Multi-Hop Network with Multiple Decision Centers under Expected-Rate Constraints

Mustapha Hamad, Michèle Wigger, Mireille Sarkiss

***Abstract*—We consider a multi-hop distributed hypothesis testing problem with multiple decision centers (DCs) for testing against independence and where the observations obey some Markov chain. For this system, we characterize the fundamental type-II error exponents region, i.e., the type-II error exponents that the various DCs can achieve simultaneously, under expected rate-constraints. Our results show that this fundamental exponents region is boosted compared to the region under maximum-rate constraints, and that it depends on the permissible type-I error probabilities. When all DCs have equal permissible type-I error probabilities, the exponents region is rectangular and all DCs can simultaneously achieve their optimal type-II error exponents. When the DCs have different permissible type-I error probabilities, a tradeoff between the type-II error exponents at the different DCs arises. New achievability and converse proofs are presented. For the achievability, a new multiplexing and rate-sharing strategy is proposed. The converse proof is based on applying different change of measure arguments in parallel and on proving asymptotic Markov chains. For the special cases $K \in \{2, 3\}$, and for arbitrary $K \geq 2$ when all permissible type-I error probabilities at the various DCs are equal, we provide simplified expressions for the exponents region; a similar simplification is conjectured for the general case.**

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

I. Introduction

Future wireless systems are driven by the exponential growth of IoT networks and applications with various requirements in terms of rate, reliability, and energy consumption. In applications such as health monitoring, security alerting or automotive car control, the sensing and decision systems aim at accurately detecting hazardous events or anomalies at the decision centers (DCs) by collecting data about the measurements at the various sensors. The different events can be considered as different hypotheses and are assumed to determine the joint probability distribution underlying the data observed at all the terminals. Our focus will be on binary hypothesis testing, i.e., situations with only two possible events, with one of the two events corresponding to the normal situation, the so called *null hypothesis* and the other to an alert situation the so called *alternative hypothesis*. There are two types of errors to distinguish here: *type-I error* and *type-II error*. Type-I error corresponds to a false alarm where the decision center decides on the alternative hypothesis when the true hypothesis is the null hypothesis. Type-II error corresponds to a missed detection where the decision center decides on the null hypothesis when the true one is the alternative hypothesis. Since our interest is in alert systems where a missed detection is more critical, we aim at maximizing the exponential decay of the type-II error probability (called *error exponent*) while only requiring the type-I error probability to stay below a given threshold.

Most of the information theoretic works studied the distributed binary hypothesis testing problem with a single sensor that communicates with a single distant DC over a noise-free link with a constraint on the maximum allowed communication rate [3]–[10]. These results were also extended to setups with noisy communication links [11]–[13], to setups with privacy and secrecy constraints [14], [15], and to more complicated networks with either interactive communication [16]–[19], multiple sensors [8], [20], [21], multiple decision centers [22]–[26], or both of them [27]–[29]. The works most closely related to this paper are [28] and [29] which considered a multi-hop setup with K sensors and K DCs. Multi-hop setups are motivated by the stringent energy constraints of IoT devices requiring short-range communication only between neighbouring sensors.

Specifically, [28] characterized a set of type-II error exponent tuples that are simultaneously achievable at the various DCs in a multi-hop network with K sensors and K DCs. For the special case of *testing against independence* and when the type-I error probabilities at all the DCs are required to vanish asymptotically, this set of exponents coincides with the fundamental exponents region, which means that in this special case no other exponent tuples are achievable. Testing against independence refers to a hypothesis test where under the alternative hypothesis the observations at the various terminals follow the product of the marginal distributions that they experience under the null hypothesis. The result in [28] further required that the joint distribution of the various observations under the null hypothesis satisfies certain Markov chains from one relay to the other. Interestingly, in this case, the set of exponent tuples that are simultaneously achievable at the K decision centers is a K-dimensional hypercube, implying that no tradeoff between the exponents arises and each DC can achieve the optimal exponent as if it was the only DC in the system. When $K = 2$, [29] proved the strong converse

Part of this material was presented at IEEE Inf. Theory Workshop (ITW) 2021 [1] and at IEEE Globecom Conference 2021 [2].

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

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

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

The works of M. Hamad and M. Wigger were supported by the European Research Council (ERC) under the European Union's Horizon 2020 programme, grant agreement number 715111.

result that the optimal exponent region does not depend on the permissible type-I error probabilities. (The work in [29] presents different expressions for the sets of achievable type-II error exponents depending on whether the sum of the two admissible type-I error probabilities exceeds 1 or not. It can however be shown that the two expressions coincide, as we argue in Remark 1.)

Above works all focused on maximum rate-constraints where the length of *any* message sent over the communication link is limited. In this paper, we consider *expected rate-constraints* as in [10], [30]–[32], where the *expected* length of the message sent over the communication link is constrained. Most closely related are the works in [30], [33] which showed that under an expected rate-constraint R, the optimal type-II error exponent for testing against independence in the single-sensor and single-DC setup coincides with the optimal type-II error exponent under a maximum-rate constraint $R/(1-\epsilon)$, for ϵ denoting the permissible type-I error constraint. In other words, the relaxed expected-rate constraint seems to allow to boost the rate by a factor $(1-\epsilon)^{-1}$ compared to the same setup under a maximum-rate constraint.

We show that the same conclusion holds for the K-hop network with K decision centers considered in [28] when all DCs obey the same type-I error constraint ϵ. In this case, the fundamental exponents region is a K-dimensional hypercube where all DCs can simultaneously achieve their optimal type-II error exponents as if they were the only DC in the system, and this exponent coincides with the exponent under a maximum-rate constraint but where the rates of *all links in the system* are boosted by a factor $(1-\epsilon)^{-1}$. In contrast, when the various DCs have different type-I error probability thresholds, a tradeoff arises between the type-II error exponents that are simultaneously achievable at the different DCs. This tradeoff, which depends on the type-I error thresholds at the different DCs, is the first of its kind and we exactly characterize it for the studied multi-hop setup. We notice hence that under expected rate-constraints a strong converse does not hold, since the optimal type-II error exponents depend on the admissible type-I error probabilities. We show this for arbitrary $K \geq 2$ and arbitrary admissible type-I error probabilities $\epsilon_1, \ldots, \epsilon_K$ by fully characterizing the fundamental type-II error exponents region.

To prove our achievability results under expected-rate constraints, we propose a new multiplexing and rate-sharing strategy that generalizes the degenerate multiplexing scheme in [30]. Specifically, we multiplex different coding schemes of different sets of rates on the various links and with different probabilities, where each multiplexed subscheme is an optimal coding and testing scheme when the maximum rates are limited by the chosen rate-tuple. For $K = 2$ and $K = 3$, we explicitly characterize the multiplexing probabilities in function of the type-I error probability thresholds at the various DCs and we show that one can restrict to only $K + 1$ subschemes, instead of 2^K. We conjecture that a similar simplification holds for arbitrary $K \geq 2$.

Our converse proofs apply several instances of the change of measure arguments in [34]–[36] in parallel, where we also restrict to jointly typical source sequences as in [35]. In contrast to the related strong converse proofs in [29], [34], no variational characterizations, or hypercontractivity arguments [37] are required to prove our desired results. Instead, we rely on arguments showing that certain Markov chains hold in an asymptotic regime of infinite blocklengths. Notice that our method to circumvent variational characterizations, or hypercontractivity, or blowing-up arguments [38], seems to extend also to other converse proofs, see for example the simplified proof of the well-known strong converses for lossless and lossy compression with side-information at the decoder [39], [40] presented in [41].

We summarize our main contributions for the $K \geq 2$-hop network with K DCs that test against independence and when the observations at the terminals obey a specific Markov chain:

- We provide an exact characterization of the general fundamental exponents region under expected-rate constraints. This result shows rate-boosts on all the links in the system, and illustrates a tradeoff between the exponents at all DCs with different type-I error thresholds.
- To prove achievability, we propose a new coding scheme based on *multiplexing and rate-sharing strategy*.
- Converses are proved by several parallel change of measure arguments and by showing certain Markov chains in the asymptotic regime of infinite blocklengths.
- We prove that our results simplify for the special cases of $K = 2$ or $K = 3$ hops or when all K DCs have same admissible type-I error probabilities ϵ. For equal type-I error probabilities, the optimal scheme multiplexes only two subschemes instead of 2^K. For general type-I error probabilities and when there are only $K = 2$ or $K = 3$ hops, then at most $K + 1$ subschemes need to be multiplexed. Multiplexing probabilities in these special cases can directly be obtained from the permissible type-I error probabilities at the various DCs. Similar simplifications are conjectured for arbitrary $K \geq 2$ and arbitrary type-I error probabilities.

Paper organization: The remainder of this paper is divided into two main parts, one focusing on the two-hop network (Sections II–VI) and one considering the general K-hop network (Sections VII–VIII). For the first part, Section II describes the two-hop system model, and Section III presents the related previous results under maximum-rate constraints. Section IV explains and analyses our proposed optimal coding schemes for the setup under expected-rate constraints. Section V contains our main results, discussion, and numerical analysis for the two-hop network. In Section VI, we provide our converse proof. For the second part, Section VII introduces the system model for K hops, presents the related previous results on maximum-rate constraints. It also describes our new optimal coding scheme and the fundamental exponents region under expected-rate constraints, and simplifications on them. The converse for K-Hops is presented in Section VIII.

Notation: We follow the notation in [30]. 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 $\mathrm{bin}(m)$ 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. The set of all bit-strings is denoted by $\{0,1\}^\star$.

Random variables are typically written with upper case symbols, e.g., A, and realizations in lower case, e.g., a. Sets show in calligraphic symbols, for example $\mathcal{A}$, and $\mathcal{A}^n$ denotes the n-fold Cartesian product of $\mathcal{A}$, for any positive integer n. We further abbreviate the random and deterministic n-tuples $(A_1,\ldots,A_n)$ and $(a_1,\ldots,a_n)$ by A^n and a^n. For given integer K, we use $\mathcal{P}(K)$ to denote the power set of $\{1,\ldots,K\}$, i.e., the set of all possible subsets of $\{1,\ldots,K\}$ including the set $\{1,\ldots,K\}$ itself, but excluding the emptyset $\emptyset$. In addition, $\mathcal{T}_\mu^{(n)}(P_{XY})$ denotes the strongly typical set as defined in [42, Definition 2.8]. That means, for any small positive number μ a pair (x^n,y^n) lies in $\mathcal{T}_\mu^{(n)}(P_{XY})$ if, and only if,

$$\left|\frac{|\{t\colon (x_t,y_t)\}|}{n} - P_{XY}(a,b)\right| \leq \mu, \quad \forall (a,b)\in\mathcal{X}\times\mathcal{Y}, \quad (1)$$

and for all pairs (a,b) of 0 probability $(x_t,y_t)\neq(a,b)$ for all $t\in\{1,\ldots,n\}$.

Throughout this manuscript, $h_b(\cdot)$ denotes the binary entropy function, and $D(\cdot\|\cdot)$ the Kullback-Leibler divergence between two probability mass functions on the same alphabet. Entropy, conditional entropy, and mutual information of random variables are denoted by $H(\cdot)$, $H(\cdot|\cdot)$, and $I(\cdot;\cdot)$. When the probability mass functions of the involved random variables are not clear from the context, we add them as a subscript and write for example $H_P(\cdot)$, $H_P(\cdot|\cdot)$, and $I_P(\cdot;\cdot)$. We use the symbols $\overline{\lim}$ and $\underline{\lim}$ to denote the limsup and the liminf of sequences. $\mathbb{1}\{\cdot\}$ denotes the indicator function and for any real value x and we write $[x]^+$ for $\max\{0,x\}$.

Finally, we abbreviate the terms independent and identically distributed by *i.i.d.* and probability mass function by *pmf*.

II. The Two-Hop System Model

Consider the distributed hypothesis testing problem in Figure 1 with a transmitter T_0, a relay R_1 and a receiver R_2 observing sequences Y_0^n, Y_1^n and Y_2^n respectively, forming the Markov chain

$$Y_0^n \to Y_1^n \to Y_2^n \quad (2)$$

In the special case of testing against independence, i.e., depending on the binary hypothesis $\mathcal{H}\in\{0,1\}$, the tuple (Y_0^n,Y_1^n,Y_2^n) is distributed as:

$$\text{under } \mathcal{H}=0: (Y_0^n,Y_1^n,Y_2^n) \text{ i.i.d.} \sim P_{Y_0Y_1}\cdot P_{Y_2|Y_1}; \quad (3a)$$

$$\text{under } \mathcal{H}=1: (Y_0^n,Y_1^n,Y_2^n) \text{ i.i.d.} \sim P_{Y_0}\cdot P_{Y_1}\cdot P_{Y_2} \quad (3b)$$

for given pmfs $P_{Y_0Y_1}$ and $P_{Y_2|Y_1}$ and where P_{Y_0}, P_{Y_1}, and P_{Y_2} denote the marginals of the joint pmf $P_{Y_0Y_1Y_2} := P_{Y_0Y_1}P_{Y_2|Y_1}$.

In this two-hop setup, the transmitter T_0 observes the source sequence Y_0^n and sends its bit-string message $\mathsf{M}_1=\phi_0^{(n)}(Y_0^n)$ to R_1, where the encoding function is of the form $\phi_0^{(n)}:\mathcal{Y}_0^n\to\{0,1\}^\star$ and satisfies the *expected* rate constraint

$$\mathbb{E}\left[\mathrm{len}\left(\mathsf{M}_1\right)\right] \leq nR_1. \quad (4)$$

The relay R_1 observes the source sequence Y_1^n and with the message M_1 received from T_0, it produces a guess $\hat{\mathcal{H}}_1$ of the

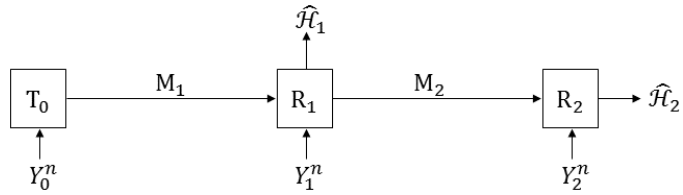

Fig. 1: Cascaded two-hop setup with two decision centers.

hypothesis $\mathcal{H}$ using a decision function $g_1^{(n)}:\mathcal{Y}_1^n\times\{0,1\}^\star\to\{0,1\}$:

$$\hat{\mathcal{H}}_1 = g_1^{(n)}\left(Y_1^n,\mathsf{M}_1\right) \in\{0,1\}. \quad (5)$$

Relay R_1 also computes a bit-string message $\mathsf{M}_2 = \phi_1^{(n)}\left(Y_1^n,\mathsf{M}_1\right)$ using some encoding function $\phi_1^{(n)}:\mathcal{Y}_1^n\times\{0,1\}^\star\to\{0,1\}^\star$ that satisfies the expected-rate constraint

$$\mathbb{E}\left[\mathrm{len}\left(\mathsf{M}_2\right)\right] \leq nR_2. \quad (6)$$

Then it sends M_2 to the receiver R_2, which guesses hypothesis $\mathcal{H}$ using its observation Y_2^n and the received message M_2, i.e., using a decision function $g_2^{(n)}:\mathcal{Y}_2^n\times\{0,1\}^\star\to\{0,1\}$, it produces the guess:

$$\hat{\mathcal{H}}_2 = g_2^{(n)}\left(Y_2^n,\mathsf{M}_2\right) \in\{0,1\}. \quad (7)$$

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

$$\alpha_{1,n} \triangleq \Pr[\hat{\mathcal{H}}_1=1|\mathcal{H}=0] \quad (8)$$

$$\alpha_{2,n} \triangleq \Pr[\hat{\mathcal{H}}_2=1|\mathcal{H}=0] \quad (9)$$

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

$$\beta_{1,n} \triangleq \Pr[\hat{\mathcal{H}}_1=0|\mathcal{H}=1] \quad (10)$$

$$\beta_{2,n} \triangleq \Pr[\hat{\mathcal{H}}_2=0|\mathcal{H}=1] \quad (11)$$

decay to 0 with largest possible exponential decay.

Definition 1: Fix maximum type-I error probabilities $(\epsilon_1,\epsilon_2)\in[0,1)^2$ and nonnegative rates (R_1,R_2). The exponent pair (θ_1,θ_2) is called $(R_1,R_2,\epsilon_1,\epsilon_2)$-*achievable* if there exists a sequence of encoding and decision functions $\{\phi_0^{(n)},\phi_1^{(n)},g_1^{(n)},g_2^{(n)}\}_{n\geq1}$ satisfying $\forall i\in\{1,2\}$:

$$\mathbb{E}[\mathrm{len}(\mathsf{M}_i)] \leq nR_i, \quad (12a)$$

$$\overline{\lim_{n\to\infty}}\ \alpha_{i,n} \leq \epsilon_i, \quad (12b)$$

$$\underline{\lim_{n\to\infty}}\ \frac{1}{n}\log\frac{1}{\beta_{i,n}} \geq \theta_i. \quad (12c)$$

Definition 2: The closure of the set of all (ϵ_1,ϵ_2)-achievable exponent pairs (θ_1,θ_2) is called the fundamental (ϵ_1,ϵ_2)-*exponents region* and is denoted $\mathcal{E}^*(R_1,R_2,\epsilon_1,\epsilon_2)$.

III. Previous Results on Maximum-Rate Constraints for Two Hops

A. The Setup

The multi-hop hypothesis testing setup of Figure 1 and Equations (3) was also considered in [27] and [29], but under *maximum-rate constraints*:

$$\text{len}(\mathsf{M}_i) \leq nR_i, \qquad i \in \{1,2\}, \tag{13}$$

instead of the *expected-rate constraints* (12a). The fundamental exponents region $\mathcal{E}^*_{\text{max}}(R_1, R_2, \epsilon_1, \epsilon_2)$ for this maximum-rate setup is defined analogously to Definition (2), but with (12a) replaced by (13).

In the following subsection, we report the fundamental exponents region $\mathcal{E}^*_{\text{max}}(R_1, R_2, \epsilon_1, \epsilon_2)$ derived in [29].

B. The Exponents Region

Define the two functions

$$\eta_1(R_1) := \max_{\substack{P_{U_1|Y_0}: \\ R_1 \geq I(U_1;Y_0)}} I(U_1;Y_1) \tag{14}$$

$$\eta_2(R_2) := \max_{\substack{P_{U_2|Y_1}: \\ R_2 \geq I(U_2;Y_1)}} I(U_2;Y_2), \tag{15}$$

where the mutual information quantities are calculated with respect to the joint pmfs $P_{U_1Y_0Y_1} := P_{U_1|Y_0}P_{Y_0Y_1}$ and $P_{U_2Y_1Y_2} := P_{U_2|Y_1}P_{Y_1Y_2}$, respectively. As stated in [3], in the above maximization problems it suffices to consider auxiliary random variables U_1 and U_2 over alphabets of sizes $|\mathcal{Y}_0|+1$ and $|\mathcal{Y}_1|+1$.

Lemma 1: The functions η_1 and η_2 are continuous, concave and monotonically non-decreasing on their entire domain $\mathbb{R}_0^+$.

Proof: Appendix A proves the desired properties for η_1. The proof for η_2 is analogous and omitted. ■

The exponents region of the two-hop setup under maximum-rate constraints was determined in [29] for the case $\epsilon_1+\epsilon_2 \neq 1$ and in [41] for the case $\epsilon_1+\epsilon_2 = 1$. Achievability was first established in [27].

Theorem 1 ([29], [41]): Fix $(\epsilon_1, \epsilon_2) \in [0,1)^2$. The fundamental exponents region under the maximum-rate constraints (13) is:

$$\begin{aligned}&\mathcal{E}^*_{\text{max}}(R_1, R_2, \epsilon_1, \epsilon_2) \\ &\quad = \{(\theta_1, \theta_2) \colon \theta_1 \leq \eta_1(R_1),\ \theta_2 \leq \eta_1(R_1) + \eta_2(R_2)\}.\end{aligned} \tag{16}$$

We notice that the fundamental exponents region does not depend on the permissible type-I error probabilities ϵ_1 and ϵ_2. We will therefore abbreviate $\mathcal{E}^*_{\text{max}}(R_1, R_2, \epsilon_1, \epsilon_2)$ by $\mathcal{E}^*_{\text{max}}(R_1, R_2)$.

Remark 1: When $\epsilon_1+\epsilon_2 > 1$, the work in [29] characterizes $\mathcal{E}^*_{\text{max}}(R_1, R_2)$ in form of an optimization problem over three auxiliary random variables U_1, U_2, V, see [29, Eq. (33)]. It can however be verified that without loss of optimality this optimization can be restricted to auxiliaries $U_1 = U_2$, in which case the characterisation in [29, Eq. (33)] reduces to the expression in (16).

Notice that $\eta_1(R_1)$ determines the optimal exponent in a point-to-point system where R_2 is not present, and $\eta_2(R_2)$ determines the optimal exponent in a point-to-point system where T_0 is not present [3]. In the studied two-hop setup, R_2 thus accumulates the optimal exponents achieved over the two links. Since the exponents region is a rectangle, each of the two decision centers, R_1 and R_2, can simultaneously achieve their optimal exponents, no tradeoff occurrs between the two exponents. We shall see that this is not always the case under expected-rate constraints.

IV. Optimal Two-Hop Coding Scheme under Expected-Rate Constraints

The optimal coding scheme under expected-rate constraints depends on whether $\epsilon_1 = \epsilon_2$, $\epsilon_1 < \epsilon_2$, or $\epsilon_1 > \epsilon_2$. The general idea of all the three schemes is that the three terminals T_0, R_1, R_2 multiplex two or three different subschemes, and the choice of which subscheme to use depends on the transmitter T_0's observations y_0^n. To inform all terminals about the choice of the subscheme, T_0 adds one or two flag bits to its message, which the relay R_1 forwards to the receiver R_2.

The main distinguishing feature of the different subschemes is the choice of the subset of terminals—no terminal, both terminals, only R_1 or only R_2–that exploit the information in the transmitted messages to produce a guess of hypothesis $\mathcal{H}$. The other terminals ignore this communication and simply declare $\hat{\mathcal{H}} = 1$. The different subschemes occupy different communication rates, and as we shall see in the following Section V, the allocation of the rates has to be chosen in function of the desired tradeoff between the exponents θ_1 and θ_2. In this section, we formulate the subschemes based on generic hypothesis testing schemes for the two-hop network and the single-hop network with vanishing type-I error probabilities and respecting given rate constraints. Replacing these generic schemes by the optimal schemes under maximum-rate constraints [4], [27] attains the optimal error exponents presented in Theorem 2 ahead.

A. The case $\epsilon_1 = \epsilon_2 = \epsilon$

We combine two subschemes, where in one subscheme both R_1 and R_2 attempt to correctly guess the hypothesis $\mathcal{H}$ and in the other subscheme both simply declare $\hat{\mathcal{H}} = 1$. To this end, we partition the set $\mathcal{Y}_0^n$ into subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{1,2\}} \subseteq \mathcal{Y}_0^n$ so that under $P_{Y_0}^n$ the probability of subset $\mathcal{D}_{\{1,2\}}$ is as large as possible but satisfies

$$\Pr\left[Y_0^n \in \mathcal{D}_{\{1,2\}}\right] \leq 1 - \epsilon. \tag{17}$$

Notice that as $n \to \infty$ the inequality turns into an equality.

Depending on whether Y_0^n lies in $\mathcal{D}_\emptyset$ or $\mathcal{D}_{\{1,2\}}$, the three terminals follow a different subscheme.

<u>If $Y_0^n \in \mathcal{D}_\emptyset$</u>: In this case, none of the terminals attempts to correctly guess the hypothesis $\mathcal{H}$. Specifically, T_0 and R_1 both send

$$\mathsf{M}_1 = \mathsf{M}_2 = [0] \tag{18}$$

and $\mathrm{R}_1$and R_2 simply declare

$$\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1. \tag{19}$$

If $Y_0^n \in \mathcal{D}_{\{1,2\}}$: In this case, both R_1 and R_2 attempt to correctly guess $\mathcal{H}$ based on the transmitted messages. Specifically, T_0, R_1, R_2 all apply the encoding/decision functions of a given two-hop hypothesis testing scheme with vanishing type-I error probabilities and respecting maximum-rate constraints $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$ on the two links,[1] where these rates are chosen to satisfy

$$(1-\epsilon)R_{\{1,2\},1} \leq R_1 \tag{20a}$$
$$(1-\epsilon)R_{\{1,2\},2} \leq R_2. \tag{20b}$$

To inform all the terminals about the event $Y_0 \in \mathcal{D}_{\{1,2\}}$ and consequently about the employed scheme, T_0 and R_1 append the [1]-flag at the beginning of their messages M_1 and M_2.

