} &= \Pr[\tilde{\mathcal{H}}_{Z,2}=0, X^n\in\mathcal{D}_2|\mathcal{H}=1]\\
&\quad+\Pr[\tilde{\mathcal{H}}_{Z,3}=0, X^n\in\mathcal{D}_3|\mathcal{H}=1] &(69)\\
&\leq \Pr[\tilde{\mathcal{H}}_{Z,2}=0|\mathcal{H}=1]+\Pr[\tilde{\mathcal{H}}_{Z,3}=0|\mathcal{H}=1] &(70)\\
&\leq 2^{-n(I(U_2;Y)+I(V_2;Z)+\delta(\mu))}\\
&\quad+2^{-n(I(U_3;Y)+I(V_3;Z)+\delta(\mu))}. &(71)
\end{aligned}$$

where (71) and (68) are proved in [11], and $\delta(\mu)\downarrow 0$ as $\mu\downarrow 0$.

The expected message lengths of the described scheme are:

$$\begin{aligned}
\mathbb{E}[\mathrm{len}(\mathsf{M}_1)] &= (\sigma+\epsilon_1+\epsilon_2-1-\mu)\cdot 2\\
&\quad+(1-\epsilon_1-\sigma)\cdot(n(I(U_1;X)+\mu)+2)\\
&\quad+(\sigma+\mu)\cdot(n(I(U_2;X)+\mu)+2) &(72)\\
&\quad+(1-\epsilon_2-\sigma)\cdot(n(I(U_3;X)+\mu)+2) &(73)\\
&\leq nR_1 &(74)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\mathrm{len}(\mathsf{M}_2)] &= (\epsilon_2-\mu)\cdot 2\\
&\quad+(\sigma+\mu)\cdot(n(I(V_2;Y)+\mu)+2)\\
&\quad+(1-\epsilon_2-\sigma)\cdot(n(I(V_3;Y)+\mu)+2) &(75)\\
&\leq nR_2. &(76)
\end{aligned}$$

Here, (74) and (76) hold for sufficiently large values of n such that $(2-\epsilon_1-\epsilon_2-\sigma+\mu)n\mu\geq 2$ and $(1-\epsilon_2+\mu)n\mu\geq 2$.

Letting first $n\to\infty$ and then $\mu\downarrow 0$, establishes the desired achievability result in (16).

APPENDIX B
PROOF OF LEMMA 1

Note first that by (28):

$$D(P_{\tilde{X}^n}||P_X^n)=D(P_{\tilde{X}^n\tilde{Y}^n}||P_{XY}^n)\leq\log\Delta_n^{-1}, \quad (77)$$

where we defined $\Delta_n:=P_{X^n}(\mathcal{D})$.

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

$$\begin{aligned}
&H(\tilde{\mathsf{M}}_1)\\
&\geq I(\tilde{\mathsf{M}}_1;\tilde{X}^n)+D(P_{\tilde{X}^n}||P_X^n)+\log\Delta_n &(78)\\
&=H(\tilde{X}^n)+D(P_{\tilde{X}^n}||P_X^n)-H(\tilde{X}^n|\tilde{\mathsf{M}}_1)+\log\Delta_n &(79)\\
&\geq n[H(\tilde{X}_T)+D(P_{\tilde{X}_T}||P_X)]-\sum_{t=1}^{n}H(\tilde{X}_t|\tilde{U}_t)+\log\Delta_n &(80)\\
&=n[H(\tilde{X}_T)+D(P_{\tilde{X}_T}||P_X)-H(\tilde{X}_T|\tilde{U}_T,T)]+\log\Delta_n &(81)\\
&\geq n[H(\tilde{X}_T)-H(\tilde{X}_T|\tilde{U}_T,T)]+\log\Delta_n &(82)\\
&=n\left[I(\tilde{X};U)+\frac{1}{n}\log\Delta_n\right]. &(83)
\end{aligned}$$

Here, (78) holds by (77); (80) holds by the super-additivity property in [16, Proposition 1], by the chain rule, and by the definition of $\tilde{U}_t$; (81) by defining T uniform over $\{1,\dots,n\}$ independent of all other random variables; and (83) by definitions of $U,\tilde{X}$.