Analysis: By (17) and (20), and because transmission of single bits hardly changes the communication rate for large blocklengths, the overall scheme satisfies the expected-rate constraints R_1 and R_2 on the two links. Appendix B proves that when the optimal two-hop hypothesis testing scheme with vanishing type-I error probability [27] is employed for $Y_0^n \in \mathcal{D}_{\{1,2\}}$, then the overall scheme meets the permissible type-I error probability ϵ and achieves the error exponent given by Equation (31) of Theorem 2.

B. The case $\epsilon_1 < \epsilon_2$

We combine three subschemes, where in each subscheme either no terminal, only R_1, or both R_1 and R_2 attempt to correctly guess $\mathcal{H}$. To this end, we partition the set $\mathcal{Y}_0^n$ into three disjoint subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{1\}}, \mathcal{D}_{\{1,2\}} \subseteq \mathcal{Y}_0^n$ so that under P_{Y_0} the two sets $\mathcal{D}_{\{1\}}$ and $\mathcal{D}_{\{1,2\}}$ have largest possible probabilities but limited by

$$\Pr\left[Y_0^n \in \mathcal{D}_{\{1\}}\right] \leq \epsilon_2 - \epsilon_1 \tag{21a}$$
$$\Pr\left[Y_0^n \in \mathcal{D}_{\{1,2\}}\right] \leq 1 - \epsilon_2. \tag{21b}$$

As a consequence,

$$\Pr\left[Y_0^n \in \mathcal{D}_\emptyset\right] \geq \epsilon_1. \tag{21c}$$

Notice that as $n \to \infty$, the three inequalities (21) can hold with equality.

Choose also nonnegative rates $R_{\{1\},1}$, $R_{\{1,2\},1}$, $R_{\{1,2\},2}$ satisfying

$$(\epsilon_2 - \epsilon_1)R_{\{1\},1} + (1-\epsilon_2)R_{\{1,2\},1} \leq R_1 \tag{22a}$$
$$(1-\epsilon_2)R_{\{1,2\},2} \leq R_2. \tag{22b}$$

Depending on whether Y_0^n lies in $\mathcal{D}_\emptyset$, $\mathcal{D}_{\{1\}}$, or $\mathcal{D}_{\{1,2\}}$, the three terminals apply a different subscheme satisfying a different pair of maximum-rate constraints, where the subscript $\mathcal{I}$ of set $\mathcal{D}_\mathcal{I}$ indicates the set of relays that attempt to correctly guess $\mathcal{H}$ in the event $Y_0^n \in \mathcal{D}_\mathcal{I}$. To communicate which of the three subschemes is used, T_0 adds a two-bit flag at the beginning of its message M_1 to R_1, which forwards this flag at the beginning of its message M_2 to inform R_2.

If $Y_0^n \in \mathcal{D}_\emptyset$: T_0 and R_1 send only the flag-bits

$$\mathsf{M}_1 = \mathsf{M}_2 = [0,0] \tag{23}$$

and R_1 and R_2 decide on

$$\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1. \tag{24}$$

If $Y_0^n \in \mathcal{D}_{\{1\}}$: T_0 and R_1 apply a given single-hop hypothesis testing scheme with vanishing type-I error probability and expected-rate constraint $R_{\{1\},1}$ for message M_1. Moreover, message M_1 is preceded by flag-bits $[1,0]$, and the relay R_1 forwards these flag-bits to R_2:

$$\mathsf{M}_2 = [1,0]. \tag{25}$$

Upon reception of these flag-bits, R_2 declares

$$\hat{\mathcal{H}}_2 = 1. \tag{26}$$

We observe that, as indicated by the subscript $\{1\}$ of set $\mathcal{D}_{\{1\}}$, only terminal R_1 attempts to correctly guess $\mathcal{H}$. Receiver R_2 produces the trivial guess in (26) because of its higher admissible type-I error probability $\epsilon_2 > \epsilon_1$. Notice also that no communication rate is required for message M_2 in the limit as $n \to \infty$.

If $Y_0^n \in \mathcal{D}_{\{1,2\}}$: T_0, R_1, R_2 apply a given two-hop hypothesis testing scheme with vanishing type-I error probabilities and satisfying the expected-rate constraints $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$.

Analysis: By (21) and (22), and because transmission of two bits hardly changes the rate for sufficiently large blocklengths, the proposed overall scheme respects the expected-rate constraints R_1 and R_2 for large values of n. Appendix C proves that when the optimal single-hop and two-hop hypothesis testing schemes under maximum-rate constraints $R_{\{1\},1}$ and $(R_{\{1,2\},1}, R_{\{1,2\},2})$ with vanishing type-I error probability [4], [27] are used, then the overall scheme satisfies the type-I error constraints ϵ_1 and ϵ_2 and achieves the error exponents in Equation (32) of Theorem 2.

C. The case $\epsilon_1 > \epsilon_2$

We combine three subschemes, where in each subscheme either no terminal, only R_2, or both R_1 and R_2 attempt to correctly guess $\mathcal{H}$. To this end, we partition the set $\mathcal{Y}_0^n$ into three disjoint subsets $\mathcal{D}_\emptyset, \mathcal{D}_{\{2\}}, \mathcal{D}_{\{1,2\}} \subseteq \mathcal{Y}_0^n$ so that under $P_{Y_0^n}$ the two sets $\mathcal{D}_{\{2\}}$ and $\mathcal{D}_{\{1,2\}}$ have largest possible probabilities but limited by

$$\Pr\left[Y_0^n \in \mathcal{D}_{\{2\}}\right] \leq \epsilon_1 - \epsilon_2 \tag{27a}$$
$$\Pr\left[Y_0^n \in \mathcal{D}_{\{1,2\}}\right] \leq 1 - \epsilon_1. \tag{27b}$$

As a consequence,

$$\Pr\left[Y_0^n \in \mathcal{D}_\emptyset\right] \geq \epsilon_2. \tag{27c}$$

Notice that as $n \to \infty$, the three inequalities (27) hold with equality.

Choose also nonnegative rates $R_{\{2\},1}$, $R_{\{1,2\},1}$, $R_{\{2\},2}$, and $R_{\{1,2\},2}$ satisfying

$$(\epsilon_1 - \epsilon_2)R_{\{1\},1} + (1-\epsilon_1)R_{\{1,2\},1} \leq R_1 \tag{28}$$

[1] As it will become clear in the subsequent analysis, for the overall scheme to respect rate constraints (4) and (6), it suffices that the two-hop scheme respects the rate constraints $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$ on expectation. However, as a consequence of our main result in Theorem 2, under vanishing type-I error probabilities, the same type-II error exponents are achievable under both expected- and maximum-rate constraints. There is thus no benefit in considering schemes with expected rates $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$, but possibly larger maximum rates.

$$(\epsilon_1 - \epsilon_2)R_{\{2\},2} + (1-\epsilon_1)R_{\{1,2\},2} \leq R_2. \tag{29}$$

Depending on whether Y_0^n lies in $\mathcal{D}_\emptyset$, $\mathcal{D}_{\{2\}}$, or $\mathcal{D}_{\{1,2\}}$, the three terminals apply a different subscheme. The subscript $\mathcal{I}$ of set $\mathcal{D}_\mathcal{I}$ again indicates the set of terminals that attempt to correctly guess $\mathcal{H}$ in the event $Y_0^n \in \mathcal{D}_\mathcal{I}$, and $R_{\mathcal{I},1}, R_{\mathcal{I},2}$ indicate the maximum rates of the subscheme employed under $Y_0^n \in \mathcal{D}_\mathcal{I}$. (An exception is the event $Y_0^n \in \mathcal{D}_\emptyset$, where both rates are 0.) Flag-bits are used at the beginning of the messages M_1 and M_2 to inform R_1 and R_2 about which of the subschemes is employed.

If $Y_0^n \in \mathcal{D}_\emptyset$: All three terminals, T_0, R_1, and R_2 apply the degenerate scheme in (23)–(24).

If $Y_0^n \in \mathcal{D}_{\{2\}}$: As indicated by the subscript of set $\mathcal{D}_{\{2\}}$, only R_2 makes a serious attempt to correctly guess $\mathcal{H}$, while R_1 always declares

$$\hat{\mathcal{H}}_1 = 1, \tag{30}$$

irrespective of the received message and its observations. This implies that under this subscheme, $\alpha_{1,n} = 1$ and $\beta_{1,n} = 0$. Besides this decision, T_0, R_1, and R_2 apply a given two-hop distributed hypothesis testing scheme with vanishing type-I error probabilities and respecting the maximum-rate constraints $R_{\{2\},1}$ and $R_{\{2\},2}$ for messages M_1 and M_2. Moreover, both T_0 and R_1 append the two-bit flag [0,1] at the beginning of these two messages to inform all the terminals about the employed scheme.

Notice that in the optimal two-hop hypothesis testing scheme [27], the relay R_1 computes a tentative decision based on M_1 and Y_1^n, which influences the message M_2 sent to R_2 and allows the latter to improve its type-I error probability. Here we propose that R_1 itself ignores its tentative decision, because the naive decision (30) is sufficient to satisfy the constraint ϵ_1 on its type-I error probability and is also the most-favorable decision to maximize the type-II error exponent.

If $Y_0^n \in \mathcal{D}_{\{1,2\}}$: Both decision centers R_1 and R_2 attempt to correctly guess $\mathcal{H}$. Specifically, T_0, R_1, and R_2 apply a given two-hop hypothesis testing scheme with vanishing type-I error probabilities and respecting the maximum-rate constraints $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$ for messages M_1 and M_2. Moreover, both T_0 and R_1 append the two-bit flag [1,1] at the beginning of these two messages to inform all the terminals about the employed scheme.

Analysis: Similarly to the case $\epsilon_1 < \epsilon_2$, it can be shown that the described scheme respects the expected-rate constraints (4) and (6) on both links, and that when the optimal two-hop scheme [27] is employed, then the described scheme achieves the error exponents in Equation (33) of Theorem 2.

V. Exponents Region for the Two-Hop Network under Expected-Rate Constraints

The fundamental exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ has a different form, depending on the three cases $\epsilon_1 = \epsilon_2$, $\epsilon_1 < \epsilon_2$, or $\epsilon_1 > \epsilon_2$.

Theorem 2: Given $\epsilon_1, \epsilon_2, R_1, R_2 \geq 0$.

If $\epsilon_1 = \epsilon_2 = \epsilon$, then $\mathcal{E}^*(R_1, R_2, \epsilon, \epsilon)$ is the set of all nonnegative (θ_1, θ_2) pairs satisfying

$$\theta_1 \leq \eta_1(R_1/(1-\epsilon)) \tag{31a}$$

$$\theta_2 \leq \eta_1(R_1/(1-\epsilon)) + \eta_2(R_2/(1-\epsilon)). \tag{31b}$$

If $\epsilon_1 < \epsilon_2$, then $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ is the set of all nonnegative (θ_1, θ_2) pairs satisfying

$$\theta_1 \leq \min\left\{\eta_1\left(R_{\{1\},1}\right), \eta_1\left(R_{\{1,2\},1}\right)\right\} \tag{32a}$$

$$\theta_2 \leq \eta_1\left(R_{\{1,2\},1}\right) + \eta_2\left(R_2/(1-\epsilon_2)\right), \tag{32b}$$

for some rates $R_{\{1\},1}, R_{\{1,2\},1} \geq 0$ so that

$$R_1 \geq (\epsilon_2 - \epsilon_1)R_{\{1\},1} + (1-\epsilon_2)R_{\{1,2\},1}. \tag{32c}$$

If $\epsilon_1 > \epsilon_2$, then $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ is the set of all nonnegative (θ_1, θ_2) pairs satisfying

$$\theta_1 \leq \eta_1(R_{\{1,2\},1}) \tag{33a}$$

$$\theta_2 \leq \min\left\{\eta_1(R_{\{1,2\},1}) + \eta_2\left(R_{\{1,2\},2}\right), \eta_1\left(R_{\{2\},1}\right) + \eta_2\left(R_{\{2\},2}\right)\right\} \tag{33b}$$

for some rates $R_{\{1,2\},1}$, $R_{\{2\},1}$, $R_{\{1,2\},2}$, $R_{\{2\},2} \geq 0$, so that

$$R_1 \geq (\epsilon_1 - \epsilon_2)R_{\{2\},1} + (1-\epsilon_1)R_{\{1,2\},1} \tag{33c}$$

$$R_2 \geq (\epsilon_1 - \epsilon_2)R_{\{2\},2} + (1-\epsilon_1)R_{\{1,2\},2}. \tag{33d}$$

Proof: Achievability is based on the schemes in Section IV, see Appendices B and C for their analyses. The converse is proved in Section VI. ■

Remark 2 (Discussion for $\epsilon_1 = \epsilon_2$): By above theorem, for $\epsilon_1 = \epsilon_2 = \epsilon$, the fundamental exponents region $\mathcal{E}^*(R_1, R_2, \epsilon, \epsilon)$ is a rectangle. Also, compared to the fundamental exponents region under maximum-rate constraints, here the rates are boosted by a factor $(1-\epsilon)^{-1}$:

$$\mathcal{E}^*(R_1, R_2, \epsilon, \epsilon) = \mathcal{E}^*_{\text{max}}\left(\frac{R_1}{(1-\epsilon)}, \frac{R_2}{(1-\epsilon)}\right). \tag{34}$$

In particular, for $\epsilon_1 = \epsilon_2 = 0$ the fundamental exponents regions under maximum- and expected-rates coincide:

$$\mathcal{E}^*(R_1, R_2, 0, 0) = \mathcal{E}^*_{\text{max}}(R_1, R_2). \tag{35}$$

Remark 3 (Discussion for $\epsilon_1 < \epsilon_2$): For $\epsilon_1 \neq \epsilon_2$, the fundamental exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ is not a rectangle, as can be verified by the numerical results in Figures 2, 3, and 4 in the next subsection. In fact, one observes a tradeoff between the two exponents θ_1 and θ_2, which is driven by the choice of the rates $R_{\mathcal{I},1}, R_{\mathcal{I},2}$ for $\mathcal{I} \in \mathcal{P}(2)$, where $\mathcal{P}(2)$ is the power set of all subsets of $\{1,2\}$ excluding the emptyset, i.e. $\mathcal{P}(2) = \{\{1\}, \{2\}, \{1,2\}\}$. More specifically, for $\epsilon_1 < \epsilon_2$ the choice

$$R_{\{1,2\},1} = R_1/(1-\epsilon_2) \tag{36a}$$

$$R_{\{1\},1} = 0 \tag{36b}$$

maximizes exponent θ_2, which then evaluates to

$$\theta_2 = \theta_{2,\max} := \eta_1\left(R_1/(1-\epsilon_2)\right) + \eta_2\left(R_2/(1-\epsilon_2)\right), \tag{37}$$

but completely degrades θ_1 to $\theta_1 = 0$. (Notice that for large $R_1/(1-\epsilon_2)$ above choice (36) might not be the unique optimizer and other optimizers will still allow to attain a positive η_1.)

On the other hand, the choice

$$R_{\{1\},1} = R_{\{1,2\},1} = R_1/(1-\epsilon_1) \tag{38}$$

maximizes exponent θ_1, which then evaluates to

$$\theta_1 = \theta_{1,\max} := \eta_1 \left(R_1/(1-\epsilon_1)\right), \tag{39}$$

but it degrades θ_2 to

$$\theta_2 = \theta_{2,\text{deg}} := \eta_1 \left(R_1/(1-\epsilon_1)\right) + \eta_2 \left(R_2/(1-\epsilon_2)\right) < \theta_{2,\max}. \tag{40}$$

Varying the rate $R_{\{1,2\},1}$ between the choices in (36) and (38), (and accordingly varying also rate $R_{\{1\},1}$ to meet (32c)) achieves the entire Pareto-optimal boundary of the fundamental exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$.

Remark 4 (Discussion for $\epsilon_1 > \epsilon_2$): For $\epsilon_1 > \epsilon_2$ the choice

$$R_{\{1,2\},1} = R_1/(1-\epsilon_1) \tag{41a}$$

$$R_{\{2\},1} = 0 \tag{41b}$$

maximizes exponent θ_1, which then evaluates to

$$\theta_1 = \theta_{1,\max} \tag{42}$$

and degrades θ_2 to

$$\theta_2 = \theta'_{2,\text{deg}} := \min\left\{\eta_1 \left(R_1/(1-\epsilon_1)\right) + \eta_2(R_{\{1,2\},2}), \eta_2(R_{\{2\},2})\right\}, \tag{43}$$

for $R_{\{2\},2}$ and $R_{\{1,2\},2}$ satisfying (33d). (Notice again that for large values of $R_1/(1-\epsilon_1)$ the optimizer in (41) might not be unique and other optimizers might lead to a larger value of θ_2.)

On the other hand, the choice

$$R_{\{2\},1} = R_{\{1,2\},1} = R_1/(1-\epsilon_2) \tag{44a}$$

$$R_{\{2\},2} = R_{\{1,2\},2} = R_2/(1-\epsilon_2) \tag{44b}$$

maximizes exponent θ_2, which then evaluates to $\theta_2 = \theta_{2,\max}$, but it degrades θ_1 to

$$\theta_1 = \theta_{1,\text{deg}} := \eta_1 \left(R_1/(1-\epsilon_2)\right) \tag{45}$$

Varying the rate $R_{\{1,2\},1}$ between the choices in (41) and (44) (and varying the rates $R_{\{1\},1}, R_{\{1,2\},2}, R_{\{1\},2}$ accordingly), achieves the entire Pareto-optimal boundary of the fundamental exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$.

Remark 5 (Rate-boosts when $\epsilon_1 \neq \epsilon_2$): Notice that in our two-hop system with expected-rate constraints, exponents $\theta_{1,\max}$ and $\theta_{2,\max}$ defined in (39) and (37), are the largest possible exponents achievable at the two decision centers, irrespective of the ordering of ϵ_1 and ϵ_2. By Theorem 3, they coincide with the optimal exponents under *maximum-rate constraints* $R_1/(1-\epsilon_1)$ and $R_2/(1-\epsilon_1)$ for the two links in case of (39), and maximum-rate constraints $R_1/(1-\epsilon_2)$ and $R_2/(1-\epsilon_2)$ in case of (37). We thus observe that whenever $\epsilon_1 \neq \epsilon_2$, the rate-boosts that *expected*-rate constraints allow to obtain over *maximum*-rate constraints depend on the permissible type-I error probabilities and also on the tradeoff between the two exponents θ_1 and θ_2. In this view, notice that when the focus is on maximizing θ_2, then for $\epsilon_1 < \epsilon_2$ one has to entirely sacrifice θ_1, whereas for $\epsilon_1 > \epsilon_2$ positive θ_1-exponents are possible but the rate-boost experienced by θ_1 is reduced from $(1-\epsilon_1)^{-1}$, which is the boost experienced for its maximum $\theta_{1,\max}$, to the smaller factor $(1-\epsilon_2)^{-1}$.

A. Numerical Simulations

In this section, we illustrate the benefits of exploiting the relaxed expected-rate constraints in (4) and (6) compared to the more stringent maximum-rate constraints (13) at hand of some examples. We also show for $\epsilon_1 < \epsilon_2$ the benefits of "Rate-sharing" on the first link and the corresponding tradeoff, where the rate R_1 is split into $(\epsilon_2 - \epsilon_1)R_{\{1\},1}$ and $(1-\epsilon_2)R_{\{1,2\},1}$ as in (32), instead of restricting to a single rate choice for the communication on the first link $R_{\{1,2\},1} = R_1/(1-\epsilon_1)$. For the case $\epsilon_1 < \epsilon_2$, "Rate-sharing" on the second link does not have any added value. However, for the case $\epsilon_1 > \epsilon_2$, we illustrate the benefits of "Rate-sharing" on both links and the resulting tradeoff from varying the choices of the rates $R_{\{1,2\},1}$, $R_{\{2\},1}$, $R_{\{1,2\},2}$ and $R_{\{2\},2}$ that satisfy (33). This tradeoff stems from multiplexing three coding subschemes among which we have two full versions of the basic two-hop scheme and one degraded subscheme as explained in Subsection IV-C.

Throughout this section we consider the following example.

Example 1: Let Y_0, S, T be independent Bernoulli random variables of parameters $p_{Y_0} = 0.4, p_S = 0.8, p_T = 0.8$ and set $Y_1 = Y_0 \oplus T$ and $Y_2 = Y_1 \oplus S$.

We first consider the case $\epsilon_1 = 0.05 < \epsilon_2 = 0.15$, and plot the optimal exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ in Figure 2 for symmetric rates $R_1 = R_2 = 0.5$. We note a tradeoff between the type-II error exponents θ_1 and θ_2, which is not present neither for the case $\epsilon_1 = \epsilon_2$, nor for the same setup under maximum-rate constraints. (This tradeoff occurs because both exponents have to be optimized over the same choices of rates $R_{\{1\},1}, R_{\{1,2\},1}$.) The figure also shows a sub-optimal version of the exponents region in Theorem 2, where we set $R_{\{1\},1} = R_{\{1,2\},1} = R_1/(1-\epsilon_1)$ and thus obtain $\mathcal{E}^*_{\max}(R_1/(1-\epsilon_1), R_2/(1-\epsilon_2))$. Comparing these two regions, we observe that using two different rates $R_{\{1\},1}$ and $R_{\{1,2\},1}$ (i.e., two different versions of the basic two-hop scheme) allows to obtain a better tradeoff between the two exponents. For futher comparison, Figure 2 also shows the exponents region $\mathcal{E}^*_{\max}(R_1, R_2)$ under maximum-rate constraints, so as to illustrate the gain provided by having the relaxed expected-rate constraints instead of maximum-rate constraints.

We then consider the case $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$. Here we consider three sub-cases for the rates: symmetric rates $R_1 = R_2 = 0.5$ or asymmetric rates $R_1 = 0.75 > R_2 = 0.25$ or $R_1 = 0.25 > R_2 = 0.75$.

In Figure 3 we plot the optimal exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ in Theorem 2 for the first sub-case $R_1 = R_2 = 0.5$, and we compare it with the exponents region under maximum-rate constraints $\mathcal{E}^*_{\max}(R_1, R_2)$ and with sub-optimal versions of Theorem 3 where we either set $R_{\{1,2\},1} = R_{\{2\},1}$, for which we obtain $\mathcal{E}^*_{\max}\left(\frac{R_1}{(1-\epsilon_2)}, \frac{R_2}{(1-\epsilon_2)}\right)$, or we set

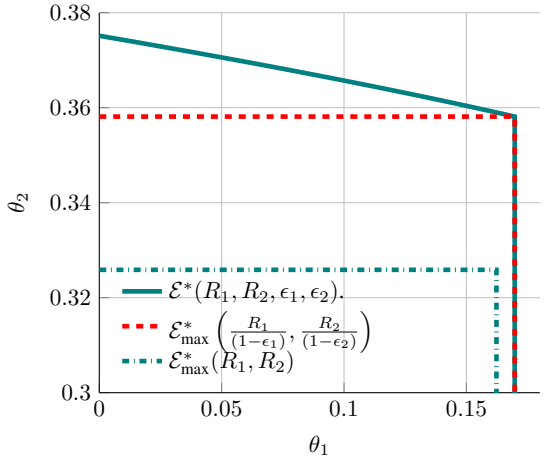

Fig. 2: Exponents regions for Example 1 when $\epsilon_1 = 0.05 < \epsilon_2 = 0.15$ and $R_1 = R_2 = 0.5$.

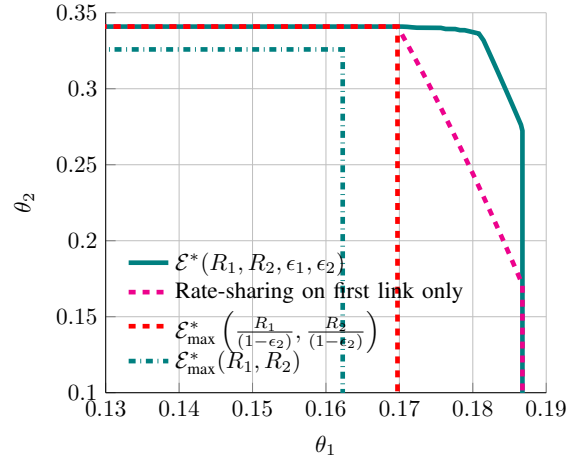

Fig. 3: Exponents regions under expected- and maximum-rate constraints for Example 1 when $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$ and $R_1 = R_2 = 0.5$.

$R_{\{1,2\},2} = R_{\{2\},2}$, for which we have a tradeoff between the type-II error exponents due to rate-sharing on the first link. Comparing all these regions, we see that rate-sharing on the first link allows to obtain a smooth tradeoff between the exponents, while rate-sharing on both links (i.e., having two full versions of the basic two-hop scheme) yields an even improved tradeoff.

Figure 4 compares the exponents regions under expected rate-constraints for all three sub-cases. Clearly, θ_1 is increasing in R_1, but θ_2 is not necessarily increasing in R_2, since it also depends on R_1. In fact, exponents region $\mathcal{E}^*(0.25, 0.75, \epsilon_1, \epsilon_2)$ is completely included in exponents region $\mathcal{E}^*(0.5, 0.5, \epsilon_1, \epsilon_2)$.

To understand these phenomena, notice that the maximum achievable exponents on each communication link are $\eta_1^*(R_1) = I(Y_0; Y_1) = 0.26766$ and $\eta_2^*(R_2) = I(Y_1; Y_2) = 0.27433$. Recall also that the θ_2-error exponent is an accumulation of the error exponents given by both functions $\eta_1(\cdot) + \eta_2(\cdot)$. The similar behaviours of the two functions $\eta_1(r) \approx \eta_2(r)$ ($r \in [0, 1]$), together with the concavity and monotonicity of these functions, induce that to obtain the largest θ_2 values in this example, the total rate should be distributed almost equally between both links. In contrast, since the θ_1-error exponent depends only on rate R_1, the largest value is achieved by putting all available rate to R_1. All of the above explains the superiority of the error exponent region obtained when $R_1 = R_2 = 0.5$ over the one obtained when $R_1 = 0.25, R_2 = 0.75$, and the tradeoff between the exponents regions for the sub-cases $R_1 = R_2 = 0.5$ and $R_1 = 0.75, R_2 = 0.25$.

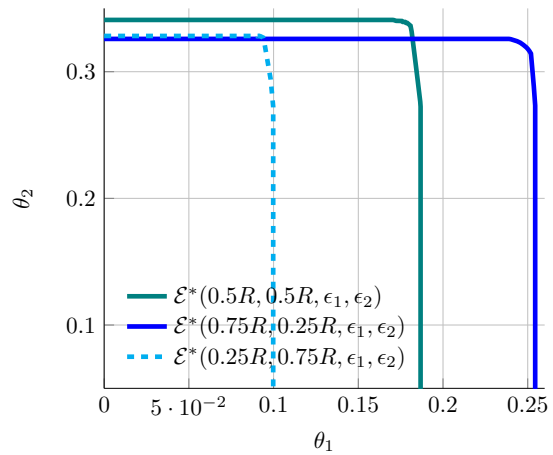

Fig. 4: Exponents regions under symmetric and asymmetric expected-rate constraints for Example 1 and when $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$ and $R = 1$.

VI. Converse Proof to Theorem 2

A. Outline of the Converse Proof

The main idea of the proof is to divide the set of strongly typical sequences $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1Y_2})$ into four subsets according to the decisions taken at the two DCs, see Figure 5. Parallel change of measure arguments are then applied to each subset and different lower bounds on the expected message lengths and upper bounds on the error exponents are derived. The (implicit) lower bounds on the expected message lengths are actually lower bounds on the conditional expectations given each of the four subsets and thus combined by means of the law of total expectation to obtain a bound on the total expected message length. The derived upper bounds on the error exponents are not conditioned on the subsets but hold in an unconditional sense. The final bound on the type-II error exponent is then simply obtained by considering the tightest of the derived bounds, which explains the minimizations in Theorem 2.

To match the so-obtained bounds on the message lengths and the type-II error exponents with our achievability result (see the scheme in Section IV) two final steps are needed: 1) Show that the introduced auxiliary random variables satisfy certain Markov chains in the asymptotic regime of infinite blocklengths;[2] and 2) Show that depending on the ordering

[2]In the weak converse proofs that do not have to apply a change of measure argument these Markov chains are naturally true for all blocklengths because of the i.i.d.-ness of the source sequences.

$\epsilon_1 = \epsilon_2$, $\epsilon_1 < \epsilon_2$, or $\epsilon_1 > \epsilon_2$, there exists an optimal scheme that under $\mathcal{H} = 0$ assigns asymptotically 0 probability to specific subsets of the introduced partition. For example, if $\epsilon_1 \leq \epsilon_2$ the probability under $\mathcal{H} = 0$ that R_2 decides on $\mathcal{H} = 0$ but R_1 does not should vanish, and if $\epsilon_1 \geq \epsilon_2$ the probability that R_1 decides on $\mathcal{H} = 0$ but R_2 does not should vanish.

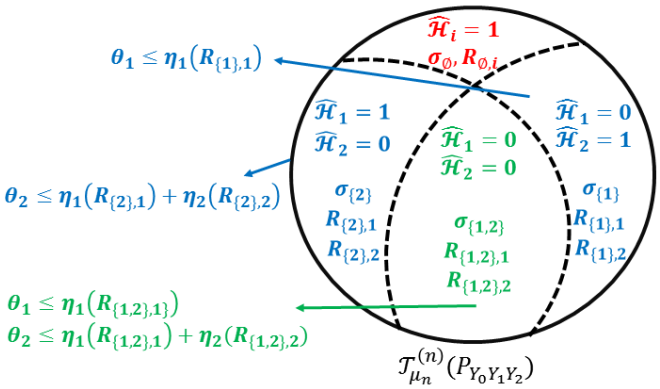

Fig. 5: Sketch of partitioning $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1Y_2})$ and applying parallel change of measure arguments in the converse proof to Theorem 2.

Technically, the main part of the proof is contained in the subsequent Proposition 3, which states the converse bound before the mentioned simplification depending on the ordering of ϵ_1 and ϵ_2, which is proved in Appendix D. The proof of Proposition 3 applies Lemma 2 ahead to each of the four subsets introduced in Figure 5 to obtain bounds on the expected message length and the error exponents, and to prove the desired Markov chains asymptotically.

B. The Converse Proof

Consider 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 rates and error probabilities in (12). The converse proof is based on the following lemma.

Lemma 2: Consider an arbitrary blocklength n and a subset $\mathcal{D} \subseteq \mathcal{Y}_0^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$. Let the tuple $(\tilde{\mathsf{M}}_1, \tilde{\mathsf{M}}_2, \tilde{Y}_0^n, \tilde{Y}_1^n, \tilde{Y}_2^n)$ follow the pmf

$$P_{\tilde{\mathsf{M}}_1\tilde{\mathsf{M}}_2\tilde{Y}_0^n\tilde{Y}_1^n\tilde{Y}_2^n}(\mathsf{m}_1, \mathsf{m}_2, y_0^n, y_1^n, y_2^n) \triangleq P_{Y_0^nY_1^nY_2^n}(y_0^n, y_1^n, y_2^n) \cdot \frac{\mathbb{1}\{(y_0^n, y_1^n, y_2^n) \in \mathcal{D}\}}{P_{Y_0^nY_1^nY_2^n}(\mathcal{D})} \cdot \mathbb{1}\{\phi_1^{(n)}(y_0^n) = \mathsf{m}_1\} \cdot \mathbb{1}\{\phi_2^{(n)}(y_1^n, \phi_1(y_0^n)) = \mathsf{m}_2\}. \quad (46)$$

Further, define

$$U_1 \triangleq (\tilde{\mathsf{M}}_1, \tilde{Y}_0^{T-1}, \tilde{Y}_1^{T-1}, \tilde{Y}_2^{T-1}, T) \quad (47)$$
$$U_2 \triangleq (\tilde{\mathsf{M}}_2, \tilde{Y}_0^{T-1}, \tilde{Y}_1^{T-1}, \tilde{Y}_2^{T-1}, T) \quad (48)$$
$$\tilde{Y}_k \triangleq \tilde{Y}_{k,T}, \quad k \in \{0, 1, 2\}, \quad (49)$$

where T is uniform over $\{1, \ldots, n\}$ and independent of all previously defined random variables.

The following (in)equalities hold:

$$\frac{1}{n} H(\tilde{\mathsf{M}}_k) \geq I(U_k; \tilde{Y}_{k-1}) + \frac{1}{n} \log P_{Y_0^nY_1^nY_2^n}(\mathcal{D}), \quad k \in \{1, 2\}, \quad (50)$$

$$I(U_1; \tilde{Y}_1 | \tilde{Y}_0) \leq -\frac{1}{n} \log P_{Y_0^nY_1^nY_2^n}(\mathcal{D}), \quad (51)$$

$$I(U_2; \tilde{Y}_2 | \tilde{Y}_1) \leq -\frac{2}{n} \log P_{Y_0^nY_1^nY_2^n}(\mathcal{D}) + D(P_{\tilde{Y}_0\tilde{Y}_1\tilde{Y}_2} \| P_{Y_0Y_1Y_2}). \quad (52)$$

Moreover, if the decision center R_k decides on the null hypothesis $\hat{\mathcal{H}}_k = 0$ for all tuples $(y_0^n, y_1^n, y_2^n) \in \mathcal{D}$, then

$$-\frac{1}{n} \log \beta_{k,n} \leq \sum_{\ell=1}^{k} I(U_\ell; \tilde{Y}_\ell) - \frac{k+1}{n} \log P_{Y_0^nY_1^nY_2^n}(\mathcal{D}). \quad (53)$$

Proof: The lemma is essentially a special case of Lemma 4 ahead, which is proved in Appendix VIII-C. The slightly stronger statement in (51) can easily be verified at hand of the proof in Appendix VIII-C. ■

Remark 6 (Discussion of Lemma 2): Inequalities (50) and (53) provide bounds on the entropy of a message and on the error exponent in case the set $\mathcal{D}$ is well chosen.

Inequalities (51) and (52) state that when the probability $P_{Y_0^nY_1^nY_2^n}(\mathcal{D})$ does not vanish exponentially fast in n and the set $\mathcal{D}$ is a subset of the typical set (in which case the divergence $D(P_{\tilde{Y}_0\tilde{Y}_1\tilde{Y}_2} \| P_{Y_0Y_1Y_2})$ is arbitrarily small), then for large blocklengths the pmfs $P_{U_k\tilde{Y}_k\tilde{Y}_{k-1}}$ are close to the product pmfs $P_{U_k|\tilde{Y}_k} P_{\tilde{Y}_{k-1}\tilde{Y}_k}$ indicating indicating the Markov chain $\tilde{Y}_{k-1} \to \tilde{Y}_k \to U_k$. Equivalent Markov chains show up in the achievability result and thus have to be recovered for the converse to be tight.

With this lemma, we can prove the desired general outer bound on the exponents region.

Proposition 3: Given $R_1, R_2, \epsilon_1, \epsilon_2, \geq 0$, the fundamental exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ is included in the set of all (θ_1, θ_2) pairs satisfying

$$\theta_1 \leq \min\{\eta_1(R_{\{1\},1}),\ \eta_1(R_{\{1,2\},1})\}, \quad (54a)$$
$$\theta_2 \leq \min\{\eta_1(R_{\{1,2\},1}) + \eta_2(R_{\{1,2\},2}), \eta_1(R_{\{2\},1}) + \eta_2(R_{\{2\},2})\}, \quad (54b)$$

for rates $R_{\{1\},1}, R_{\{1,2\},1}, R_{\{1,2\},2}, R_{\{2\},1}, R_{\{2\},2} \geq 0$ and numbers $\sigma_{\{1\}}, \sigma_{\{2\}}, \sigma_{\{1,2\}} \geq 0$ so that $\sigma_{\{1\}} + \sigma_{\{2\}} + \sigma_{\{1,2\}} \leq 1$ and

$$\sigma_{\{1\}} + \sigma_{\{1,2\}} \geq 1 - \epsilon_1 \quad (54c)$$
$$\sigma_{\{2\}} + \sigma_{\{1,2\}} \geq 1 - \epsilon_2 \quad (54d)$$
$$\sigma_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}, \quad (54e)$$

and so that the following rate constraints are satisfied:

$$R_1 \geq \sigma_{\{1\}} R_{\{1\},1} + \sigma_{\{1,2\}} R_{\{1,2\},1} + \sigma_{\{2\}} R_{\{2\},1}, \quad (54f)$$
$$R_2 \geq \sigma_{\{1,2\}} R_{\{1,2\},2} + \sigma_{\{2\}} R_{\{2\},2}. \quad (54g)$$

In Appendix D we show that the fundamental exponents region presented in this proposition is contained in the simpler region in Theorem 2, which concludes the proof. (Actually

the region in Proposition 3 is not only included in the simpler region of Theorem 2 but coincides with it. The proof of the reverse direction, i.e., inclusion of the region in Theorem 2 in the region of Proposition 3, is however trivial because Proposition 3 is a converse result and the region in Theorem 2 is achievable.)

Proof of Proposition 3: Choose μ_n as a sequence satisfying

$$\lim_{n\to\infty} \mu_n = 0 \tag{55}$$

$$\lim_{n\to\infty} \mu_n n^2 = \infty \tag{56}$$

and define for each blocklenth n the sets

$$\mathcal{B}_{1,n} \triangleq \{(y_0^n, y_1^n, y_2^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1Y_2}) : g_1^{(n)}(y_1^n, \phi_0^{(n)}(y_0^n)) = 0\}, \tag{57}$$

$$\mathcal{B}_{2,n} \triangleq \{(y_0^n, y_1^n, y_2^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1Y_2}) : g_2^{(n)}(y_2^n, \phi_1^{(n)}(y_1^n, \phi_0^{(n)}(y_0^n))) = 0\}, \tag{58}$$

$$\mathcal{D}_{\{1,2\},n} \triangleq \mathcal{B}_{1,n} \cap \mathcal{B}_{2,n}, \tag{59}$$

$$\mathcal{D}_{\{1\},n} \triangleq \mathcal{B}_{1,n} \backslash \mathcal{D}_{\{1,2\},n}, \tag{60}$$

$$\mathcal{D}_{\{2\},n} \triangleq \mathcal{B}_{2,n} \backslash \mathcal{D}_{\{1,2\},n}. \tag{61}$$

Define also the set

$$\mathcal{D}_{\emptyset,n} = \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1Y_2}) \backslash (\mathcal{D}_{\{1\},n} \cup \mathcal{D}_{\{2\},n} \cup \mathcal{D}_{\{1,2\},n}). \tag{62}$$

Notice that the sets $\mathcal{D}_{\emptyset,n}, \mathcal{D}_{\{1\},n}, \mathcal{D}_{\{2\},n}, \mathcal{D}_{\{1,2\},n}$ partition the strongly typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0Y_1Y_2})$ and in each set $\mathcal{D}_{\mathcal{I},n}$ only the terminals R_k with $k \in \mathcal{I}$ declare $\hat{\mathcal{H}}_k = 0$, while terminals R_k with $k \notin \mathcal{I}$ declare $\hat{\mathcal{H}}_k = 1$.

Define the probabilities

$$\Delta_{\mathcal{I},n} \triangleq P_{Y_0^nY_1^nY_2^n}(\mathcal{D}_{\mathcal{I},n}), \quad \mathcal{I} \in \mathcal{P}(2), \tag{63}$$

and notice that by the laws of probability

$$\Delta_{\{1,2\}} + \Delta_{\{1\}} = P_{Y_0^nY_1^nY_2^n}(\mathcal{B}_1) \tag{64}$$

$$\Delta_{\{1,2\}} + \Delta_{\{2\}} = P_{Y_0^nY_1^nY_2^n}(\mathcal{B}_2) \tag{65}$$

$$\Delta_{\{1,2\}} \geq P_{Y_0^nY_1^nY_2^n}(\mathcal{B}_1) + P_{Y_0^nY_1^nY_2^n}(\mathcal{B}_2) - 1. \tag{66}$$

Notice further that by the type-I error probability constraints (12b) and by [42, Remark to Lemma 2.12] and basic laws of probability:

$$P_{Y_0^nY_1^nY_2^n}(\mathcal{B}_k) \geq 1 - \epsilon_k - \frac{|\mathcal{Y}_0||\mathcal{Y}_1||\mathcal{Y}_2|}{4\mu_n^2 n}, \quad k \in \{1,2\}, \tag{67}$$

so that we conclude that:

$$\varliminf_{n\to\infty} (\Delta_{\{1,2\},n} + \Delta_{\{1\},n}) \geq 1 - \epsilon_1 \tag{68a}$$

$$\varliminf_{n\to\infty} (\Delta_{\{1,2\},n} + \Delta_{\{2\},n}) \geq 1 - \epsilon_2 \tag{68b}$$

$$\varliminf_{n\to\infty} \Delta_{\{1,2\},n} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\} \tag{68c}$$

$$\varliminf_{n\to\infty} \sum_{\mathcal{I}\in\mathcal{P}(2)} \Delta_{\mathcal{I},n} \leq 1, \tag{68d}$$

where (68d) holds (for any blocklength) by the basic laws of probability because the sets $\mathcal{D}_{\mathcal{I},n}$ are disjoint.

To simplify exposition, we assume that for any blocklength n the probabilities $\Delta_{\mathcal{I},n} > 0$ for all sets $\mathcal{I} \in \mathcal{P}(2)$. (Otherwise a corresponding subsequence of blocklengths should be considered if it exists. And if no such subsequence exists, the corresponding sets should be discarded. Details of the proofs in these other cases are omitted for brevity.) Applying Lemma 2 to the four subsets allows to conclude that for any $\mathcal{I} \in \mathcal{P}(2)$ there exists a pair $(U_{\mathcal{I},1}, U_{\mathcal{I},2})$ satisfying the (in)equalities

$$H(\tilde{\mathsf{M}}_{\mathcal{I},1}) \geq nI(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},0}) + \log \Delta_{\mathcal{I},n}, \tag{69}$$

$$H(\tilde{\mathsf{M}}_{\mathcal{I},2}) \geq nI(U_{\mathcal{I},2}; \tilde{Y}_{\mathcal{I},1}) + \log \Delta_{\mathcal{I},n} \tag{70}$$

$$-\frac{1}{n}\log \Delta_{\mathcal{I},n} \geq I(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},1} | \tilde{Y}_{\mathcal{I},0}), \tag{71}$$

$$-\frac{2}{n}\log \Delta_{\mathcal{I},n} + D(P_{\tilde{Y}_0\tilde{Y}_1\tilde{Y}_2} \| P_{Y_0Y_1Y_2}) \geq I(U_{\mathcal{I},2}; \tilde{Y}_{\mathcal{I},2} | \tilde{Y}_{\mathcal{I},1}) \tag{72}$$

and

$$-\frac{1}{n}\log \beta_{1,n} \leq I(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},1}) - \frac{2}{n}\log \Delta_{\mathcal{I}}, \quad \text{if } \mathcal{I} \in \{\{1\},\{1,2\}\}, \tag{73}$$

$$-\frac{1}{n}\log \beta_{2,n} \leq I(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},1}) + I(U_{\mathcal{I},2}; \tilde{Y}_{\mathcal{I},2}) - \frac{3}{n}\log \Delta_{\mathcal{I}}, \quad \text{if } \mathcal{I} \in \{\{2\},\{1,2\}\}. \tag{74}$$

Define the following random variables

$$\tilde{L}_{\mathcal{I},j} \triangleq \mathrm{len}(\tilde{\mathsf{M}}_{\mathcal{I},j}), \quad j \in \{1,2\},\ \mathcal{I} \in \mathcal{P}(2). \tag{75}$$

By the rate constraints (4) and (6), and the definition of the random variables $\tilde{L}_{\mathcal{I},j}$, we obtain by the total law of expectations:

$$nR_1 \geq \mathbb{E}[L_1] \tag{76}$$

$$\geq \sum_{\mathcal{I}\in\mathcal{P}(2)} \mathbb{E}[\tilde{L}_{\mathcal{I},1}]\Delta_{\mathcal{I},n}. \tag{77}$$

Moreover,

$$H(\tilde{\mathsf{M}}_{\mathcal{I},1}) = H(\tilde{\mathsf{M}}_{\mathcal{I},1}, \tilde{L}_{\mathcal{I},1}) \tag{78}$$

$$= \sum_{l_{\mathcal{I}}=1}^{\infty} \Pr[\tilde{L}_{\mathcal{I},1} = l_{\mathcal{I}}] H(\tilde{\mathsf{M}}_{\mathcal{I},1} | \tilde{L}_{\mathcal{I},1} = l_{\mathcal{I}}) + H(\tilde{L}_{\mathcal{I},1}) \tag{79}$$

$$\leq \sum_{l_{\mathcal{I}}=1}^{\infty} \Pr[\tilde{L}_{\mathcal{I},1} = l_{\mathcal{I}}] l_{\mathcal{I}} + H(\tilde{L}_{\mathcal{I},1}) \tag{80}$$

$$= \mathbb{E}[\tilde{L}_{\mathcal{I},1}] + H(\tilde{L}_{\mathcal{I},1}), \tag{81}$$

which combined with (69) and (77) and the nonnegativity of entropy establishes

$$\sum_{\mathcal{I}\in\mathcal{P}(2)} \Delta_{\mathcal{I},n} \left[I(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},0}) + \frac{1}{n}\log \Delta_{\mathcal{I},n} \right]^+ \leq \sum_{\mathcal{I}\in\mathcal{P}(2)} \left(\frac{1}{n}\Delta_{\mathcal{I},n}\mathbb{E}[\tilde{L}_{\mathcal{I},1}] + \frac{1}{n}\Delta_{\mathcal{I},n} H(\tilde{L}_{\mathcal{I},1}) \right) \tag{82}$$

$$\leq R_1 \left(1 + \sum_{\mathcal{I}\in\mathcal{P}(2)} h_b\left(\frac{\Delta_{\mathcal{I},n}}{nR_1} \right) \right), \tag{83}$$

where (83) holds by (77) and because the entropy of the discrete and positive random variable $\tilde{L}_{\mathcal{I},1}$ of mean $\mathbb{E}[\tilde{L}_{\mathcal{I},1}] \leq \frac{nR_1}{\Delta_{\mathcal{I},n}}$ is bounded by $H(\tilde{L}_{\mathcal{I},1}) \leq \frac{nR_1}{\Delta_{\mathcal{I},n}} \cdot h_b\left(\frac{\Delta_{\mathcal{I},n}}{nR_1}\right)$, which is easily seen by the following two facts: 1) The KL-divergence $D(P\|Q)$ between an arbitrary pmf $P(x)$ over the positive integers of mean s and the geometric distribution of mean s equals $-H_P(X) + H_Q(X)$ and is nonnegative; and 2) The entropy of a geometric distribution of mean s is $H_Q(X) = s \cdot h_b(\frac{1}{s})$.

In a similar way, we obtain

$$\sum_{\substack{\mathcal{I}\in \\ \{\{1,2\},\{2\}\}}} \Delta_{\mathcal{I},n} \left[I(U_{\mathcal{I},1}; \tilde{Y}_{\mathcal{I},0}) + \frac{1}{n} \log \Delta_{\mathcal{I},n} \right]^+ \leq nR_2 \left(1 + \sum_{\substack{\mathcal{I}\in \\ \{\{1,2\},\{2\}\}}} h_b\left(\frac{\Delta_{\mathcal{I}}}{R_2}\right) \right). \quad (84)$$

The proof then follows from (71)–(74) and from (83) and (84) and by letting $n \to \infty$. Details are as follows. By Carathéodory's theorem, we can replace above auxiliary random variables $\{(U_{\mathcal{I},1}, U_{\mathcal{I},2})\}_{\mathcal{I}}$, which are of increasing alphabet sizes, by random variables over constant alphabets $\mathcal{U}_{\mathcal{I},1}$ and $\mathcal{U}_{\mathcal{I},2}$ of sizes

$$|\mathcal{U}_{\mathcal{I},1}| \leq |\mathcal{Y}_0| \cdot |\mathcal{Y}_1| + 2, \qquad \mathcal{I} \in \mathcal{P}(2), \quad (85)$$

$$|\mathcal{U}_{\mathcal{I},2}| \leq |\mathcal{U}_{\mathcal{I},1}| \cdot |\mathcal{Y}_0| \cdot |\mathcal{Y}_1| + 1, \quad \mathcal{I} \in \{\{1,2\},\{2\}\}, \quad (86)$$

while still preserving inequalities (71)–(74), (83), and (84).