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

$$H(\tilde{\mathsf{M}}_2) \geq n \left[I(\tilde{Y}; V) + \frac{1}{n} \log \Delta_n \right]. \tag{84}$$

We next upper bound the error exponent at the receiver. Define for each x^n, the set

$$\mathcal{A}_{YZ,n}(x^n) \triangleq \{(y^n, z^n) \colon g_2(\phi_2(\phi_1(x^n), y^n), z^n) = 0\}, \tag{85}$$

and its Hamming neighborhood:

$$\hat{\mathcal{A}}^{\ell_n}_{YZ,n}(x^n) \triangleq \{(\tilde{y}^n, \tilde{z}^n) : \exists\, (y^n, z^n) \in \mathcal{A}_{YZ,n}(x^n) \text{ s.t. } d_H((y^n, z^n), (\tilde{y}^n, \tilde{z}^n)) \leq \ell_n\} \tag{86}$$

for some real number ℓ_n satisfying $\lim_{n\to\infty} \ell_n/n = 0$ and $\lim_{n\to\infty} \ell_n/\sqrt{n} = \infty$. Since by condition (31),

$$P_{\tilde{Y}^n \tilde{Z}^n | \tilde{X}^n}(\mathcal{A}_{YZ,n}(x^n)|x^n) \geq \eta, \quad \forall x^n \in \mathcal{D}, \tag{87}$$

the blowing-up lemma [17] yields

$$P_{\tilde{Y}^n \tilde{Z}^n | \tilde{X}^n}(\hat{\mathcal{A}}^{\ell_n}_{YZ,n}(x^n)|x^n) \geq 1 - \zeta_n, \quad \forall x^n \in \mathcal{D}, \tag{88}$$

for a real number $\zeta_n > 0$ such that $\lim\limits_{n\to\infty} \zeta_n = 0$.
Further define:

$$\hat{\mathcal{A}}^{\ell_n}_{M_2Z,n}(x^n, y^n) \triangleq \{(m_2, z^n) : (y^n, z^n) \in \hat{\mathcal{A}}^{\ell_n}_{YZ,n}(x^n), \; \mathsf{m}_2 = \phi_2(\phi_1(x^n), y^n)\}, \tag{89}$$

and

$$\hat{\mathcal{A}}^{\ell_n}_{M_2Z,n} \triangleq \bigcup_{x^n, y^n} \hat{\mathcal{A}}^{\ell_n}_{M_2Z,n}(x^n, y^n), \tag{90}$$

Then:

$$\begin{aligned} P_{\tilde{\mathsf{M}}_2 \tilde{Z}^n}(\hat{\mathcal{A}}^{\ell_n}_{M_2Z,n}) &= \sum_{\substack{x^n \in \mathcal{D}, \\ (y^n, z^n) \in \hat{\mathcal{A}}^{\ell_n}_{YZ,n}(x^n), \\ \mathsf{m}_2 = \phi_2(\phi_1(x^n), y^n)}} P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{\mathsf{M}}_2}(x^n, y^n, z^n, \mathsf{m}_2) && (91) \\ &\geq (1 - \zeta_n). && (92) \end{aligned}$$

Defining

$$Q_{\tilde{\mathsf{M}}_2}(\mathsf{m}_2) \triangleq \sum_{y^n, \mathsf{m}_1} P_{\tilde{\mathsf{M}}_1}(\mathsf{m}_1) P_{\tilde{Y}^n}(y^n) \cdot \mathbb{1}\{\phi_2(\mathsf{m}_1, y^n) = \mathsf{m}_2\}, \tag{93}$$

we can write

$$\begin{aligned} &Q_{\tilde{\mathsf{M}}_2} P_{\tilde{Z}^n}\left(\hat{\mathcal{A}}^{\ell_n}_{MZ,n}\right) \\ &\quad \leq \underbrace{Q_{\tilde{\mathsf{M}}_2} P_{\tilde{Z}^n}\left(\mathcal{A}_{MZ,n}\right)}_{\Pr[\hat{\mathcal{H}}_Z = 0 | \mathcal{H} = 1, X^n \in \mathcal{D}]} e^{n h_b(\ell_n/n)} |\mathcal{Y}|^{\ell_n} |\mathcal{Z}|^{\ell_n} k_n^{\ell_n} && (94) \end{aligned}$$

where $k_n \triangleq \min\limits_{\substack{y, y', z, z': \\ P_Y(y'), P_Z(z') > 0}} \frac{P_Y(y) P_Z(z)}{P_Y(y') P_Z(z')}$. Here, (94) holds by [14, Proof of Lemma 5.1].