Let $P^{(n)}_{\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}U_{\mathcal{I},1}U_{\mathcal{I},2}}$ denote the pmf of these new tuples $(\tilde{Y}_{\mathcal{I},0}, \tilde{Y}_{\mathcal{I},1}, \tilde{Y}_{\mathcal{I},2}, U_{\mathcal{I},1}, U_{\mathcal{I},2})$ at blocklength n, and abbreviate it as $P^*_{\mathcal{I}}$. We invoke the Bolzano-Weierstrass theorem and consider an increasing subsequence of positive blocklengths $\{n_i\}_{i=1}^{\infty}$ such that for all $\mathcal{I} \in \mathcal{P}(2)$ the subsequences of probabilities $\Delta_{\mathcal{I},n_i}$ and pmfs $P^{(n_i)}_{\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}U_{\mathcal{I},1}U_{\mathcal{I},2}}$ converge. Define the convergence points as:

$$\sigma_{\mathcal{I}} \triangleq \lim_{i\to\infty} \Delta_{\mathcal{I},n_i}, \qquad \mathcal{I} \in \mathcal{P}(2), \quad (87)$$

$$P^*_{\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}U_{\mathcal{I},1}U_{\mathcal{I},2}} \triangleq \lim_{i\to\infty} P^{(n)}_{\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}U_{\mathcal{I},1}U_{\mathcal{I},2}}. \quad (88)$$

Notice next that the right-hand sides of (83) and (84) tend to R_1 and R_2, respectively as $n \to \infty$. Moreover, the terms $\frac{1}{n}\Delta_{\mathcal{I},n} \log \Delta_{\mathcal{I},n}$ on the left-hand sides of these inequalities vanish as $n \to \infty$ because $t \mapsto t \log t$ is bounded over the interval $[0,1]$. Restrict then attention to subsets $\mathcal{I} \in \mathcal{P}(2)$ for which $\sigma_{\mathcal{I}} > 0$, which implies that

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_{\mathcal{I},n} = 0, \quad (89)$$

We denote the set of these subsets $\mathcal{I}$ satisfying (89) by $\mathcal{P} \subseteq \mathcal{P}(2)$, and also notice that inequalities (83) and (84) remain valid if summation is only over sets in $\mathcal{P}$ because all terms are nonnegative. For subsets in $\mathcal{P}$ the last terms in (73) and (74) vanish in the asymptotic regime $n_i \to \infty$.

By all these considerations, by (68), (71)–(74), (83), and (84), in the limit $i \to \infty$:

$$R_1 \geq \sum_{\mathcal{I}\in\mathcal{P}} \sigma_{\mathcal{I}} \cdot I_{P^*_{\mathcal{I}}}(U_{\mathcal{I},1}; Y_{\mathcal{I},0}) \quad (90a)$$

$$R_2 \geq \sum_{\mathcal{I}\in\mathcal{P}:\, 2\in\mathcal{I}} \sigma_{\mathcal{I}} \cdot I_{P^*_{\mathcal{I}}}(U_{\mathcal{I},2}; Y_{\mathcal{I},1}) \quad (90b)$$

$$\theta_1 \leq \min_{\mathcal{I}\in\mathcal{P}} \left\{ I_{P^*_{\mathcal{I}}}(U_{\mathcal{I},1}; Y_{\mathcal{I},1}) \right\} \quad (90c)$$

$$\theta_2 \leq \min_{\mathcal{I}\in\mathcal{P}} \left\{ I_{P^*_{\mathcal{I}}}(U_{\mathcal{I},1}; Y_{\mathcal{I},1}) + I_{P^*_{\mathcal{I}}}(U_{\mathcal{I},2}; Y_{\mathcal{I},2}) \right\}. \quad (90d)$$

for some nonnegative numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{P}}$ satisfying

$$\sum_{\mathcal{I}\in\mathcal{P}} \sigma_{\mathcal{I}} \leq 1 \quad (90e)$$

$$\sum_{\mathcal{I}\in\mathcal{P}:\, 1\in\mathcal{I}} \sigma_{\mathcal{I}} \geq 1 - \epsilon_1, \quad (90f)$$

$$\sigma_{\{1,2\}} \geq \max\{1 - \epsilon_1 - \epsilon_2, 0\}, \quad (90g)$$

$$\sum_{\mathcal{I}\in\mathcal{P}:\, 2\in\mathcal{I}} \sigma_{\mathcal{I}} \geq 1 - \epsilon_2. \quad (90h)$$

Notice further that since for any $\mathcal{I} \in \mathcal{P}(2)$ and any i the triple $(\tilde{Y}^{(n_i)}_{\mathcal{I},0}, \tilde{Y}^{(n_i)}_{\mathcal{I},1}, \tilde{Y}^{(n_i)}_{\mathcal{I},2})$ lies in the jointly typical set $\mathcal{T}^{(n_i)}_{\mu_{n_i}}(P_{Y_0Y_1Y_2})$ we have $|P_{\tilde{Y}_{\mathcal{I},0}\tilde{Y}_{\mathcal{I},1}\tilde{Y}_{\mathcal{I},2}}(y_0,y_1,y_2) - P_{Y_0Y_1Y_2}(y_0,y_1,y_2)| \leq \mu_{n_i}$ and since and $\mu_{n_i} \to 0$ as $i \to \infty$, the limiting pmfs satisfy $P^*_{Y_{\mathcal{I},0}Y_{\mathcal{I},1}Y_{\mathcal{I},2}} = P_{Y_0Y_1Y_2}$. Finally, by continuity considerations, by (89), and by (71) the following Markov chains must hold for all $\mathcal{I} \in \mathcal{P}$ under $P^*_{Y_{\mathcal{I},0}Y_{\mathcal{I},1}Y_{\mathcal{I},2}U_{\mathcal{I},1}U_{\mathcal{I},2}}$:

$$U_{\mathcal{I},1} \to Y_{\mathcal{I},0} \to Y_{\mathcal{I},1}, \quad (91)$$

$$U_{\mathcal{I},2} \to Y_{\mathcal{I},1} \to Y_{\mathcal{I},2}. \quad (92)$$

Notice that at this point and when $\mathcal{P} \subsetneq \mathcal{P}(2)$, one can trivially add dummy random variables (namely $U_{\mathcal{I},j} = Y_{\mathcal{I},j-1}$ combined with $\sigma_{\mathcal{I}} = 0$ for all $\mathcal{I} \in \mathcal{P}(2)\backslash\mathcal{P}$) to extend conditions (90) to all subsets $\mathcal{I} \in \mathcal{P}(2)$. The desired result in Proposition 3 is then obtained by plugging the definitions of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$ into above expressions (90) and by relaxing some constraints (because any element is no larger than the maximum).. ■

VII. A SYSTEM WITH K-HOPS

We generalize our setup and results to K hops, i.e., to $K-1$ relays.

A. System Model

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

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

$$\text{if } \mathcal{H} = 0 : (Y_0^n, Y_1^n, \ldots, Y_K^n) \text{ i.i.d. } \sim P_{Y_0Y_1} \cdot P_{Y_2|Y_1} \cdots P_{Y_K|Y_{K-1}}; \quad (93a)$$

$$\text{if } \mathcal{H} = 1 : (Y_0^n, Y_1^n, \ldots, Y_K^n) \text{ i.i.d. } \sim P_{Y_0} \cdot P_{Y_1} \cdots P_{Y_K}. \quad (93b)$$

Communication takes place over K hops as illustrated in Figure 6. The transmitter T_0 sends a message $\mathsf{M}_1 = \phi_0^{(n)}(Y_0^n)$ to the first relay R_1, which sends a message

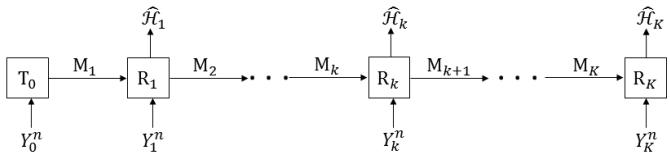

Fig. 6: Cascaded K-hop setup with K decision centers.

$\mathsf{M}_2 = \phi_1^{(n)}(Y_1^n, \mathsf{M}_1)$ to the second relay and so on. The communication is thus described by encoding functions

$$\phi_0^{(n)} \colon \mathcal{Y}_0^n \to \{0,1\}^\star \tag{94}$$

$$\phi_k^{(n)} \colon \mathcal{Y}_k^n \times \{0,1\}^\star \to \{0,1\}^\star, \quad k \in \{1,\dots,K-1\}, \tag{95}$$

so that the produced message strings

$$\mathsf{M}_1 = \phi_0^{(n)}(Y_0^n) \tag{96}$$

$$\mathsf{M}_{k+1} = \phi_k^{(n)}(Y_k^n, \mathsf{M}_k), \quad k \in \{1,\dots,K-1\}, \tag{97}$$

satisfy the expected-rate constraints

$$\mathbb{E}\left[\operatorname{len}(\mathsf{M}_k)\right] \le nR_k, \quad k \in \{1,\dots,K\}. \tag{98}$$

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

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

where we request that the guesses

$$\hat{\mathcal{H}}_{k,n} = g_k^{(n)}(Y_k^n, \mathsf{M}_k), \quad k \in \{1,\dots,K\}, \tag{100}$$

have type-I error probabilities

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

not exceeding given thresholds $\epsilon_1, \epsilon_2, \dots, \epsilon_K > 0$, and type-II error probabilities

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

decaying to 0 exponentially fast with largest possible exponents.

Definition 3: Given maximum type-I error probabilities $\epsilon_1, \epsilon_2, \dots, \epsilon_K \in [0,1)$ and rates $R_1, R_2, \dots, R_K \ge 0$. The exponent tuple $(\theta_1, \theta_2, \dots, \theta_K)$ is called $(\epsilon_1, \epsilon_2, \dots, \epsilon_K)$-*achievable* if there exists a sequence of encoding and decision functions $\left\{\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_{K-1}^{(n)}, g_1^{(n)}, g_2^{(n)}, \cdots g_K^{(n)}\right\}_{n\ge1}$ satisfying for each $k \in \{1,\dots,K\}$:

$$\mathbb{E}[\operatorname{len}(\mathsf{M}_k)] \le nR_k, \tag{103a}$$

$$\varlimsup_{n\to\infty} \alpha_{k,n} \le \epsilon_k, \tag{103b}$$

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

Definition 4: The fundamental exponents region $\mathcal{E}^*(R_1, R_2, \dots, R_K, \epsilon_1, \epsilon_2, \dots, \epsilon_K)$ is defined as the closure of the set of all $(\epsilon_1, \epsilon_2, \dots, \epsilon_K)$-achievable exponent pairs $(\theta_1, \theta_2, \dots, \theta_K)$ for given rates $R_1, \dots, R_K \ge 0$.

B. Previous Results under Maximum-Rate Constraints

The K-hop hypothesis testing setup of Figure 6 and Equations (93) was also considered in [28], but under *maximum-rate constraints*:

$$\operatorname{len}(\mathsf{M}_i) \le nR_i, \qquad i \in \{1,\dots,K\}, \tag{104}$$

instead of the *expected-rate constraints* (98). The fundamental exponents region $\mathcal{E}^*_{\max}(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$ for this maximum-rate setup is defined analogously to Definition 4, but with (98) replaced by (104).

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

$$\eta_\ell \colon \mathbb{R}_0^+ \to \mathbb{R}_0^+ \tag{105}$$

$$R \mapsto \max_{\substack{P_{U|Y_{\ell-1}}:\\ R \ge I(U;Y_{\ell-1})}} I\left(U; Y_\ell\right). \tag{106}$$

The functions $\eta_1, \dots, \eta_K$ are concave and monotonically non-decreasing. The proof is analogous to the proof of Lemma 1 presented in Appendix A, and omitted for brevity. Notice further that in the maximization determining $\eta_\ell(R)$ it suffices to consider distributions $P_{U|Y_{\ell-1}}$ on alphabets of sizes $|\mathcal{Y}_{\ell-1}| + 1$, see [3].

Theorem 4 ([41]): Given $(\epsilon_1, \dots, \epsilon_K) \in [0,1)^K$, the fundamental exponents region under the maximum-rate constraints (104) and vanishing type-I error constraints satisfies

$$\mathcal{E}^*_{\max}(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K) = \left\{ (\theta_1, \dots, \theta_K) \colon \theta_k \le \sum_{\ell=1}^{k} \eta_\ell(R_\ell),\ k \in \{1,\dots,K\} \right\} \tag{107}$$

Notice that in this K-hop setup, each decision center accumulates all the error exponents on the various links from the transmitter to this decision center. The fundamental exponents region is thus given by a K-dimensional hyperrectangle. That means, each decision center can simultaneously achieve its optimal error exponent as if the other decision centers were not present in the system.

We abbreviate $\mathcal{E}^*_{\max}(R_1, \dots, R_K, \epsilon_1, \dots, \epsilon_K)$ by $\mathcal{E}^*_{\max}(R_1, \dots, R_K)$.

C. Optimal Coding Scheme for K Hops under Expected-Rate Constraints

Similarly to the two-hop scheme, the terminals multiplex different subschemes depending on the sequence Y_0^n observed at the transmitter T_0. To this end, partition the set $\mathcal{Y}_0^n$ into disjoint subsets $\mathcal{D}_\emptyset$ and $\{\mathcal{D}_\mathcal{I}\}_{\mathcal{I}\in\mathcal{P}(K)}$ so that the probabilities

$$\sigma_\mathcal{I} := \Pr[Y_0^n \in \mathcal{D}_\mathcal{I}] \tag{108}$$

satisfy

$$1 - \sum_{k\in\mathcal{S}} \epsilon_k \le \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ \mathcal{S}\subseteq\mathcal{I}}} \sigma_\mathcal{I}, \quad \mathcal{S} \subseteq \{1,\dots,K\}, \tag{109a}$$

$$\sum_{\mathcal{I}\in\mathcal{P}(K)} \sigma_\mathcal{I} \le 1. \tag{109b}$$

In our multiplexed schemes, the index $\mathcal{I}$ of $\mathcal{D}_\mathcal{I}$ indicates that if T_0's observation Y_0^n lies in $\mathcal{D}_\mathcal{I}$, then all decision centers

R_k, for $k \in \mathcal{I}$, attempt to correctly guess hypothesis $\mathcal{H}$, while all decision centers R_k, for $k \notin \mathcal{I}$, simply declare $\hat{\mathcal{H}}_k = 1$. If $Y_0^n \in \mathcal{D}_\emptyset$, then *all* decision centers $\mathrm{R}_1, \ldots, \mathrm{R}_K$ simply declare $\hat{\mathcal{H}} = 1$.

The transmitter T_0 adds K flag-bits to its message M_1 to inform R_1 about the set $\mathcal{D}_\mathcal{I}$ containing its observation Y_0^n, and thus about the choice of the employed coding scheme. These flag-bits are forwarded by all relays $\mathrm{R}_1, \ldots, \mathrm{R}_{K-1}$ at the beginning of their messages $\mathsf{M}_2, \ldots, \mathsf{M}_K$ so as to pass the information to all terminals in the network.

We describe the different multiplexed coding schemes in more detail. Let $\ell_\mathcal{I}^*$ be the largest index in set $\mathcal{I}$:

$$\ell_\mathcal{I}^* := \max_{k \in \mathcal{I}} k, \tag{110}$$

and chooses a set of rates

$$\{R_{\mathcal{I},\ell} \colon \quad \mathcal{I} \in \mathcal{P}(K),\ \ell \in \{1, \ldots, \ell_\mathcal{I}^*\}\} \tag{111}$$

satisfying

$$R_\ell > \sum_{\substack{\mathcal{I} \in \mathcal{P}(K)\colon \\ \ell_\mathcal{I}^* \geq \ell}} \sigma_\mathcal{I} \cdot R_{\mathcal{I},\ell}, \quad \ell \in \{1, \ldots, K\}. \tag{112}$$

We will see that the choice of the various rates determines the tradeoff between the different exponents $\theta_1, \ldots, \theta_K$. Rates $\{R_{\mathcal{I}_\ell} \colon \ell \in \{1, \ldots, \ell_\mathcal{I}^*\}\}$ are used in the subscheme employed when $Y_0^n \in \mathcal{D}_\mathcal{I}$, where under this event only the messages on the first $\ell_\mathcal{I}^*$ links have positive rates, while messages on the last $K - \mathcal{I}_\ell^*$ links are of zero rate. The reason is that decision center $\mathrm{R}_{\mathcal{I}_\ell^*+1}, \ldots, \mathrm{R}_K$ simply declare $\hat{\mathcal{H}} = 1$ and thus messages $\mathsf{M}_{\ell_\mathcal{I}^*+1}, \ldots, \mathsf{M}_K$ only have to convey the zero-rate information that $Y_0^n \in \mathcal{D}_\mathcal{I}$.

Subscheme for $Y_0^n \in \mathcal{D}_\emptyset$: All terminals T_0 and $\mathrm{R}_1, \ldots, \mathrm{R}_{K-1}$ send the length-K all-zero bit string over the respective communication links:

$$\mathsf{M}_1 = \cdots = \mathsf{M}_K = [0, 0, \ldots, 0]. \tag{113}$$

Upon receiving this all-zero flag, relays $\mathrm{R}_1, \ldots, \mathrm{R}_{K-1}$ and receiver R_K all declare

$$\hat{\mathcal{H}}_1 = \cdots = \hat{\mathcal{H}}_K = 1. \tag{114}$$

Communication is thus only used to inform the relays and the receiver about the scheme to employ, or equivalently the event $Y_0^n \in \mathcal{D}_\emptyset$, without providing any further information about the correct hypothesis.

Subscheme for $Y_0^n \in \mathcal{D}_\mathcal{I}$, for $\mathcal{I} \in \mathcal{P}(K)$: In this case, only decision centers R_k, for $k \in \mathcal{I}$, attempt to correctly guess hypothesis $\mathcal{H}$; all other decision centers R_k, for $k \notin \mathcal{I}$, directly declare $\hat{\mathcal{H}}_k = 1$.

Terminals $\mathrm{T}_0, \mathrm{R}_1, \ldots, \mathrm{R}_{\ell_\mathcal{I}^*}$ apply a given $\ell_\mathcal{I}^*$-hop hypothesis testing scheme with vanishing type-I error probabilities and respecting the maximum-rate constraints $R_{\mathcal{I},1}, \ldots, R_{\mathcal{I},\ell_\mathcal{I}^*}$ on the first $\ell_\mathcal{I}^*$ links. To inform all relays and the receiver about the scheme to use, terminals $\mathrm{T}_0, \mathrm{R}_1, \ldots, \mathrm{R}_{K-1}$ append a K-length flag sequence describing set $\mathcal{I}$ at the beginning of their messages. We propose that this flag sequence shows bit 1 at all positions $k \in \mathcal{I}$ and bit 0 at all positions $k \notin \mathcal{I}$. Notice that Messages $\mathsf{M}_{\ell_\mathcal{I}^*+1}, \ldots, \mathsf{M}_K$ consist of only the flag sequence.

All decision centers R_k with $k \in \mathcal{I}$ declare the hypothesis indicated by the employed multi-hop hypothesis testing scheme. The remaining decision centers R_k with $k \notin \mathcal{I}$ simply declare

$$\hat{\mathcal{H}}_k = 1, \quad k \notin \mathcal{I}. \tag{115}$$

Analysis: By (108) and (112), and because transmission of K bits hardly changes the rate for sufficiently large block-lengths, the proposed overall scheme respects the expected-rate constraints $R_1, \ldots, R_K$ on the K links for large values of n. Appendix E proves that when the optimal multi-hop hypothesis testing schemes with vanishing type-I error probability [28] are used as the various subschemes, then the overall scheme satisfies the type-I error constraints $\epsilon_1, \ldots, \epsilon_K$ and achieves the error exponents in the following Theorem 5.

D. Results on the Exponents Region

Theorem 5: The fundamental exponents region $\mathcal{E}^*(R_1, \ldots, R_K, \epsilon_1, \ldots, \epsilon_K)$ is equal to the set of all nonnegative tuples $(\theta_1, \ldots, \theta_K)$ satisfying

$$\theta_k \leq \min_{\substack{\mathcal{I} \in \mathcal{P}(K)\colon \\ k \in \mathcal{I}}} \sum_{\ell=1}^{k} \eta_\ell(R_{\mathcal{I},\ell}), \tag{116a}$$

for some nonnegative rates $\{R_{\mathcal{I},1}, \ldots, R_{\mathcal{I},\ell_\mathcal{I}^*}\}_{\mathcal{I} \in \mathcal{P}(K)}$ and nonnegative numbers $\{\sigma_\mathcal{I}\}_{\mathcal{I} \in \mathcal{P}(K)}$ satisfying

$$R_k \geq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K)\colon \\ k \leq \ell_\mathcal{I}^*}} \sigma_\mathcal{I} R_{\mathcal{I},k}, \quad k \in \{1, \ldots, K\}, \tag{116b}$$

$$\max\left\{0, 1 - \sum_{k \in \mathcal{S}} \epsilon_k\right\} \leq \sum_{\substack{\mathcal{I} \in \mathcal{P}(K)\colon \\ \mathcal{S} \subseteq \mathcal{I}}} \sigma_\mathcal{I}, \quad \mathcal{S} \subseteq \{1, \ldots, K\}, \tag{116c}$$

$$\sum_{\mathcal{I} \in \mathcal{P}(K)} \sigma_\mathcal{I} \leq 1. \tag{116d}$$

Proof: Achievability is based on the coding scheme presented in the previous subsection and analyzed in Appendix E. The converse is proved in the next Section VIII. ■

We conjecture that similarly to the case of $K = 2$ hops, the optimal exponents region $\mathcal{E}^*(R_1, \ldots, R_K, \epsilon_1, \ldots, \epsilon_K)$ in Theorem 5 can be simplified depending on the ordering of the ϵ-values. To state our conjecture, we define a permutation $\pi \colon \{1, \ldots, K\} \to \{1, \ldots, K\}$ that orders the ϵ-values in decreasing order:

$$\epsilon_{\pi(1)} \geq \epsilon_{\pi(2)} \geq \cdots \geq \epsilon_{\pi(K)}, \tag{117}$$

and sets $\epsilon_{\pi(0)} := 1$. We conjecture then that in Theorem 5 without loss of optimality one can set

$$\sigma_{\{\pi(i), \cdots, \pi(K)\}} = \epsilon_{\pi(i-1)} - \epsilon_{\pi(i)}, \quad i \in \{1, \ldots, K\}, \tag{118}$$

and all other σ-values to 0. Renaming rates $R_{\{\pi(i), \cdots, \pi(K)\},\ell}$ to $R_{i,\ell}$ and $\ell_{\{\pi(i), \cdots, \pi(K)\}}^*$ to ℓ_i^*, with the choice in (118) Theorem 5 evaluates to the following region:

Conjecture 6: The fundamental exponents region $\mathcal{E}^*(R_1,\dots,R_K,\epsilon_1,\dots,\epsilon_K)$ is the set of all exponent tuples $(\theta_1,\dots,,\theta_K)$ that satisfy

$$\theta_k \leq \min_{i\in\{1,\dots,\pi(k)\}} \left[\sum_{\ell=1}^{k} \eta_\ell\left(R_{i,\ell}\right)\right], \quad k\in\{1,\dots,K\}, \tag{119a}$$

for some nonnegative rates $\{R_{i,\ell}\}$ satisfying

$$R_\ell \geq \sum_{\substack{i\in\{1,\dots,K\}:\\ \ell_i^*\geq \ell}} \left(\epsilon_{\pi(i-1)}-\epsilon_{\pi(i)}\right) R_{i,\ell}, \quad \ell\in\{1,\dots,K\}, \tag{119b}$$

where

$$\ell_i^* := \max_\ell\{\ell\colon \ell\in\{\pi(i),\dots,\pi(K)\}\}. \tag{119c}$$

The importance of this conjecture lies in proving that the optimal coding scheme in the previous Subsection VII-C only has to multiplex $K+1$ subschemes, where the choice of the subschemes depends on the permutation π that orders the ϵ-values. Moreover, no optimization over the probabilities of the various subschemes is required anymore, as the probability of each subscheme is determined by the ϵ-values and the ordering permutation π. More specifically, if the conjecture is correct, the optimal scheme multiplexes $K+1$ schemes, where the i-th scheme is applied with probability $\epsilon_{\pi(i-1)}-\epsilon_{\pi(i)}$ and serves only the DCs with $(K-i+1)$-th smallest type-I error constraints, while all other DCs directly declare $\mathcal{H}=1$.