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

$$\begin{aligned} &-\frac{1}{n} \log \Pr[\hat{\mathcal{H}}_Z = 0 | \mathcal{H} = 1, X^n \in \mathcal{D}] \\ &\quad \leq \frac{1}{n(1 - \zeta_n)} D(P_{\tilde{\mathsf{M}}_2 \tilde{Z}^n} || Q_{\tilde{\mathsf{M}}_2} P_{\tilde{Z}^n}) + \delta_n && (95) \end{aligned}$$

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

$$\begin{aligned} &D(P_{\tilde{\mathsf{M}}_2 \tilde{Z}^n} || Q_{\tilde{\mathsf{M}}_2} P_{\tilde{Z}^n}) \\ &\quad = I(\tilde{\mathsf{M}}_2; \tilde{Z}^n) + D(P_{\tilde{\mathsf{M}}_2} || Q_{\tilde{\mathsf{M}}_2}) && (96) \\ &\quad \leq I(\tilde{\mathsf{M}}_2; \tilde{Z}^n) + D(P_{\tilde{Y}^n \tilde{\mathsf{M}}_1} || P_{\tilde{Y}^n} P_{\tilde{\mathsf{M}}_1}) && (97) \\ &\quad = I(\tilde{\mathsf{M}}_2; \tilde{Z}^n) + I(\tilde{\mathsf{M}}_1; \tilde{Y}^n) && (98) \\ &\quad = \sum_{t=1}^n I(\tilde{\mathsf{M}}_2; \tilde{Z}_t | \tilde{Z}^{t-1}) + I(\tilde{\mathsf{M}}_1; \tilde{Y}_t | \tilde{Y}^{t-1}) && (99) \\ &\quad \leq \sum_{t=1}^n I(\tilde{\mathsf{M}}_2 \tilde{Y}^{t-1} \tilde{Z}^{t-1}; \tilde{Z}_t) + I(\tilde{\mathsf{M}}_1 \tilde{X}^{t-1} \tilde{Y}^{t-1}; \tilde{Y}_t) && (100) \\ &\quad = \sum_{t=1}^n I(\tilde{\mathsf{M}}_2 \tilde{Y}^{t-1}; \tilde{Z}_t) + I(\tilde{\mathsf{M}}_1 \tilde{X}^{t-1}; \tilde{Y}_t) && (101) \\ &\quad = \sum_{t=1}^n I(\tilde{V}_t; \tilde{Z}_t) + I(\tilde{U}_t; \tilde{Y}_t) && (102) \\ &\quad = n[I(\tilde{V}_T; \tilde{Z}_T | T) + I(\tilde{U}_T; \tilde{Y}_T | T)] && (103) \\ &\quad \leq n[I(\tilde{V}_T T; \tilde{Z}_T) + I(\tilde{U}_T T; \tilde{Y}_T)] && (104) \\ &\quad = n[I(V; \tilde{Z}) + I(U; \tilde{Y})]. && (105) \end{aligned}$$

Here (97) is obtained by data processing inequality for relative entropy; (99) by the chain rule; (101) by the Markov chains $\tilde{Z}^{t-1} \to \tilde{Y}^{t-1} \to \tilde{Z}_t$ and $\tilde{Y}^{t-1} \to \tilde{X}^{t-1} \to \tilde{Y}_t$; and (102)–(105) 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 $\Pr[\hat{\mathcal{H}}_Y = 0 | \mathcal{H} = 1, X^n \in \mathcal{D}]$ if (33) is satisfied. In this latter proof it suffices to blow up the set of y^n sequences.

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] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed.* Wiley, 2006.
[16] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 689–703, 2019.
[17] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.