While a general proof of the conjecture is still missing, we can prove it in two special cases.

Proposition 7: Conjecture 6 holds when

$$K\in\{2,3\} \tag{120}$$

or when

$$\epsilon_1=\dots=\epsilon_K. \tag{121}$$

In particular, for $\epsilon_1=\ldots=\epsilon_K=\epsilon$, the exponents region $\mathcal{E}^*(R_1,\dots,R_K,\epsilon,\dots,\epsilon)$ is the set of all nonnegative tuples $(\theta_1,\dots,\theta_K)$ satisfying

$$\theta_k \leq \sum_{\ell=1}^{k}\eta_\ell\left(\frac{R_\ell}{1-\epsilon}\right), \quad k\in\{1,\dots,K\}. \tag{122a}$$

Proof: Achievability of the region in (119) for any value of K follows by specializing the region in Theorem 5 to the parameter choice in (118) and setting all other σ-values to 0, and by renaming rates $R_{\{\pi(i),\cdots,\pi(K)\},\ell}$ as $R_{i,\ell}$. For $K=2$ the result recovers Theorem 2. The converse for $K=3$ is proved in Appendix G. The converse for $\epsilon_1=\dots=\epsilon_K$ is proved in Appendix H. ■

As we discuss next, similar observations apply for the general case $K\geq 2$ as we have presented for $K=2$ in Remarks 2–5.

Remark 7 (Discussion for $\epsilon_1=\dots=\epsilon_K=$): For equal type-I error probabilities $\epsilon_1=\dots=\epsilon_K=\epsilon$, there is a rate-boost on each link of $(1-\epsilon)^{-1}$ compared to the scenario with maximum-rate constraints.

Remark 8 (Discussion for the general case): In the general case, irrespective of the ordering of the permissible type-I error probabilities, the largest exponent achievable at a decision center k is given by

$$\theta_{k,\max} := \sum_{\ell=1}^{k}\eta_\ell\left(\frac{R_\ell}{1-\epsilon_k}\right), \tag{123}$$

so that for this exponent all rates are boosted by $(1-\epsilon_k)^{-1}$. In fact, $\theta_k=\theta_{k,\max}$ is achieved by choosing the first k rates as:[3]

$$R_{\mathcal{I},\ell}=\frac{R_\ell}{1-\epsilon_k}, \quad k\in\mathcal{I},\ \ell\in\{1,\dots,k\}. \tag{124}$$

This choice imposes that $\sigma_{\mathcal{I}}R_{\mathcal{I},\ell}=0$ for all $\mathcal{I}$ not containing k and all $\ell\in\{1,\dots,k\}$. As a consequence, the optimal performance for a decision center $\mathrm{R}_{k'}$, for $k'<k$, is

$$\theta_{k'}=\sum_{\ell=1}^{k'}\eta_\ell\left(\frac{R_\ell}{1-\epsilon_k}\right), \quad \text{if } \epsilon_{k'}>\epsilon_k \tag{125}$$

$$\theta_{k'}=0, \quad \text{if } \epsilon_{k'}<\epsilon_k, \tag{126}$$

where the performance in (125) is obtained by setting $\sigma_{\mathcal{I}}=0$ for all $\mathcal{I}$ containing an index $k'<k$ with $\epsilon_k'>\epsilon_k$ and by setting the corresponding rates to infinity. Notice that $\sigma_{\mathcal{I}}$ cannot be chosen equal to 0 for all sets $\mathcal{I}$ containing index $k'<k$ when $\epsilon_{k'}<\epsilon_k$ because Constraint (116c) implies that at least one of these σ-values is positive, which by $\sigma_{\mathcal{I}}R_{\mathcal{I},\ell}=0$ implies that the corresponding rates $R_{\mathcal{I},\ell}=0$, for all $\ell=1,\dots,k$, causing $\theta_{k'}$ to degrade to 0. We conclude that under (123), for any $k'<k$, when $\epsilon_k'\geq\epsilon_k$ then exponent $\theta_{k'}$ is degraded from its maximum value because all rates are only boosted by the factor $(1-\epsilon_k)^{-1}$ and not by the larger factor $(1-\epsilon_{k'})^{-1}$, and when $\epsilon_k'<\epsilon_k$ the exponent $\theta_{k'}$ completely degrades to 0.

With appropriate choices for the rates on the last $(K-k)$ links, different tradeoffs between the exponents $\theta_{k+1},\dots,\theta_K$ can be achieved. In particular, it is possible that an exponent $\theta_{k'}$, for $k'>k$, experiences its maximum rate-boost $(1-\epsilon_{k'})^{-1}$ on some of these links. On the first k links, any exponent $\theta_{k+1},\dots,\theta_K$ experiences a rate-boost of $(1-\epsilon_k)^{-1}$ if the corresponding $\epsilon_{k'}>\epsilon_k$, whereas the contributions of the first k links completely degrade to 0 if $\epsilon_{k'}<\epsilon_k$.

The following lemma indicates that in the evaluation of the fundamental exponents region $\mathcal{E}^*(R_1,\dots,R_K,\epsilon_1,\dots,\epsilon_K)$ in Theorem 5 one can restrict to sets of parameters $\{\sigma_{\mathcal{I}}\}$ that satisfy some of the constraints (116c) with equality and set certain σ-values to 0. It is a first step towards proving Conjecture 6 in the general case or at least towards simplifying the evaluation of the exponents region in Theorem 5.

[3]This choice assumes that the ordering (117) is strict, i.e., no two ϵ-values coincide. Moreover, when some of the available rates $R_1,\dots,R_k$ are sufficiently large so as to saturate the functions $\eta_\ell(R_\ell)$, then other choices are possible.

Lemma 3: Consider a set of nonnegative numbers $\{R_{\mathcal{I},1},\dots,R_{\mathcal{I},\ell_{\mathcal{I}}^*}\}_{\mathcal{I}\in\mathcal{P}(K)}$ and $\{\sigma_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{P}(K)}$ satisfying (116) for exponents $(\theta_1,\dots,\theta_K)$. Let $\mathcal{I}',\mathcal{I}''\in\mathcal{P}(K)$ and $\Gamma\in[0,\sigma_{\mathcal{I}''}]$ be so that

$$\mathcal{I}'\subseteq\mathcal{I}'' \tag{127}$$

and

$$\max\left\{0,1-\sum_{k\in\mathcal{S}}\epsilon_k\right\}+\Gamma\leq\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ \mathcal{S}\subseteq\mathcal{I}}}\sigma_{\mathcal{I}},\quad \mathcal{S}\subseteq\mathcal{I}'',\mathcal{S}\nsubseteq\mathcal{I}'. \tag{128}$$

Then, the new nonnegative numbers

$$\tilde{\sigma}_{\mathcal{I}'}=\sigma_{\mathcal{I}'}+\Gamma \tag{129}$$

$$\tilde{\sigma}_{\mathcal{I}''}=\sigma_{\mathcal{I}''}-\Gamma \tag{130}$$

$$\tilde{\sigma}_{\mathcal{I}}=\sigma_{\mathcal{I}},\quad \mathcal{I}\in\mathcal{P}(K)\backslash\{\mathcal{I}',\mathcal{I}''\}, \tag{131}$$

and rates, for $\ell\in\{1,\dots,K\}$,

$$\tilde{R}_{\mathcal{I}',\ell}=\frac{\sigma_{\mathcal{I}'}\cdot R_{\mathcal{I}',\ell}+\Gamma\cdot R_{\mathcal{I}'',\ell}}{\tilde{\sigma}_{\mathcal{I}'}}, \tag{132}$$

$$\tilde{R}_{\mathcal{I},\ell}=R_{\mathcal{I},\ell},\quad \mathcal{I}\in\mathcal{P}(K)\backslash\{\mathcal{I}'\}. \tag{133}$$

also satisfy (116) for exponents $(\theta_1,\dots,\theta_K)$.

Proof: Above rate-definitions essentially only shift the term $\Gamma\cdot R_{\mathcal{I}'',\ell}$ from $\sigma_{\mathcal{I}''}R_{\mathcal{I}'',\ell}$ to $\tilde{\sigma}_{\mathcal{I}'}R_{\mathcal{I}',\ell}$, and therefore the rate constraints (116b) remain valid also for the new numbers. Similarly, constraint (116d) remains valid since the sum of all σ-values is preserved. Notice further that the σ-values included in Constraint (116c) for $\mathcal{S}\nsubseteq\mathcal{I}''$ remain unchanged by (131) and for $\mathcal{S}\subseteq\mathcal{I}'$ their sum is preserved by (129) and (130). For $\mathcal{S}\nsubseteq\mathcal{I}'$ but $\mathcal{S}\subseteq\mathcal{I}''$, Constraint (116c) is satisfied by Assumption (128). It remains to check the validity of (116a) for the new rate-values. By (133) the constraint remains unchanged for all $k\notin\mathcal{I}'$. For $k\in\mathcal{I}'$, we notice that by (127) the minimum in (116a) includes both sets $\mathcal{I}'$ and $\mathcal{I}''$ and this minimum cannot be smaller for the new rates because:

$$\begin{aligned}&\min\left\{\sum_{\ell=1}^{k}\eta_\ell\left(R_{\mathcal{I}',\ell}\right),\ \sum_{\ell=1}^{k}\eta_\ell\left(R_{\mathcal{I}'',\ell}\right)\right\}\\ &\leq\min\left\{\sum_{\ell=1}^{k}\left(\frac{\sigma_{\mathcal{I}'}}{\tilde{\sigma}_{\mathcal{I}'}}\eta_\ell\left(R_{\mathcal{I}',\ell}\right)+\frac{\Gamma}{\tilde{\sigma}_{\mathcal{I}'}}\eta_\ell\left(R_{\mathcal{I}'',\ell}\right)\right),\right.\\ &\qquad\left.\sum_{\ell=1}^{k}\eta_\ell\left(R_{\mathcal{I}'',\ell}\right)\right\} \qquad (134)\\ &\leq\min\left\{\sum_{\ell=1}^{k}\eta_\ell\left(\tilde{R}_{\mathcal{I}',\ell}\right),\ \sum_{\ell=1}^{k}\eta_\ell\left(\tilde{R}_{\mathcal{I}'',\ell}\right)\right\}, \qquad (135)\end{aligned}$$

where the first inequality holds because the minimum of two numbers cannot exceed any convex combination of the numbers, and the second inequality holds by the concavity and monotonicity of the functions $\{\eta_\ell(\cdot)\}_\ell$. ■

VIII. Converse Proof to Theorem 5

A. Proof Outline

The converse proof is similar as for $K=2$ users. We partition the strongly typical set $\mathcal{T}_\mu(P_{Y_0\cdots Y_K})$ into 2^K subsets according to the decisions taken at the various DCs. A change of measure argument is then applied in parallel to each of these subsets and based on this change of measure, by applying Lemma 4 ahead, bounds on the conditionally expected message lengths and on the overall type-II error exponent are derived, as well as a bound on the proximity of the new measure with some Markov pmf. The desired converse proof can then be derived by combining the bounds for the different subsets (either through the law of total expectation or simply by considering the most stringent one) and by proving that there exists a subsequence of blocklengths for which the changed measure on each set converges to a distribution whose $(Y_0,\dots,Y_K)$-marginal equals $P_{Y_0\cdots Y_K}$ and satisfies desired Markov properties with the auxiliary random variables defined by Lemma 4.

B. Converse Proof

Fix an exponent-tuple $(\theta_1,\dots,\theta_K)$ in the exponents region $\mathcal{E}^*(R_1,\dots,R_K,\epsilon_1,\dots,\epsilon_K)$, and a sequence (in n) of encoding and decision functions $\{(\phi_0^{(n)},\phi_1^{(n)},\dots,\phi_K^{(n)},g_1^{(n)},\dots,g_K^{(n)})\}_{n\geq1}$ achieving this tuple, i.e., satisfying constraints (103).

Our proof relies on the following lemma:

Lemma 4: Fix a blocklength n and a set $\mathcal{D}\subseteq\mathcal{Y}_0^n\times\mathcal{Y}_1^n\times\cdots\times\mathcal{Y}_K^n$ of positive probability, and let the tuple $(\tilde{\mathsf{M}}_1,\tilde{\mathsf{M}}_2,\dots,\tilde{\mathsf{M}}_K,\tilde{Y}_0^n,\tilde{Y}_1^n,\dots,\tilde{Y}_K^n)$ follow the pmf

$$\begin{aligned}&P_{\tilde{\mathsf{M}}_1\tilde{\mathsf{M}}_2\cdots\tilde{\mathsf{M}}_K\tilde{Y}_0^n\tilde{Y}_1^n\cdots\tilde{Y}_K^n}(\mathsf{m}_1,\mathsf{m}_2,\dots,\mathsf{m}_K,y_0^n,y_1^n,\dots,y_K^n)\triangleq\\ &\quad P_{Y_0^nY_1^n\cdots Y_K^n}(y_0^n,y_1^n,\dots,y_K^n)\cdot\frac{\mathbb{1}\{(y_0^n,y_1^n,\dots,y_K^n)\in\mathcal{D}\}}{P_{Y_0^nY_1^n\cdots Y_K^n}(\mathcal{D})}\\ &\quad\cdot\mathbb{1}\{\phi_1(y_0^n)=\mathsf{m}_1\}\cdot\mathbb{1}\{\phi_2(y_1^n,\phi_1(y_0^n))=\mathsf{m}_2\}\cdot\cdots\\ &\quad\cdot\mathbb{1}\{\phi_K(y_{K-1}^n,\phi_{K-1}(y_{K-2}^n,\phi_{K-2}(\cdots,\phi_1(y_0^n)))=\mathsf{m}_K\}.\end{aligned} \tag{136}$$

Further, define the auxiliary random variables

$$U_k\triangleq(\tilde{\mathsf{M}}_k,\tilde{Y}_0^{T-1}\tilde{Y}_1^{T-1},\dots,\tilde{Y}_K^{T-1},T),\qquad k\in\{1,\dots,K\}, \tag{137}$$

$$\tilde{Y}_k\triangleq\tilde{Y}_{k,T},\quad k\in\{0,1,\dots,K\}, \tag{138}$$

where T is uniform over $\{1,\dots,n\}$ and independent of the tuple $(\tilde{\mathsf{M}}_1,\tilde{\mathsf{M}}_2,\dots,\tilde{\mathsf{M}}_K,\tilde{Y}_0^n,\tilde{Y}_1^n,\dots,\tilde{Y}_K^n)$.

For any $k\in\{1,\dots,K\}$ the following (in)equalities hold:

$$\frac{1}{n}H(\tilde{\mathsf{M}}_k)\geq I(U_k;\tilde{Y}_{k-1})+\frac{1}{n}\log P_{Y_0^nY_1^n\cdots Y_K^n}(\mathcal{D}), \tag{139}$$

$$I(U_k;\tilde{Y}_k|\tilde{Y}_{k-1})\leq-\frac{1}{n}\log P_{Y_0^nY_1^n\cdots Y_K^n}(\mathcal{D})+D(P_{\tilde{Y}_0\cdots\tilde{Y}_K}\|P_{Y_0\cdots Y_K}). \tag{140}$$

If further for some $k\in\{1,\dots,K\}$, decision center R_k decides on the null hypothesis for all tuples $(y_0^n,\dots,y_K^n)\in\mathcal{D}$[4],

$$\hat{\mathcal{H}}_k=0, \tag{141}$$

[4]Notice that once we fix the realizations of all observed sequences $Y_0^n,\dots,Y_K^n$, the decision $\hat{\mathcal{H}}_k$ is either determinstically 0 or 1.

then

$$\beta_{k,n} \leq \sum_{\ell=1}^{k} I(U_\ell; \tilde{Y}_\ell) + \frac{(k+1)}{n} \log P_{Y_0^n Y_1^n \cdots Y_K^n}(\mathcal{D}). \quad (142)$$

Proof: See Subsection VIII-C at the end of this section. ■

We continue to prove Theorem 5. Let μ_n be a sequence satisfying

$$\lim_{n\to\infty} \mu_n = 0 \quad (143)$$

$$\lim_{n\to\infty} \mu_n n^2 = \infty. \quad (144)$$

For each blocklength n, define for each index $k \in \{1, \ldots, K\}$ the set

$$\mathcal{B}_{k,n} \triangleq \{(y_0^n, \ldots, y_K^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 \cdots Y_K}) \colon \quad \hat{\mathcal{H}}_k = 0\} \quad (145)$$

and for each subset $\mathcal{I} \in \mathcal{P}(K)$ the set

$$\mathcal{D}_{\mathcal{I},n} \triangleq \Big\{(y_0^n, \ldots, y_K^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0 \cdots Y_K}) \colon \hat{\mathcal{H}}_k = 0 \quad \forall k \in \mathcal{I} \quad \text{and} \quad \hat{\mathcal{H}}_k = 1 \quad \forall k \notin \mathcal{I}\Big\}. \quad (146)$$

Notice that the sets $\{\mathcal{D}_{\mathcal{I},n}\}_{\mathcal{I}}$ are disjoint and

$$\bigcup_{\substack{\mathcal{I}\in\mathcal{P}(K)\colon \\ k\in\mathcal{I}}} \mathcal{D}_{\mathcal{I},n} = \mathcal{B}_{k,n}. \quad (147)$$

Moreover, by [42, Remark to Lemma 2.12] and the type-I error probability constraints in (103b), for any $k \in \{1, \ldots, K\}$:

$$P_{Y_0^n Y_1^n \cdots Y_K^n}(\mathcal{B}_{k,n}) \geq 1 - \epsilon_k - \frac{|\mathcal{Y}_0| \cdots |\mathcal{Y}_K|}{4\mu_n^2 n}. \quad (148)$$

Defining

$$\Delta_{\mathcal{I},n} := P_{Y_0^n Y_1^n \cdots Y_K^n}(\mathcal{D}_{\mathcal{I},n}), \quad (149)$$

we conclude by (147), by standard laws of probability, and the disjointness of the sets $\{\mathcal{D}_{\mathcal{I},n}\}_{\mathcal{I}}$, that in the limit as $n \to \infty$, for any subset $\mathcal{S} \subseteq \{1, \ldots, K\}$:

$$\varliminf_{n\to\infty} \sum_{\substack{\mathcal{I}\in\mathcal{P}(K)\colon \\ \mathcal{S}\subseteq\mathcal{I}}} \Delta_{\mathcal{I},n} \geq \max\left\{1 - \sum_{k\in\mathcal{S}} \epsilon_k, 0\right\}. \quad (150)$$

We now assume that for every blocklength n and every subset $\mathcal{I} \in \mathcal{P}(K)$ we have $\Delta_{\mathcal{I},n} > 0$. (Otherwise we restrict to an appropriate subsequence of blocklengths or eliminate sets $\mathcal{I} \in \mathcal{P}(K)$ alltogether.) By Lemma 2 we can then conclude that for any $\mathcal{I}$: for any $k \in \{1, \ldots, K\}$ the (in)equalities

$$\frac{1}{n} H(\tilde{\mathsf{M}}_{\mathcal{I},k}) \geq I(U_{\mathcal{I},k}; \tilde{Y}_{k-1}) + \frac{1}{n} \log \Delta_{\mathcal{I},n}, \quad (151a)$$

$$I(U_{\mathcal{I},k}; \tilde{Y}_{\mathcal{I},k} | \tilde{Y}_{\mathcal{I},k-1}) = -\frac{1}{n} \log \Delta_{\mathcal{I},n} + D(P_{\tilde{Y}_0 \cdots \tilde{Y}_K} \| P_{Y_0 \cdots Y_K}) \quad (151b)$$

where $\{U_{\mathcal{I},1}, \ldots, U_{\mathcal{I},\ell_{\mathcal{I}}^*}, \tilde{\mathsf{M}}_{\mathcal{I},1}, \tilde{\mathsf{M}}_{\mathcal{I},2}, \ldots, \tilde{\mathsf{M}}_{\mathcal{I},\ell_{\mathcal{I}}^*}, \tilde{Y}_{\mathcal{I},0}^n, \tilde{Y}_{\mathcal{I},1}^n, \ldots, \tilde{Y}_{\mathcal{I},K}^n\}$ are defined in the lemma. Moreover, for indices $k \in \mathcal{I}$:

$$-\frac{1}{n} \log \beta_{k,n} \leq \sum_{\ell=1}^{k} I(U_{\mathcal{I},\ell}; \tilde{Y}_{\mathcal{I},\ell}) + \frac{(k+1)}{n} \log \Delta_{\mathcal{I},n}. \quad (151c)$$

We define the following random variables for $\mathcal{I} \in \mathcal{P}(K)$ and $k \in \{1, \ldots, \ell_{\mathcal{I}}^*\}$:

$$\tilde{L}_{\mathcal{I},k} \triangleq \operatorname{len}(\tilde{\mathsf{M}}_{\mathcal{I},k}). \quad (152)$$

By the rate constraints (98) and the total law of expectations:

$$nR_k \geq \sum_{\substack{\mathcal{I}\in\mathcal{P}(K)\colon \\ \ell_{\mathcal{I}}^*\geq k}} \mathbb{E}[\tilde{L}_{\mathcal{I},k}] \Delta_{\mathcal{I},n}, \quad (153)$$

and, similarly to (84), we obtain with (151a) and the nonnegativity of entropy:

$$\sum_{\substack{\mathcal{I}\in\mathcal{P}(K)\colon \\ \ell_{\mathcal{I}}^*\geq k}} \Delta_{\mathcal{I},n} \left[I(U_{\mathcal{I},k}; \tilde{Y}_{k-1}) + \frac{1}{n} \log \Delta_{\mathcal{I},n} \right]^+ \leq R_k \left(1 + \sum_{\substack{\mathcal{I}\in\mathcal{P}(K)\colon \\ \ell_{\mathcal{I}}^*\geq k}} h_b\left(\frac{\Delta_{\mathcal{I},n}}{nR_k}\right) \right). \quad (154)$$

The desired converse can then be concluded based on (151b), (151c), and (154), in a similar as we did for $K = 2$. Details are omitted.

C. Proof of Lemma 4

Note first that by (136):

$$D(P_{\tilde{Y}_0^n \cdots \tilde{Y}_K^n} \| P_{Y_0 \cdots Y_K}^n) = -\log \Delta \quad (155)$$

$$P_{\tilde{Y}_k^n}(y_k^n) \leq P_{Y_k^n}(y_k^n), \quad k \in \{1, \ldots, K\},\ y_k^n \in \mathcal{Y}_k^n, \quad (156)$$

where we defined $\Delta \triangleq P_{Y_0^n \cdots Y_K^n}(\mathcal{D})$. Further define $\tilde{U}_{k,t} \triangleq (\tilde{\mathsf{M}}_k, \tilde{Y}_0^{t-1}, \ldots, \tilde{Y}_K^{t-1})$ for $k \in \{1, \ldots, K\}$.

Proof of (139)*:*

Similarly to [34] we introduce the divergence $D(P_{\tilde{Y}_0^n \cdots \tilde{Y}_K^n} \| P_{Y_0 \cdots Y_K}^n)$, which is bounded by $\log \Delta$ as help in the single-letterization of $H(\tilde{\mathsf{M}}_k)$. For each $i \in \{1, \ldots, K\}$, we have:

$$\begin{aligned}
H(\tilde{\mathsf{M}}_k) &\geq I(\tilde{\mathsf{M}}_k; \tilde{Y}_0^n \cdots \tilde{Y}_K^n) + D(P_{\tilde{Y}_0^n \cdots \tilde{Y}_K^n} \| P_{Y_0 \cdots Y_K}^n) + \log \Delta && (157)\\
&= H(\tilde{Y}_0^n \cdots \tilde{Y}_K^n) + D(P_{\tilde{Y}_0^n \cdots \tilde{Y}_K^n} \| P_{Y_0 \cdots Y_K}^n) - H(\tilde{Y}_0^n \cdots \tilde{Y}_K^n | \tilde{\mathsf{M}}_k) + \log \Delta && (158)\\
&\geq n[H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}) + D(P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}} \| P_{Y_0 \cdots Y_K})] - \sum_{t=1}^{n} H(\tilde{Y}_{0,t} \cdots \tilde{Y}_{K,t} | \tilde{U}_{k,t}) + \log \Delta && (159)\\
&= n[H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}) + D(P_{\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}} \| P_{Y_0 \cdots Y_K}) - H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T} | \tilde{U}_{k,T}, T)] + \log \Delta && (160)\\
&\geq n[H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T}) - H(\tilde{Y}_{0,T} \cdots \tilde{Y}_{K,T} | \tilde{U}_{k,T}, T)] + \log \Delta && (161)\\
&= n[I(\tilde{Y}_0 \cdots \tilde{Y}_K; U_k)] + \log \Delta_n && (162)
\end{aligned}$$

$$\geq n\left[I(\tilde{Y}_{k-1};U_k)+\frac{1}{n}\log\Delta_n\right]. \tag{163}$$

Here, (157) holds by (155); (159) holds by the super-additivity property in [34, Proposition 1], by the chain rule, by the definition of $\tilde{U}_{k,t}$ and by defining T uniform over $\{1,\dots,n\}$ independent of the previously defined random variables; and (162) by the definitions of $U_k,\tilde{Y}_k,\tilde{Y}_{k-1}$ in the lemma. This proves Inequality (139) in the lemma.

Proof of (140)*:*

We start by noticing the Markov chain $\tilde{\mathsf{M}}_1 \to \tilde{Y}_0^n \to (\tilde{Y}_1^n,\cdots,\tilde{Y}_K^n)$, and thus similar to the analysis in [43, Section V.C]:

$$\begin{aligned}
0 &= I(\tilde{\mathsf{M}}_k;\tilde{Y}_k^n\cdots\tilde{Y}_K^n|\tilde{Y}_0^n\cdots\tilde{Y}_{k-1}^n) && (164)\\
&\geq H(\tilde{Y}_k^n\cdots\tilde{Y}_K^n|\tilde{Y}_0^n\cdots\tilde{Y}_{k-1}^n)\\
&\quad -H(\tilde{Y}_k^n\cdots\tilde{Y}_K^n|\tilde{Y}_0^n\cdots\tilde{Y}_{k-1}^n\tilde{M}_k)\\
&\quad +D(P_{\tilde{Y}_0^n\cdots\tilde{Y}_K^n}\|P^n_{Y_0\cdots Y_K})+\log\Delta && (165)\\
&\geq n[H(\tilde{Y}_{k,T}\cdots\tilde{Y}_{K,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{k-1,T})\\
&\qquad +D(P_{\tilde{Y}_{0,T}\cdots\tilde{Y}_{K,T}}\|P_{Y_0\cdots Y_K})]+\log\Delta\\
&\quad -H(\tilde{Y}_k^n\cdots\tilde{Y}_K^n|\tilde{Y}_0^n\cdots\tilde{Y}_{k-1}^n\tilde{\mathsf{M}}_k) && (166)\\
&\geq n[H(\tilde{Y}_{k,T}\cdots\tilde{Y}_{K,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{k-1,T})\\
&\quad +D(P_{\tilde{Y}_{0,T}\cdots\tilde{Y}_{K,T}}\|P_{Y_0\cdots Y_K})]+\log\Delta\\
&\quad -\sum_{t=1}^n H(\tilde{Y}_{k,t}\cdots\tilde{Y}_{K,t}|\tilde{Y}_{0,t}\cdots\tilde{Y}_{k-1,t}\\
&\qquad\qquad \tilde{Y}_0^{t-1}\cdots\tilde{Y}_K^{t-1}\tilde{Y}_{0,t+1}^n\cdots\tilde{Y}_{k-1,t+1}^n\tilde{\mathsf{M}}_k) && (167)\\
&\geq nH(\tilde{Y}_{k,T}\cdots\tilde{Y}_{K,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{k-1,T})+\log\Delta\\
&\quad -nH(\tilde{Y}_{k,T}\cdots\tilde{Y}_{K,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{k-1,T}\\
&\qquad\qquad \tilde{Y}_0^{T-1}\cdots\tilde{Y}_K^{T-1}\tilde{Y}_{0,T+1}^n\cdots\tilde{Y}_{k-1,T+1}^n\tilde{\mathsf{M}}_kT) && (168)\\
&\geq nI(\tilde{Y}_{k,T}\cdots\tilde{Y}_{K,T};\tilde{Y}_0^{T-1}\cdots\tilde{Y}_K^{T-1}\tilde{\mathsf{M}}_kT|\tilde{Y}_{0,T}\cdots\tilde{Y}_{k-1,T})\\
&\qquad +\log\Delta_n && (169)\\
&= nI(\tilde{Y}_k\cdots\tilde{Y}_K;U_k|\tilde{Y}_0\cdots\tilde{Y}_{k-1})+\log\Delta && (170)\\
&\geq nI(\tilde{Y}_k;U_k|\tilde{Y}_0\cdots\tilde{Y}_{k-1})+\log\Delta && (171)\\
&\geq nI(\tilde{Y}_k;U_k|\tilde{Y}_{k-1})-nI(\tilde{Y}_k;\tilde{Y}_0\cdots\tilde{Y}_{k-2}|\tilde{Y}_{k-1})\\
&\qquad +\log\Delta. && (172)
\end{aligned}$$

where (166) holds by the super-additivity property in [34, Proposition 1]; (167) by the chain rule; (168) by the non-negativity of the Kullback-Leibler divergence.

The desired inequality (140) is then obtained by combining (172) with the following lower bound:

$$\begin{aligned}
&D(P_{\tilde{Y}_0\cdots\tilde{Y}_K}\|P_{Y_0\cdots Y_K})\\
&\quad\geq D(P_{\tilde{Y}_0\cdots\tilde{Y}_k}\|P_{Y_0\cdots Y_k}) && (173)\\
&\quad= D(P_{\tilde{Y}_0\cdots\tilde{Y}_k}\|P_{Y_0\cdots Y_{k-1}}P_{Y_k|Y_{k-1}}) && (174)\\
&\quad= D(P_{\tilde{Y}_0\cdots\tilde{Y}_{k-1}}\|P_{\tilde{Y}_0\cdots\tilde{Y}_{k-1}}P_{\tilde{Y}_k|\tilde{Y}_{k-1}})\\
&\qquad +\mathbb{E}_{P_{\tilde{Y}_{k-1}}}\left[D(P_{\tilde{Y}_k|\tilde{Y}_{k-1}}\|P_{Y_k|Y_{k-1}})\right]\\
&\qquad\quad +D(P_{\tilde{Y}_0\cdots\tilde{Y}_{k-1}}\|P_{Y_0\cdots Y_{k-1}}) && (175)\\
&\quad\geq D(P_{\tilde{Y}_0\cdots\tilde{Y}_k}\|P_{\tilde{Y}_0\cdots\tilde{Y}_{k-1}}P_{\tilde{Y}_k|\tilde{Y}_{k-1}}) && (176)\\
&\quad\geq I(\tilde{Y}_0\cdots\tilde{Y}_{k-2};\tilde{Y}_k|\tilde{Y}_{k-1}). && (177)
\end{aligned}$$

Proof of (142)*:* For each $k\in\{1,\dots,K\}$, define R_k's acceptance region

$$\mathcal{A}_k \triangleq \{(\mathsf{m}_k,y_k^n)\colon g_k(\mathsf{m}_k,y_k^n)=0\}. \tag{178}$$

Consider an index k satisfying Condition (141), i.e., satisfying $\mathcal{D}\subseteq\mathcal{A}_k$. Obviously,

$$P_{\tilde{\mathsf{M}}_k\tilde{Y}_k^n}(\mathcal{A}_k)=1. \tag{179}$$

Define for any $k\in\{1,\dots,K\}$ the pmfs

$$\begin{aligned}
&Q_{\tilde{\mathsf{M}}_k}(\mathsf{m}_k)\\
&\triangleq \sum_{y_0^n,y_1^n,\dots,y_{k-1}^n} P_{\tilde{Y}_0^n}(y_0^n)\cdots P_{\tilde{Y}_{k-1}^n}(y_{k-1}^n)\\
&\qquad\cdot\mathbb{1}\{\mathsf{m}_k=\phi_k(\phi_{k-1}(\cdots(\phi_1(y_0^n)\cdots)),y_{k-1}^n)\}, && (180)
\end{aligned}$$

and

$$\begin{aligned}
&Q_{\mathsf{M}_k}(\mathsf{m}_K)\\
&\triangleq \sum_{y_0^n,y_1^n,\dots,y_{k-1}^n} P_{Y_0^n}(y_0^n)\cdots P_{Y_{k-1}^n}(y_{k-1}^n)\\
&\qquad\cdot\mathbb{1}\{\mathsf{m}_k=\phi_k(\phi_{k-1}(\cdots(\phi_1(y_0^n)\cdots)),y_{k-1}^n)\}. && (181)
\end{aligned}$$

and notice that by the definitions of $P_{\tilde{M}_k\tilde{Y}_k^n},Q_{\tilde{\mathsf{M}}_k}$ and by (156):

$$Q_{\tilde{\mathsf{M}}_k}P_{\tilde{Y}_k^n}(\mathcal{A}_k)\leq Q_{\mathsf{M}_k}P_{Y_k^n}(\mathcal{A}_k)\,\Delta^{-(k+1)}. \tag{182}$$

By (179) and (182) and the definition of $\beta_{k,n}$ we have:

$$\begin{aligned}
&-\frac{1}{n}\log\beta_{k,n}\\
&\quad= -\frac{1}{n}\log Q_{M_k}P^n_{Y_k}(\mathcal{A}_k) && (183)\\
&\quad= -\frac{1}{n}\log\left(Q_{\tilde{\mathsf{M}}_k}P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right)-\frac{(k+1)}{n}\log\Delta && (184)\\
&\quad= \frac{1}{n}D\left(P_{\tilde{\mathsf{M}}_k\tilde{Y}_k^n}(\mathcal{A}_k)\parallel Q_{\tilde{\mathsf{M}}_k}P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right)-\frac{(k+1)}{n}\log\Delta && (185)\\
&\quad\leq \frac{1}{n}D(P_{\tilde{\mathsf{M}}_k\tilde{Y}_k^n}\|Q_{\tilde{\mathsf{M}}_k}P_{\tilde{Y}_k^n})-\frac{(k+1)}{n}\log\Delta, && (186)
\end{aligned}$$

where the inequality holds by the data processing inequality for KL-divergence.

We continue to upper bound the divergence term as

$$\begin{aligned}
&D(P_{\tilde{\mathsf{M}}_k\tilde{Y}_k^n}\|Q_{\tilde{\mathsf{M}}_k}P_{\tilde{Y}_k^n})\\
&\quad= I(\tilde{\mathsf{M}}_k;\tilde{Y}_k^n)+D(P_{\tilde{\mathsf{M}}_k}\|Q_{\tilde{\mathsf{M}}_k}) && (187)\\
&\quad\leq I(\tilde{\mathsf{M}}_k;\tilde{Y}_k^n)+D(P_{\tilde{Y}_{k-1}^n\tilde{\mathsf{M}}_{k-1}}\|P_{\tilde{Y}_{k-1}^n}Q_{\tilde{\mathsf{M}}_{k-1}}) && (188)\\
&\quad\leq I(\tilde{\mathsf{M}}_k;\tilde{Y}_k^n)+I(\tilde{\mathsf{M}}_{k-1};\tilde{Y}_{k-1}^n)\\
&\qquad\qquad +D(P_{\tilde{Y}_{k-2}^n\tilde{\mathsf{M}}_{k-2}}\|P_{\tilde{Y}_{k-2}^n}Q_{\tilde{\mathsf{M}}_{k-2}}) && (189)\\
&\quad\vdots\\
&\quad\leq \sum_{i=1}^k I(\tilde{\mathsf{M}}_i;\tilde{Y}_i^n) && (190)\\
&\quad= \sum_{i=1}^k\sum_{t=1}^n I(\tilde{\mathsf{M}}_i;\tilde{Y}_{i,t}|\tilde{Y}_i^{t-1}) && (191)\\
&\quad\leq \sum_{i=1}^k\sum_{t=1}^n I(\tilde{\mathsf{M}}_i\tilde{Y}_0^{t-1}\cdots\tilde{Y}_K^{t-1};\tilde{Y}_{i,t}) && (192)
\end{aligned}$$

$$= \sum_{i=1}^{k} \sum_{t=1}^{n} I(\tilde{U}_{i,t}; \tilde{Y}_{i,t}) \tag{193}$$

$$= \sum_{i=1}^{k} nI(\tilde{U}_{i,T}; \tilde{Y}_{i,T}|T) \tag{194}$$

$$\leq n \sum_{i=1}^{k} I(U_i; \tilde{Y}_i). \tag{195}$$

Here (188) is obtained by the data processing inequality for KL-divergence; (191) by the chain rule; and (193)–(195) by the definitions of $\tilde{U}_{i,t}, U_i, \tilde{Y}_i$ and T.

Combined with (183), Inequality (195) establishes Inequality (142) for $k \in \{1, \ldots, K\}$.

IX. Discussion and Outlook

We derived the optimal type-II exponents region under expected-rate constraints for the K-hop network with K decision centers (DC) for testing against independence and when the observations at the sensors respect some Markov chain. Equivalent simplified expressions were proved for $K \in \{2, 3\}$ and when all DCs have same admissible type-I error probabilities. Similar simplifications are conjectured to hold also in the general case. When the various DCs have different permissible type-I errors, then the derived exponents region illustrates a tradeoff between the error exponents that are simultaneously achievable at the various DCs. In general, an increase in exponents region is observed compared to the setup with maximum-rate constraints. When all DCs have equal permissible type-I error probability ϵ, then the exponents region degenerates to a K-dimensional hypercube meaning that all DCs can simultaneously achieve their optimal error exponents. This optimal exponent coincides with the optimal exponent under maximum-rate constraint where the rates have to be boosted by the factor $(1-\epsilon)^{-1}$.

To achieve the optimal tradeoff, a novel coding and testing scheme based on multiplexing and rate-sharing is proposed. The idea is that the transmitter chooses one of 2^K subschemes with appropriate probabilities and applies each subscheme with a well-chosen rate tuple. Notice that the various rate-tuples determine the error exponents achieved at the various DCs, and thus steer the tradeoff between the error exponents at the different DCs. We multiplex schemes in a way that each of the subschemes is meant to help only a subset of the DCs in their decision; all other DCs simply raise an alarm so as not to compromise their type-II error exponents. The probabilities of the various subschemes then have to be chosen such that the probability of each DC raising an alarm does not exceed its permissible type-I error probability. We conjecture that it suffices to multiplex only $K+1$ subschemes and that they should be chosen with probabilities determined by the type-I error probabilities. We managed to prove this conjecture for $K \in \{2, 3\}$ and when all DCs have same permissible type-I error probabilities, but proofs for the general case seem cumbersome.

Notice that the proposed multiplexing and rate-sharing strategy is also optimal for other multi-terminal hypothesis testing setups, as we show in [44].

Our converse proof methods rely on applying 2^K change of measure arguments in parallel, and to separately bound the achievable error exponents and the required rates for each of them. Moreover, we prove the desired Markov chains of the auxiliary random variables that arise in the typical single-letterization steps, in the asymptotic regimes of infinite blocklengths. The proof technique of using asymptotic Markov chains in connection with change of measure arguments can also be used to prove strong converse results of source coding and channel coding theorems, see [41] for first results.

Interesting future research directions include results for other types of hypothesis testing, not necessarily testing against independence or not assuming a Markov chain under the null hypothesis. Other network structures are also of practical importance. Intriguing following-up questions exist also from an optimization perspective. For example, finding the optimal rate-distribution across the various links so as to maximize a weighted sum of the exponents.

Acknowledgment

M. Wigger and M. Hamad have been supported by the European Union's Horizon 2020 Research And Innovation Programme under grant agreement no. 715111.

References

[1] M. Hamad, M. Wigger, and M. Sarkiss, "Optimal exponents in cascaded hypothesis testing under expected rate constraints," in *2021 IEEE Information Theory Workshop (ITW)*, pp. 1–6, 2021.

[2] M. Hamad, M. Wigger, and M. Sarkiss, "Two-hop network with multiple decision centers under expected-rate constraints," in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2021.

[3] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.

[4] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, pp. 759–772, Nov. 1987.

[5] T. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 2–14, 1989.

[6] H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 254–267, 1992.

[7] H. Shimokawa, T. Han, and S. I. Amari, "Error bound for hypothesis testing with data compression," in *Proc. ISIT*, p. 114, Jul. 1994.

[8] 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.

[9] N. Weinberger, Y. Kochman, and M. Wigger, "Exponent trade-off for hypothesis testing over noisy channels," in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 1852–1856, 2019.

[10] S. Watanabe, "On sub-optimality of random binning for distributed hypothesis testing." [Online]. Available: https://arxiv.org/abs/2201.13005, 2022.

[11] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2044–2066, 2020.

[12] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *IEEE Trans. Inf. Theory*, vol. 66, pp. 4150–41820, Jul. 2020.

[13] S. Sreekumar and D. Gündüz, "Strong converse for testing against independence over a noisy channel," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1283–1288, 2020.

[14] S. Sreekumar, D. Gündüz, and A. Cohen, "Distributed hypothesis testing under privacy constraints," in *2018 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2018.

[15] A. Gilani, S. Belhadj Amor, S. Salehkalaibar, and V. Y. F. Tan, "Distributed hypothesis testing with privacy constraints," *Entropy*, vol. 21, p. 478, May 2019.

[16] Y. Xiang and Y.-H. Kim, "Interactive hypothesis testing with communication constraints," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1065–1072, IEEE, 2012.

[17] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing against independence," in *Proc. ISIT*, pp. 2840–2844, Jun. 2013.

[18] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing with general hypotheses," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 1705–1709, IEEE, 2016.

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

[20] W. Zhao and L. Lai, "Distributed detection with vector quantizer," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 2, pp. 105–119, 2016.

[21] W. Zhao and L. Lai, "Distributed testing with cascaded encoders," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7339–7348, 2018.

[22] M. Wigger and R. Timo, "Testing against independence with multiple decision centers," in *2016 International Conference on Signal Processing and Communications (SPCOM)*, pp. 1–5, 2016.

[23] S. Salehkalaibar, M. Wigger, and R. Timo, "On hypothesis testing against independence with multiple decision centers," *IEEE Trans. on Comm.* , vol. 66, pp. 2409–2420, Jan. 2018.

[24] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detection," in *Proc. ISIT*, Jun. 2018.

[25] P. Escamilla, A. Zaidi, and M. Wigger, "Distributed hypothesis testing with collaborative detection," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 512–518, 2018.

[26] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing: cooperation and concurrent detection," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7550–7564, 2020.

[27] 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.

[28] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing in multi-hop networks." [Online]. Available: https://arxiv.org/abs/1708.05198v1, 2017.

[29] 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.

[30] 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.

[31] S. Salehkalaibar and V. Y. Tan, "Distributed sequential hypothesis testing with zero-rate compression," in *2021 IEEE Information Theory Workshop (ITW)*, pp. 1–5, IEEE, 2021.

[32] Y. Inan, M. Kayaalp, A. H. Sayed, and E. Telatar, "A fundamental limit of distributed hypothesis testing under memoryless quantization," p. 8, 2021.

[33] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," in *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pp. 1–5, 2020.

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

[35] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *2009 IEEE International Symposium on Information Theory*, pp. 2316–2320, IEEE, 2009.

[36] W. Gu and M. Effros, "A strong converse in source coding for super-source networks," in *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 395–399, IEEE, 2011.

[37] J. Liu, R. Van Handel, and S. Verdú, "Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 943–947, IEEE, 2017.

[38] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.

[39] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[40] Y. Oohama, "Exponential strong converse for source coding with side information at the decoder," *Entropy*, vol. 20, no. 5, p. 352, 2018.

[41] M. Hamad, M. Wigger, and M. Sarkiss, "Strong converses using change of measure and asymptotic Markov chains." [Online]. Available: https://arxiv.org/abs/2205.08910, May 2022.

[42] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[43] 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.

[44] M. Hamad, M. Sarkiss, and M. Wigger, "Benefits of rate-sharing for distributed hypothesis testing." [Online]. Available: https://arxiv.org/pdf/2202.02282.pdf, 2022.

[45] I. B. Gattegno, Z. Goldfeld, and H. H. Permuter, "Fourier-motzkin elimination software for information theoretic inequalities." [Online]. Available: https://arxiv.org/abs/1610.03990, 2016.

Appendix A
Proof of Lemma 1: Concavity and Monotonicity of the Function η_1

The function $\eta_1(R)$ is monotonically non-decreasing because larger values of R imply larger optimization domains. Continuity follows simply by the continuity of mutual information.

The concavity of $\eta_1(R)$ follows by the following arguments. Consider rates R and $\tilde{R}$, and let U^* and $\tilde{U}^*$ be the corresponding solutions to the optimizations in the definition of η_1. Pick any $\lambda \in [0,1]$, define $Q \sim \text{Bern}(\lambda)$ independent of $(Y_0, Y_1, U^*, \tilde{U}^*)$, and set

$$U_Q^* = \begin{cases} U^* & \text{if } Q = 0 \\ \tilde{U}^* & \text{if } Q = 1. \end{cases} \tag{196}$$

Defining the random variable $V := (U_Q^*, Q)$, we obtain

$$\begin{aligned} \lambda \cdot \eta_1(R) + (1-\lambda) \cdot \eta_1(\tilde{R}) & \\ &= \lambda I(U^*; Y_1) + (1-\lambda) I(\tilde{U}^*; Y_1) && (197)\\ &= I(U_Q^*; Y_1 | Q) && (198)\\ &= I(U_Q^*, Q; Y_1) && (199)\\ &= I(V; Y_1) && (200)\\ &\leq \eta_1(I(V; Y_0)) && (201)\\ &\leq \eta_1(\lambda R + (1-\lambda)\tilde{R}) && (202) \end{aligned}$$

where (199) holds because Q is independent of Y_1, (201) holds by the definition of the function η_1, and (202) holds by the monotonicity of the function η_1 and the following set of (in)equalities:

$$\begin{aligned} I(V; Y_0) = I(U_Q^*, Q; Y_0) = I(U_Q^*; Y_0 | Q) && (203)\\ = \lambda I(U^*; Y_0) + (1-\lambda) I(\tilde{U}^*; Y_0) && (204)\\ \leq \lambda R + (1-\lambda)\tilde{R}. && (205) \end{aligned}$$

■

Appendix B
Analysis of the coding scheme in Subsection IV-A for $\epsilon_1 = \epsilon_2 = \epsilon$

Consider the two-hop scheme employed when $Y_0^n \in \mathcal{D}_{\{1,2\}}$, and let $\hat{\mathcal{H}}_{\{1,2\},1}$ and $\hat{\mathcal{H}}_{\{1,2\},2}$ denote the guesses produced at R_1 and R_2 when employing this scheme for any $Y_0^n \in \mathcal{Y}_0^n$. Notice that by assumption the type-I error probabilities of this scheme tend to 0 as $n \to \infty$:

$$\lim_{n\to\infty} \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1 | \mathcal{H} = 0] = 0, \quad k \in \{1,2\}. \tag{206}$$

Noticing that when $Y_0^n \in \mathcal{D}_\emptyset$, then $\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1$, and applying the total law of probability, we can write for $k \in \{1,2\}$:

$$\begin{aligned}\alpha_{k,n} &= \Pr[\hat{\mathcal{H}}_k = 1|\mathcal{H}=0] &(207)\\ &= \Pr[\hat{\mathcal{H}}_k = 1, Y_0^n \in \mathcal{D}_\emptyset|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_k = 1, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=0] &(208)\\ &= \Pr[Y_0^n \in \mathcal{D}_\emptyset|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=0] &(209)\\ &\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1|\mathcal{H}=0] &(210)\end{aligned}$$

Combining these inequalities with (206), and because in the limit $n \to \infty$ Inequality (17) turns into an equality, we conclude that the overall scheme satisfies the type-I error constraints:

$$\overline{\lim_{n\to\infty}}\, \alpha_{k,n} \leq \epsilon, \qquad k \in \{1,2\}. \tag{211}$$

For the type-II error probabilities of the overall scheme we observe for $k \in \{1,2\}$:

$$\begin{aligned}\beta_{1,n} &= \Pr[\hat{\mathcal{H}}_k = 0|\mathcal{H}=1] &(212)\\ &= \Pr[\hat{\mathcal{H}}_k = 0, Y_0^n \in \mathcal{D}_\emptyset|\mathcal{H}=1]\\ &\quad + \Pr[\hat{\mathcal{H}}_k = 0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] &(213)\\ &= \Pr[\hat{\mathcal{H}}_k = 0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] &(214)\\ &= \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] &(215)\\ &\leq \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 0|\mathcal{H}=1]. &(216)\end{aligned}$$

The type-II error exponents of the overall scheme are thus given by the error exponents of the two-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1,2\}}$. By [27] and because the two-hop scheme has to have vanishing type-I error probabilities and respect the rate constraints $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$, the exponents in (31) are proved achievable.

Appendix C
Analysis of the coding scheme in Subsection IV-B for $\epsilon_2 > \epsilon_1$

Consider the two-hop scheme employed when $Y_0^n \in \mathcal{D}_{\{1,2\}}$, and let $\hat{\mathcal{H}}_{\{1,2\},1}$ and $\hat{\mathcal{H}}_{\{1,2\},2}$ denote the guesses produced at R_1 and R_2 when employing this scheme for any $y_0^n \in \mathcal{Y}_0^n$. Similarly, let $\hat{\mathcal{H}}_{\{1\},1}$ and $\hat{\mathcal{H}}_{\{1\},2}$ denote the guesses produced at R_1 and R_2 when employing the scheme for $Y_0^n \in \mathcal{D}_{\{1\}}$, where we again extend the scheme to the entire set $\mathcal{Y}_0$.

By assumption, the type-I error probabilities of these schemes tend to 0 as $n \to \infty$:

$$\lim_{n\to\infty} \Pr[\hat{\mathcal{H}}_{\{1\},k} = 1|\mathcal{H}=0] = 0, \quad k \in \{1,2\} \tag{217a}$$

$$\lim_{n\to\infty} \Pr[\hat{\mathcal{H}}_{\{1,2\},k} = 1|\mathcal{H}=0] = 0, \quad k \in \{1,2\}. \tag{217b}$$

Notice that for $Y_0^n \in \mathcal{D}_\emptyset$ both R_1 and R_2 decide on $\hat{\mathcal{H}}_1 = \hat{\mathcal{H}}_2 = 1$. Applying the total law of probability, we can write

$$\begin{aligned}\alpha_{1,n} &= \Pr[\hat{\mathcal{H}}_1 = 1|\mathcal{H}=0] &(218)\\ &= \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_\emptyset|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1\}}|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_1 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=0] &(219)\\ &= \Pr[Y_0^n \in \mathcal{D}_\emptyset|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1, Y_0^n \in \mathcal{D}_{\{1\}}|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=0] &(220)\\ &\leq \Pr[Y_0^n \in \mathcal{D}_\emptyset|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1\},1} = 1|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 1|\mathcal{H}=0] &(221)\end{aligned}$$

Combining this inequality with (217), and because in the limit $n \to \infty$ Inequality (128) turns into an equality, we conclude that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim_{n\to\infty}}\, \alpha_{1,n} \leq \epsilon_1. \tag{222}$$

Similarly we have:

$$\begin{aligned}\alpha_{2,n} &= \Pr[\hat{\mathcal{H}}_2 = 1|\mathcal{H}=0] &(223)\\ &= \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{1\}})|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_2 = 1, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=0] &(224)\\ &= \Pr[Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{1\}})|\mathcal{H}=0]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=0] &(225)\\ &\leq \Pr[Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{1\}})|\mathcal{H}=0] &(226)\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1,2\},2} = 1|\mathcal{H}=0]. &(227)\end{aligned}$$

Combining this inequality with (217), and because in the limit $n \to \infty$ Inequalities (21a) and (128) turn into equalities, we conclude that the overall scheme satisfies the type-I error constraint:

$$\overline{\lim_{n\to\infty}}\, \alpha_{2,n} \leq \epsilon_2. \tag{228}$$

For the relay's type-II error probability in the overall scheme we observe:

$$\begin{aligned}\beta_{1,n} &= \Pr[\hat{\mathcal{H}}_1 = 0|\mathcal{H}=1] &(229)\\ &= \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_\emptyset|\mathcal{H}=1]\\ &\quad + \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1\}}|\mathcal{H}=1]\\ &\quad + \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] &(230)\\ &= \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1\}}|\mathcal{H}=1]\\ &\quad + \Pr[\hat{\mathcal{H}}_1 = 0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] &(231)\\ &= \Pr[\hat{\mathcal{H}}_{\{1\},1} = 0, Y_0^n \in \mathcal{D}_{\{1\}}|\mathcal{H}=1]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] &(232)\\ &\leq \Pr[\hat{\mathcal{H}}_{\{1\},1} = 0|\mathcal{H}=1]\\ &\quad + \Pr[\hat{\mathcal{H}}_{\{1,2\},1} = 0|\mathcal{H}=1]. &(233)\end{aligned}$$

The relay's type-II error exponent of the overall scheme is thus given by the minimum of the error exponents of the single-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1\}}$ and of two-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1,2\}}$. By [4] and [27] and because these schemes have vanishing type-I error probabilities and respect the rate constraints $R_{\{1\},1}$ and $(R_{\{1,2\},1}, R_{\{1,2\},2})$, respectively, the exponent θ_1 in (32) is proved achievable.

It remains to analyze the receiver's type-II error exponent:

$$\begin{aligned}\beta_{2,n} &= \Pr[\hat{\mathcal{H}}_2 = 0|\mathcal{H}=1] &(234)\\ &= \Pr[\hat{\mathcal{H}}_2 = 0, Y_0^n \in (\mathcal{D}_\emptyset \cup \mathcal{D}_{\{1\}})|\mathcal{H}=1]\end{aligned}$$

$$+\Pr[\hat{\mathcal{H}}_2=0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] \tag{235}$$
$$=\Pr[\hat{\mathcal{H}}_2=0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] \tag{236}$$
$$=\Pr[\hat{\mathcal{H}}_{\{1,2\},2}=0, Y_0^n \in \mathcal{D}_{\{1,2\}}|\mathcal{H}=1] \tag{237}$$
$$\leq \Pr[\hat{\mathcal{H}}_{\{1,2\},2}=0|\mathcal{H}=1]. \tag{238}$$

The receiver's type-II error exponent of the overall scheme is thus given by the error exponent of the two-hop scheme employed under $Y_0^n \in \mathcal{D}_{\{1,2\}}$. By [27] and because this scheme has vanishing type-I error probabilities and respects the rate constraints $(R_{\{1,2\},1}, R_{\{1,2\},2})$, the exponent θ_2 in (32) is proved achievable.

Appendix D
Simplification of the Outer Bound in Proposition 3

We proceed to simplify the outer bound in Proposition 3 depending on the cases $\epsilon_1=\epsilon_2$, $\epsilon_1<\epsilon_2$, or $\epsilon_1>\epsilon_2$. To this end, fix an exponent pair (θ_1,θ_2) in $\mathcal{E}^\star(R_1,R_2,\epsilon_1,\epsilon_2)$, rates $R_{\{1\},1}, R_{\{1,2\},1}, R_{\{1,2\},2}, R_{\{2\},1}, R_{\{2\},2} \geq 0$ and numbers $\sigma_{\{1\}},\sigma_{\{2\}},\sigma_{\{1,2\}} \geq 0$ summing to less than 1 and satisfying constraints (54).

1) The case $\epsilon_1=\epsilon_2$: By (54):

$$\theta_1 \leq \min\{\eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1})\} \tag{239}$$
$$\leq \frac{\sigma_{\{1\}}\eta_1(R_{\{1\},1})+\sigma_{\{1,2\}}\eta_1(R_{\{1,2\},1})}{\sigma_{\{1\}}+\sigma_{\{1,2\}}} \tag{240}$$
$$\leq \eta_1\left(\frac{\sigma_{\{1\}}R_{\{1\},1}+\sigma_{\{1,2\}}R_{\{1,2\},1}}{\sigma_{\{1\}}+\sigma_{\{1,2\}}}\right) \tag{241}$$
$$\leq \eta_1\left(R_1/(1-\epsilon)\right), \tag{242}$$

where (240) holds because the minimum is never larger than any linear combination; (241) holds by the concavity of the function $\eta_1(\cdot)$; and (242) holds by the monotonicity of the function $\eta_1(\cdot)$ and because by (54) we have $\sigma_{\{1\}}R_{\{1\},1}+\sigma_{\{1,2\}}R_{\{1,2\},1} \leq R_1$ and $\sigma_{\{1\}}+\sigma_{\{1,2\}} \geq 1-\epsilon$.

Following similar steps, one can prove that

$$\theta_2 \leq \min\Big\{\eta_1\left(R_{\{1,2\},1}\right)+\eta_2\left(R_{\{1,2\},2}\right), \eta_1\left(R_{\{2\},1}\right)+\eta_2\left(R_{\{2\},2}\right)\Big\} \tag{243}$$
$$\leq \frac{\sigma_{\{2\}}\eta_1\left(R_{\{2\},1}\right)+\sigma_{\{2\}}\eta_2\left(R_{\{2\},2}\right)}{\sigma_{\{2\}}+\sigma_{\{1,2\}}} + \frac{\sigma_{\{1,2\}}\eta_1\left(R_{\{1,2\},1}\right)+\sigma_{\{1,2\}}\eta_2\left(R_{\{1,2\},2}\right)}{\sigma_{\{2\}}+\sigma_{\{1,2\}}} \tag{244}$$
$$\leq \eta_1\left(\frac{\sigma_{\{2\}}R_{\{2\},1}+\sigma_{\{1,2\}}R_{\{1,2\},1}}{\sigma_{\{2\}}+\sigma_{\{1,2\}}}\right) + \eta_2\left(\frac{\sigma_{\{2\}}R_{\{2\},2}+\sigma_{\{1,2\}}R_{\{1,2\},2}}{\sigma_{\{2\}}+\sigma_{\{1,2\}}}\right) \tag{245}$$
$$\leq \eta_1\left(R_1/(1-\epsilon)\right)+\eta_2\left(R_2/(1-\epsilon)\right), \tag{246}$$

where (244) holds again because the minimum is never larger than any linear combination; (245) holds by the concavity of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$; and (246) holds because by (54) we have $\sigma_{\{2\}}R_{\{2\},i}+\sigma_{\{1,2\}}R_{\{1,2\},i} \leq R_i$, for $i\in\{1,2\}$, and $\sigma_{\{2\}}+\sigma_{\{1,2\}} \geq 1-\epsilon$.

This concludes the converse proof to (31).

2) The case $\epsilon_1<\epsilon_2$: Choose nonnegative numbers $a_1, a_{1,2}, b_1, b_{1,2}, c_{1,2}$ satisfying

$$a_1+a_{1,2} \leq \sigma_{\{1\}} \tag{247a}$$
$$b_1+b_{1,2} \leq \sigma_{\{1,2\}} \tag{247b}$$
$$c_{1,2} \leq \sigma_{\{2\}} \tag{247c}$$
$$a_{1,2}+b_{1,2}=b_{1,2}+c_{1,2}=1-\epsilon_2 \tag{247d}$$
$$a_1+b_1=\epsilon_2-\epsilon_1. \tag{247e}$$

Notice that this set of (in)equalities is equivalent to the two equalities $a_{1,2}=c_{1,2}=1-\epsilon_2-b_{1,2}$ and $a_1=\epsilon_2-\epsilon_1-b_1$ and the three inequalities:

$$1-\epsilon_1-b_{1,2}-b_1 \leq \sigma_{\{1\}} \tag{248a}$$
$$b_1+b_{1,2} \leq \sigma_{\{1,2\}} \tag{248b}$$
$$1-\epsilon_2-b_{1,2} \leq \sigma_{\{2\}}. \tag{248c}$$

Through the Fourier-Motzkin Elimination (FME) Algorithm, it can be verified that above three inequalities (248) have a nonnegative solution pair $(b_1,b_{1,2})$, with corresponding nonnegative values for $a_{1,2}, c_{1,2}, a_1$, whenever

$$0 \leq \sigma_{\mathcal{I}}, \qquad \mathcal{I}\in\mathcal{P}(2), \tag{249a}$$
$$1-\epsilon_i \leq \sigma_{\{i\}}+\sigma_{\{1,2\}}, \quad i\in\{1,2\}, \tag{249b}$$
$$0 \leq \epsilon_2-\epsilon_1, \tag{249c}$$

which hold by assumption, see (54). The existence of the desired nonnegative numbers $a_1, a_{1,2}, b_1, b_{1,2}, c_{1,2}$ satisfying (247) is thus established.

With the chosen numbers, we form

$$\tilde{R}_{\{1,2\},1} := \max\left\{\frac{a_{1,2}R_{\{1\},1}+b_{1,2}R_{\{1,2\},1}}{1-\epsilon_2}, \frac{b_{1,2}R_{\{1,2\},1}+c_{1,2}R_{\{2\},1}}{1-\epsilon_2}\right\}, \tag{250a}$$
$$\tilde{R}_{\{1,2\},2} := \frac{b_{1,2}R_{\{1,2\},2}+c_{1,2}R_{\{2\},2}}{1-\epsilon_2}, \tag{250b}$$
$$\tilde{R}_{\{1\},1} := \frac{a_1R_{\{1\},1}+b_1R_{\{1,2\},1}}{\epsilon_2-\epsilon_1}. \tag{250c}$$

We show that exponents (θ_1,θ_2) and rates $\tilde{R}_{\{1\},1}$, $\tilde{R}_{\{1,2\},1}$ and $\tilde{R}_{\{1,2\},2}$ satisfy constraints (32). To this end, notice that

$$\theta_1 \leq \min\{\eta_1(R_{\{1\},1}), \eta_1(R_{\{1,2\},1})\} \tag{251}$$
$$\leq \frac{a_1\eta_1(R_{\{1\},1})+b_1\eta_1(R_{\{1,2\},1})}{\epsilon_2-\epsilon_1} \tag{252}$$
$$\leq \eta_1\left(\frac{a_1R_{\{1\},1}+b_1R_{\{1,2\},1}}{\epsilon_2-\epsilon_1}\right) \tag{253}$$
$$\leq \eta_1\left(\tilde{R}_{\{1\},1}\right), \tag{254}$$

where (252) holds because the minimum is smaller than any linear combination and because $a_1+b_1=\epsilon_2-\epsilon_1$; (253) holds by the concavity of the function $\eta_1(\cdot)$; and (254) holds by the definition of rate $R_{\{1\},1}$. In a similar way we have:

$$\theta_1 \leq \min\left\{\eta_1\left(R_{\{1\},1}\right), \eta_1\left(R_{\{1,2\},1}\right)\right\} \tag{255}$$
$$\leq \frac{a_{1,2}\eta_1\left(R_{\{1\},1}\right)+b_{1,2}\eta_1\left(R_{\{1,2\},1}\right)}{1-\epsilon_2} \tag{256}$$

$$\leq \eta_1\left(\frac{a_{1,2}R_{\{1\},1}+b_{1,2}R_{\{1,2\},1}}{1-\epsilon_2}\right) \tag{257}$$

$$\leq \eta_1\left(\tilde{R}_{\{1,2\},1}\right), \tag{258}$$

where the last step holds by the monotonicity of the function $\eta_1(\cdot)$ and because by definition $\tilde{R}_{\{1,2\},1} \geq \frac{a_{1,2}R_{\{1\},1}+b_{1,2}R_{\{1,2\},1}}{1-\epsilon_2}$. Thus, by (254) and (258):

$$\theta_1 \leq \min\left\{\eta_1\left(\tilde{R}_{\{1\},1}\right), \eta_1\left(\tilde{R}_{\{1,2\},1}\right)\right\}. \tag{259}$$

We continue to notice

$$\theta_2 \leq \min\Big\{\eta_1\left(R_{\{1,2\},1}\right)+\eta_2\left(R_{\{1,2\},2}\right), \eta_1\left(R_{\{2\},1}\right)+\eta_2\left(R_{\{2\},2}\right)\Big\} \tag{260}$$

$$\leq \frac{b_{1,2}\eta_1\left(R_{\{1,2\},1}\right)+b_{1,2}\eta_2\left(R_{\{1,2\},2}\right)}{1-\epsilon_2} + \frac{c_{1,2}\eta_1\left(R_{\{2\},1}\right)+c_{1,2}\eta_2\left(R_{\{2\},2}\right)}{1-\epsilon_2} \tag{261}$$

$$\leq \eta_1\left(\frac{b_{1,2}R_{\{1,2\},1}+c_{1,2}R_{\{2\},1}}{1-\epsilon_2}\right) + \eta_2\left(\frac{b_{1,2}R_{\{1,2\},2}+c_{1,2}R_{\{2\},2}}{1-\epsilon_2}\right) \tag{262}$$

$$\leq \eta_1\left(\tilde{R}_{\{1,2\},1}\right)+\eta_2\left(\tilde{R}_{\{1,2\},2}\right), \tag{263}$$

where (261) holds because the minimum is smaller than any linear combination and because $b_{1,2}+c_{1,2}=1-\epsilon_2$; (262) holds concavity of the functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$; and (263) holds by the definitions of rates $R_{\{1,2\},1}$ and $R_{\{1,2\},2}$ and by the monotonicity of the function $\eta_1(\cdot)$.

From the rate constraints in (54), we further obtain

$$R_1 \geq \sigma_{\{1\}}R_{\{1\},1}+\sigma_{\{2\}}R_{\{2\},1}+\sigma_{\{1,2\}}R_{\{1,2\},1} \tag{264}$$

$$\geq (a_1+a_{1,2})R_{\{1\},1}+c_{1,2}R_{\{2\},1}+(b_1+b_{1,2})R_{\{1,2\},1} \tag{265}$$

$$= (\epsilon_2-\epsilon_1)\left(\frac{a_1R_{\{1\},1}+b_1R_{\{1,2\},1}}{\epsilon_2-\epsilon_1}\right) + (1-\epsilon_2)\left(\frac{a_{1,2}R_{\{1\},1}+c_{1,2}R_{\{2\},1}+b_{1,2}R_{\{1,2\},1}}{1-\epsilon_2}\right) \tag{266}$$

$$\geq (\epsilon_2-\epsilon_1)\tilde{R}_{\{1\},1}+(1-\epsilon_2)\tilde{R}_{\{1,2\},1} \tag{267}$$

and

$$R_2 \geq \sigma_{\{1,2\}}R_{\{1,2\},2}+\sigma_{\{2\}}R_{\{2\},2} \tag{268}$$

$$\geq b_{1,2}R_{\{1,2\},2}+c_{1,2}R_{\{2\},2} \tag{269}$$

$$= (1-\epsilon_2)\left(\frac{b_{1,2}R_{\{1,2\},2}+c_{1,2}R_{\{2\},2}}{1-\epsilon_2}\right) \tag{270}$$

$$= (1-\epsilon_2)\tilde{R}_{\{1,2\},2}, \tag{271}$$

where inequalities (265) and (269) hold because $a_1+a_{1,2} \leq \sigma_{\{1\}}$, $c_{1,2} \leq \sigma_{\{2\}}$, and $b_1+b_{1,2} \leq \sigma_{\{1,2\}}$, see (247); and (267) and (271) hold by the definitions of rates $\tilde{R}_{\{1\},1}$, $\tilde{R}_{\{1,2\},1}$, and $\tilde{R}_{\{1,2\},2}$ and because

$$a_{1,2}R_{\{1\},1}+c_{1,2}R_{\{2\},1}+b_{1,2}R_{\{1,2\},1} \geq \max\{a_{1,2}R_{\{1\},1}+b_{1,2}R_{\{1,2\},1}; b_{1,2}R_{\{1,2\},1}+c_{1,2}R_{\{2\},1}\}. \tag{272}$$

The desired converse result to (32) then follows by combining (259), (263), (267), and (271), and by noticing that by the monotonicity of the function $\eta_2(\cdot)$ there is no loss in optimality to restrict to rates $\tilde{R}_{\{1,2\},2}=R_2/(1-\epsilon_2)$.

3) The case $\epsilon_1 > \epsilon_2$: The proof is similar to the case $\epsilon_1 < \epsilon_2$. We present it here for completeness.

Choose nonnegative numbers $a_{1,2}, b_2, b_{1,2}, c_2, c_{1,2}$ satisfying

$$a_{1,2} \leq \sigma_{\{1\}} \tag{273a}$$

$$b_2+b_{1,2} \leq \sigma_{\{1,2\}} \tag{273b}$$

$$c_2+c_{1,2} \leq \sigma_{\{2\}} \tag{273c}$$

$$a_{1,2}+b_{1,2}=b_{1,2}+c_{1,2}=1-\epsilon_1 \tag{273d}$$

$$b_2+c_2=\epsilon_1-\epsilon_2, \tag{273e}$$

which is equivalent to the three equalities $a_{1,2}=c_{1,2}=1-\epsilon_1-b_{1,2}$ and $c_2=\epsilon_1-\epsilon_2-b_2$ and the three inequalities

$$1-\epsilon_1-b_{1,2} \leq \sigma_{\{1\}} \tag{274a}$$

$$b_2+b_{1,2} \leq \sigma_{\{1,2\}} \tag{274b}$$

$$1-\epsilon_2-b_2-b_{1,2} \leq \sigma_{\{2\}}. \tag{274c}$$

Through FME it can be shown that a nonnegative pair $(b_2, b_{1,2})$ satisfying (274) exists and the corresponding values for $a_{1,2}, c_{1,2}, c_2$ are nonnegative whenever

$$0 \leq \sigma_{\mathcal{I}}, \qquad \mathcal{I} \in \mathcal{P}(2), \tag{275a}$$

$$1-\epsilon_i \leq \sigma_{\{i\}}+\sigma_{\{1,2\}}, \quad i \in \{1,2\}, \tag{275b}$$

$$0 \leq \epsilon_1-\epsilon_2, \tag{275c}$$

which hold by assumption, see (54).

Define the new rates

$$\tilde{R}_{\{1,2\},1} := \max\left\{\frac{a_{1,2}R_{\{1\},1}+b_{1,2}R_{\{1,2\},1}}{1-\epsilon_1}, \frac{b_{1,2}R_{\{1,2\},1}+c_{1,2}R_{\{2\},1}}{1-\epsilon_1}\right\}, \tag{276}$$

$$\tilde{R}_{\{1,2\},2} := \frac{b_{1,2}R_{\{1,2\},2}+c_{1,2}R_{\{2\},2}}{1-\epsilon_1}, \tag{277}$$

$$\tilde{R}_{\{2\},i} := \frac{b_2R_{\{1,2\},i}+c_2R_{\{2\},i}}{\epsilon_1-\epsilon_2}, \qquad i \in \{1,2\} \tag{278}$$

We show that the exponents θ_1, θ_2 and the rates $\tilde{R}_{\{2\},1}$,$\tilde{R}_{\{2\},2}$, $\tilde{R}_{\{1,2\},1}$ and $\tilde{R}_{\{1,2\},2}$ satisfy constraints (33). To this end, notice that by similar arguments as in the preceding subsections:

$$\theta_1 \leq \min\left\{\eta_1\big(R_{\{1\},1}\big), \eta_1\big(R_{\{1,2\},1}\big)\right\} \tag{279}$$

$$\leq \frac{a_{1,2}\eta_1\big(R_{\{1\},1}\big)+b_{1,2}\eta_1\big(R_{\{1,2\},1}\big)}{1-\epsilon_1} \tag{280}$$

$$\leq \eta_1\left(\frac{a_{1,2}R_{\{1\},1}+b_{1,2}R_{\{1,2\},1}}{1-\epsilon_1}\right) \tag{281}$$

$$\leq \eta_1\left(\tilde{R}_{\{1,2\},1}\right). \tag{282}$$

Moreover,

$$\theta_2 \leq \min\left\{\eta_1\big(R_{\{1,2\},1}\big)+\eta_2\big(R_{\{1,2\},2}\big), \eta_1\big(R_{\{2\},1}\big)+\eta_2\big(R_{\{2\},2}\big)\right\} \tag{283}$$

$$\leq \frac{b_{1,2}\eta_1\big(R_{\{1,2\},1}\big)+b_{1,2}\eta_2\big(R_{\{1,2\},2}\big)}{1-\epsilon_1} + \frac{c_{1,2}\eta_1\big(R_{\{2\},1}\big)+c_{1,2}\eta_2\big(R_{\{2\},2}\big)}{1-\epsilon_1} \tag{284}$$

$$\leq \eta_1\left(\frac{b_{1,2}R_{\{1,2\},1}+c_{1,2}R_{\{2\},1}}{1-\epsilon_1}\right) + \eta_2\left(\frac{b_{1,2}R_{\{1,2\},2}+c_{1,2}R_{\{2\},2}}{1-\epsilon_1}\right) \tag{285}$$

$$\leq \eta_1\left(\tilde{R}_{\{1,2\},1}\right)+\eta_2\left(\tilde{R}_{\{1,2\},2}\right) \tag{286}$$

and

$$\theta_2 \leq \frac{b_2\eta_1\big(R_{\{1,2\},1}\big)+b_2\eta_2\big(R_{\{1,2\},2}\big)}{\epsilon_1-\epsilon_2} + \frac{c_2\eta_1\big(R_{\{2\},1}\big)+c_2\eta_2\big(R_{\{2\},2}\big)}{\epsilon_1-\epsilon_2} \tag{287}$$

$$\leq \eta_1\left(\frac{b_2R_{\{1,2\},1}+c_2R_{\{2\},1}}{\epsilon_1-\epsilon_2}\right) + \eta_2\left(\frac{b_2R_{\{1,2\},2}+c_2R_{\{2\},2}}{\epsilon_1-\epsilon_2}\right) \tag{288}$$

$$\leq \eta_1\left(\tilde{R}_{\{2\},1}\right)+\eta_2\left(\tilde{R}_{\{2\},2}\right). \tag{289}$$

Combining (286) and (289) we obtain:

$$\theta_2 \leq \min\Bigg\{\eta_1\left(\tilde{R}_{\{1,2\},1}\right)+\eta_1\left(\tilde{R}_{\{1,2\},2}\right), \eta_1\left(\tilde{R}_{\{2\},1}\right)+\eta_1\left(\tilde{R}_{\{2\},2}\right)\Bigg\}. \tag{290}$$

From the rate constraints in (54), inequalities (273), and the definitions of the rates $\tilde{R}_{\{2\},1}, \tilde{R}_{\{2\},2}, \tilde{R}_{\{1,2\},1}, \tilde{R}_{\{1,2\},2}$, we obtain:

$$R_1 \geq \sigma_{\{1\}}R_{\{1\},1}+\sigma_{\{1,2\}}R_{\{1,2\},1}+\sigma_{\{2\}}R_{\{2\},1} \tag{291}$$

$$\geq a_{1,2}R_{\{1\},1}+(c_2+c_{1,2})R_{\{2\},1}+(b_2+b_{1,2})R_{\{1,2\},1} \tag{292}$$

$$= (\epsilon_1-\epsilon_2)\left(\frac{b_2R_{\{1,2\},1}+c_2R_{\{2\},1}}{\epsilon_1-\epsilon_2}\right) + (1-\epsilon_1)\left(\frac{a_{1,2}R_{\{1\},1}+c_{1,2}R_{\{2\},1}+b_{1,2}R_{\{1,2\},1}}{1-\epsilon_1}\right) \tag{293}$$

$$\geq (\epsilon_1-\epsilon_2)\tilde{R}_{\{2\},1}+(1-\epsilon_1)\tilde{R}_{\{1,2\},1} \tag{294}$$

and

$$R_2 \geq \sigma_{\{1,2\}}R_{\{1,2\},2}+\sigma_{\{2\}}R_{\{2\},2} \tag{295}$$

$$\geq (b_2+b_{1,2})R_{\{1,2\},2}+(c_2+c_{1,2})R_{\{2\},2} \tag{296}$$

$$= (1-\epsilon_1)\left(\frac{b_{1,2}R_{\{1,2\},2}+c_{1,2}R_{\{2\},2}}{1-\epsilon_1}\right) + (\epsilon_1-\epsilon_2)\left(\frac{b_2R_{\{1,2\},2}+c_2R_{\{2\},2}}{1-\epsilon_1}\right) \tag{297}$$

$$= (1-\epsilon_1)\tilde{R}_{\{1,2\},2}+(\epsilon_1-\epsilon_2)\tilde{R}_{\{2\},2}. \tag{298}$$

Combining (282), (290), (294), and (298) establishes the desired converse result in (33).

Appendix E
Analysis of the coding scheme in Section VII-C

Consider the $\ell^*_{\mathcal{I}}$-hop hypothesis testing scheme employed when $Y_0^n \in \mathcal{D}_{\mathcal{I}}$, for $\mathcal{I} \in \mathcal{P}(K)$. For any $\mathcal{I} \in \mathcal{P}(K)$, let $\hat{\mathcal{H}}_{\mathcal{I},1},\ldots,\hat{\mathcal{H}}_{\mathcal{I},\ell^*_{\mathcal{I}}}$ denote the guesses produced at terminals $1,\ldots,\ell^*_{\mathcal{I}}$ when employing this scheme.

By assumption, the type-I error probabilities of these decisions tend to 0 as $n\to\infty$ for any $\mathcal{I}\in\mathcal{P}(K)$:

$$\lim_{n\to\infty}\Pr[\hat{\mathcal{H}}_{\mathcal{I},k}=1|\mathcal{H}=0,Y_0^n\in\mathcal{D}_{\mathcal{I}}]=0, \quad k\in\mathcal{I}. \tag{299}$$

Recalling that decision center k declares $\hat{\mathcal{H}}_k=1$ whenever $Y_0^n\in\mathcal{D}_\emptyset$ or $Y_0^n\in\mathcal{D}_{\mathcal{I}}$ for a set $\mathcal{I}$ not containing k, and applying the total law of probability, we can write

$$\alpha_{k,n} = \Pr[\hat{\mathcal{H}}_k=1|\mathcal{H}=0] \tag{300}$$

$$= \sum_{\mathcal{I}\in(\mathcal{P}(K)\cup\emptyset)}\Pr[\hat{\mathcal{H}}_k=1,Y_0^n\in\mathcal{D}_{\mathcal{I}}|\mathcal{H}=0] \tag{301}$$

$$= \Pr[Y_0^n\in\mathcal{D}_\emptyset|\mathcal{H}=0]+\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\notin\mathcal{I}}}\Pr[Y_0^n\in\mathcal{D}_{\mathcal{I}}|\mathcal{H}=0] + \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}}\Pr[\hat{\mathcal{H}}_k=1,Y_0^n\in\mathcal{D}_{\mathcal{I}}|\mathcal{H}=0] \tag{302}$$

$$\leq \Pr[Y_0^n\in\mathcal{D}_\emptyset|\mathcal{H}=0]+\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\notin\mathcal{I}}}\Pr[Y_0^n\in\mathcal{D}_{\mathcal{I}}|\mathcal{H}=0] + \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}}\Pr[\hat{\mathcal{H}}_{\mathcal{I},k}=1|\mathcal{H}=0,Y_0^n\in\mathcal{D}_{\mathcal{I}}]. \tag{303}$$

Combining this inequality with (299), and by Inequalities (109), we conclude that the overall scheme satisfies the type-I error constraints:

$$\overline{\lim_{n\to\infty}}\,\alpha_{k,n}\leq\epsilon_k, \quad k\in\{1,\ldots,K\}. \tag{304}$$

For the type-II error exponent at a decision center k we observe:

$$\beta_{k,n} = \Pr[\hat{\mathcal{H}}_k=0|\mathcal{H}=1] \tag{305}$$

$$= \sum_{\mathcal{I}\in(\mathcal{P}(K)\cup\emptyset)}\Pr[\hat{\mathcal{H}}_k=0,Y_0^n\in\mathcal{D}_{\mathcal{I}}|\mathcal{H}=1] \tag{306}$$

$$= \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}}\Pr[\hat{\mathcal{H}}_{\mathcal{I},k}=0,Y_0^n\in\mathcal{D}_{\mathcal{I}}|\mathcal{H}=1] \tag{307}$$

$$\leq \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}}\Pr[\hat{\mathcal{H}}_{\mathcal{I},k}=0|\mathcal{H}=1,Y_0^n\in\mathcal{D}_{\mathcal{I}}]. \tag{308}$$

Defining

$$\theta_{k,\mathcal{I}} := \underline{\lim_{n\to\infty}} -\frac{1}{n}\log\Pr[\hat{\mathcal{H}}_{\mathcal{I},k}=0|\mathcal{H}=1,Y_0^n\in\mathcal{D}_{\mathcal{I}}], \tag{309}$$

we conclude by (308) that the exponent

$$\min_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}}\theta_{k,\mathcal{I}} \tag{310}$$

is achievable at decision center k. This proves in particular that when applying an instance of the multi-hop scheme in [28] for each set $\mathcal{I}\in\mathcal{P}(K)$, the exponents $\theta_1,\ldots,\theta_K$ in (5) are proved achievable.

Appendix F
Proof of Lemma 5

To show sufficiency of (368), start by fixing any set of nonnegative numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{P}(3)}$, and $\{R_{\mathcal{I},1},\dots,R_{\mathcal{I},\ell^*_{\mathcal{I}}}\}_{\mathcal{I}\in\mathcal{P}(3)}$ satisfying (116) for $K=3$, (and possibly violating (368)). Choose new nonnegative numbers $\tilde{\sigma}_{\{1,2,3\}},\tilde{\sigma}_{\{\pi(1),\pi(2)\}},\tilde{\sigma}_{\{\pi(1),\pi(3)\}},\tilde{\sigma}_{\{\pi(1)\}}$ satisfying

$$\tilde{\sigma}_{\mathcal{I}}\leq\sigma_{\mathcal{I}},\quad\forall\mathcal{I}\colon\pi(1)\in\mathcal{I},\tag{311}$$

$$\tilde{\sigma}_{\{1,2,3\}}+\tilde{\sigma}_{\{\pi(1),\pi(2)\}}\geq 1-\epsilon_{\pi(1)}-\epsilon_{\pi(2)}\tag{312}$$

$$\tilde{\sigma}_{\{1,2,3\}}+\tilde{\sigma}_{\{\pi(1),\pi(3)\}}\geq 1-\epsilon_{\pi(1)}-\epsilon_{\pi(3)}\tag{313}$$

and

$$\tilde{\sigma}_{\{1,2,3\}}+\tilde{\sigma}_{\{\pi(1),\pi(2)\}}+\tilde{\sigma}_{\{\pi(1),\pi(3)\}}+\tilde{\sigma}_{\{\pi(1)\}}=1-\epsilon_{\pi(1)}.\tag{314}$$

The existence of the desired numbers can be checked by applying the Fourier-Motzkin Elimination algorithm [45] and by noting Constraints (116). Further choose for any set $\mathcal{I}$ containing $\pi(1)$ and $\ell\in\{1,2,3\}$ the rate:

$$\tilde{R}_{\mathcal{I},\ell}:=R_{\mathcal{I},\ell},\tag{315}$$

and for any set $\mathcal{I}$ not containing $\pi(1)$ and $\ell\in\{1,2,3\}$:

$$\tilde{\sigma}_{\mathcal{I}}:=\sigma_{\mathcal{I}}+\sigma_{\mathcal{I}_{\pi(1)}}-\tilde{\sigma}_{\mathcal{I}_{\pi(1)}}\tag{316}$$

$$\tilde{R}_{\mathcal{I},\ell}:=\frac{\sigma_{\mathcal{I}}}{\tilde{\sigma}_{\mathcal{I}}}R_{\mathcal{I},\ell}+\frac{\sigma_{\mathcal{I}_{\pi(1)}}-\tilde{\sigma}_{\mathcal{I}_{\pi(1)}}}{\tilde{\sigma}_{\mathcal{I}}}R_{\mathcal{I}_{\pi(1)},\ell},\tag{317}$$

where we defined $\mathcal{I}_{\pi(1)}:=\mathcal{I}\cup\{\pi(1)\}$.

By Lemma 3, the new set of numbers $\{\tilde{\sigma}_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{P}(3)}$, and $\{\tilde{R}_{\mathcal{I},1},\dots,\tilde{R}_{\mathcal{I},\ell^*_{\mathcal{I}}}\}_{\mathcal{I}\in\mathcal{P}(3)}$ also satisfies Constraints (116), which proves that one can restrict to numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{P}(3)}$ satisfying (368). Since $\epsilon_{\pi(1)}\geq\epsilon_{\pi(2)}$ and

$$\sigma_{\{1,2,3\}}+\sigma_{\{\pi(1),\pi(2)\}}+\sigma_{\{\pi(2),\pi(3)\}}+\sigma_{\{\pi(2)\}}\geq 1-\epsilon_{\pi(2)},\tag{318}$$

this further implies that one can restrict to numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{P}(3)}$ satisfying

$$\sigma_{\{\pi(2),\pi(3)\}}\geq\sigma_{\{\pi(1),\pi(3)\}}+\sigma_{\{\pi(1)\}}-\sigma_{\{\pi(2)\}}\tag{319}$$

$$\geq\sigma_{\{\pi(1),\pi(3)\}}-\sigma_{\{\pi(2)\}}-\sigma_{\{\pi(1),\pi(2)\}}.\tag{320}$$

We next show that one can further restrict to nonnegative numbers satisfying also (369). To this end, assume that (369) is violated and define

$$a:=\tilde{\sigma}_{\{\pi(1),\pi(3)\}}-\tilde{\sigma}_{\{\pi(2)\}}-\tilde{\sigma}_{\{\pi(1),\pi(2)\}}>0.\tag{321}$$

Define also the new parameters

$$\sigma'_{\{1,2,3\}}:=\tilde{\sigma}_{\{1,2,3\}}+a\tag{322}$$

$$\sigma'_{\{\pi(3)\}}:=\tilde{\sigma}_{\{\pi(3)\}}+a\tag{323}$$

$$\sigma'_{\{\pi(1),\pi(3)\}}:=\tilde{\sigma}_{\{\pi(1),\pi(3)\}}-a\tag{324}$$

$$\sigma'_{\{\pi(2),\pi(3)\}}:=\tilde{\sigma}_{\{\pi(2),\pi(3)\}}-a\tag{325}$$

$$\sigma'_{\mathcal{I}}:=\tilde{\sigma}_{\mathcal{I}},\quad\pi(3)\notin\mathcal{I},\tag{326}$$

and the new rates

$$R'_{\{1,2,3\},\ell}=\frac{a\left(\lambda_\ell\tilde{R}_{\{\pi(1),\pi(3)\},\ell}+(1-\lambda_\ell)\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right)}{\sigma'_{\{1,2,3\}}}+\frac{\tilde{\sigma}_{\{1,2,3\}}\tilde{R}_{\{1,2,3\},\ell}}{\sigma'_{\{1,2,3\}}},\qquad\ell\in\{1,2,3\},\tag{327a}$$

$$R'_{\{\pi(3)\},\ell}=\frac{a\left((1-\lambda_\ell)\tilde{R}_{\{\pi(1),\pi(3)\},\ell}+\lambda_\ell\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right)}{\sigma'_{\{\pi(3)\}}}+\frac{\tilde{\sigma}_{\{\pi(3)\}}\tilde{R}_{\{\pi(3)\},\ell}}{\sigma'_{\{\pi(3)\}}},\qquad\ell\in\{1,\dots,\pi(3)\}\tag{327b}$$

$$R'_{\mathcal{I},\ell}=\tilde{R}_{\mathcal{I},\ell},\quad\mathcal{I}\in\mathcal{P}(3)\backslash\{\{1,2,3\},\{\pi(3)\}\}.\tag{327c}$$

Notice that by the definition of a and by (320), the parameters $\{\sigma'_{\mathcal{I}}\}$ are all nonnegative, and it is easily verified that they continue to satisfy (116) for any choice of $\lambda_1,\lambda_2,\lambda_3\in[0,1]$.

We next choose the parameters $\lambda_1,\lambda_2,\lambda_3\in[0,1]$ in function of the rates $\{\tilde{R}_{\mathcal{I},\ell}\}$ and the ordering $\pi(\cdot)$, and show that for the proposed choice of rates in (327), the exponents $\theta_1,\theta_2,\theta_3$ are only increased. We distinguish three cases.

For notational simplicity we assume $\pi(1)<\pi(2)$. (The proof for $\pi(1)>\pi(2)$ is analogous.) This implies that

$$1=\pi(1)<\pi(3)\qquad\text{or}\qquad 1=\pi(3)<\pi(2)=2\tag{328}$$

and

$$2=\pi(2)<\pi(3)=3\qquad\text{or}\qquad\pi(3)<\pi(2)=3.\tag{329}$$

Case 1: If

$$\eta_1\left(\tilde{R}_{\{\pi(1),\pi(3)\},1}\right)\leq\eta_1\left(\tilde{R}_{\{\pi(2),\pi(3)\},1}\right),\tag{330}$$

choose

$$\lambda_\ell=0,\qquad\ell\in\{1,\dots,\pi(3)\},\tag{331}$$

$$\lambda_\ell=\mathbb{1}\left\{\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\geq\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right\},\qquad\ell\in\{\pi(3)+1,\dots,3\}.\tag{332}$$

Using the same proof steps as in Lemma 3, it can be shown that for this choice of the λs the new rates in (327) still satisfy Constraint (116a) for $\theta_{\pi(3)}$ because $\lambda_1=\dots=\lambda_{\pi(3)}$.

To see that they satisfy (116a) also for $\theta_{\pi(1)}$, notice that:

$$\begin{aligned}&\min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\}\\&\leq\min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right),\ \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right)+\frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\}\\&\leq\min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right),\ \frac{a}{\sigma'_{\{1,2,3\}}}\eta_1\left(\tilde{R}_{\{\pi(2),\pi(3)\},1}\right)\right.\end{aligned}\tag{333}$$

$$+\frac{a}{\sigma'_{\{1,2,3\}}}\mathbb{1}\{\pi(1)=2\} \cdot \max\left\{\eta_2\left(\tilde{R}_{\{\pi(1),\pi(3)\},2}\right), \eta_2\left(\tilde{R}_{\{\pi(2),\pi(3)\},2}\right)\right\} + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\Bigg\} \tag{334}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(R'_{\{\pi(1),\pi(3)\},\ell}\right), \sum_{\ell=1}^{\pi(1)}\eta_\ell\left(R'_{\{1,2,3\},\ell}\right)\right\}. \tag{335}$$

where the second inequality holds by Assumption (330) and the third inequality holds by the definitions of the rates $\{R'_{\{1,2,3\},\ell}\}$ and by the concavity and monotonicity of the functions $\{\eta_\ell(\cdot)\}$.

Similarly, we notice for $\theta_{\pi(2)}$:

$$\min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\} \tag{336}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\min\{\pi(2),\pi(3)\}}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right) + \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=\pi(3)+1}^{\pi(2)}\max\left\{\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right), \eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right)\right\} + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\} \tag{337}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(R'_{\{\pi(2),\pi(3)\},\ell}\right), \sum_{\ell=1}^{\pi(2)}\eta_\ell\left(R'_{\{1,2,3\},\ell}\right)\right\}, \tag{338}$$

where notice that the sum in the second line of (337) is empty when $\pi(2) \leq \pi(3)$. Here, the last inequality holds by the definitions of the rates $\{R'_{\{1,2,3\},\ell}\}$ and by the choice of the λs and the concavity and monotonicity of the functions $\{\eta_\ell(\cdot)\}_\ell$.

Case 2: If

$$\sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right) \leq \sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right), \tag{339}$$

choose

$$\lambda_\ell = 1, \qquad \ell \in \{1,\ldots,\max\{2,\pi(3)\}\}, \tag{340}$$

$$\lambda_\ell = \mathbb{1}\left\{\tilde{R}_{\{\pi(1),\pi(3)\},\ell} \geq \tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right\}, \qquad \ell \in \{\max\{2,\pi(3)\}+1,\ldots,3\}. \tag{341}$$

Using similar arguments as in the previous case, one can conclude that the new rates in (327) still satisfy (116a). More specifically, since $\lambda_1 = \cdots = \lambda_{\pi(3)} = 1$ by (340), similar proof steps as in Lemma 3 can be used to show that (116a) holds for $\theta_{\pi(3)}$.

To see that (116a) holds for $\theta_{\pi(2)}$, recall that $\pi(2) \geq 2$ and notice:

$$\min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\} \tag{342}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right) + \frac{a}{\sigma'_{\{1,2,3\}}}\mathbb{1}\{\pi(2)=3\} \cdot \max\left\{\eta_3\left(\tilde{R}_{\{\pi(1),\pi(3)\},3}\right), \eta_3\left(\tilde{R}_{\{\pi(2),\pi(3)\},3}\right)\right\} + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\} \tag{343}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(R'_{\{\pi(2),\pi(3)\},\ell}\right), \sum_{\ell=1}^{\pi(2)}\eta_\ell\left(R'_{\{1,2,3\},\ell}\right)\right\}, \tag{344}$$

where the second inequality holds by our assumption (339) and since $\pi(2) \geq 2$.

Finally, (116a) holds for $\theta_{\pi(1)}$, because:

$$\min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right), \sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right)\right.$$

$$+\frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\Bigg\} \tag{345}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(R'_{\{\pi(1),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(1)}\eta_\ell\left(R'_{\{1,2,3\},\ell}\right)\right\}. \tag{346}$$

where the second inequality holds by the assumption $\pi(1) < \pi(2)$ and thus $\pi(1) \leq 2$.

Case 3: Else, i.e., if

$$\eta_1\left(\tilde{R}_{\{\pi(1),\pi(3)\},1}\right) > \eta_1\left(\tilde{R}_{\{\pi(2),\pi(3)\},1}\right) \tag{347}$$

and

$$\sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right) < \sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right). \tag{348}$$

Choose

$$\lambda_1 = 1, \quad \lambda_2 = \lambda, \quad \text{and} \quad \lambda_3 = 0, \tag{349}$$

for a value of $\lambda \in [0,1]$ so that the auxiliary rates

$$\bar{R}_{\{\pi(1),\pi(3)\},2} := \lambda R_{\{\pi(1),\pi(3)\},2} + (1-\lambda)R_{\{\pi(2),\pi(3)\},2} \tag{350}$$

$$\bar{R}_{\{\pi(2),\pi(3)\},2} := (1-\lambda)R_{\{\pi(1),\pi(3)\},2} + \lambda R_{\{\pi(2),\pi(3)\},2} \tag{351}$$

satisfy

$$\eta_1\left(\tilde{R}_{\{\pi(1),\pi(3)\},1}\right) + \eta_2\left(\bar{R}_{\{\pi(1),\pi(3)\},2}\right) = \sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right) \tag{352}$$

$$\eta_1\left(\tilde{R}_{\{\pi(2),\pi(3)\},1}\right) + \eta_2\left(\bar{R}_{\{\pi(2),\pi(3)\},2}\right) \geq \sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right). \tag{353}$$

Existence of the desired choice of λ can be seen as follows. Notice first that for $\lambda = 0$, relation (352) holds with a $>$ sign because of Assumption (347). For $\lambda = 1$, relation (352) holds with a $<$ sign because of Assumption (348). By the continuity of the functions $\{\eta_\ell(\cdot)\}$ and the intermediate value theorem, there is thus a value $\lambda \in (0,1)$ such that (352) holds with equality. Let λ be this value and notice that by the concavity of the functions $\{\eta_\ell(\cdot)\}$:

$$\eta_1\left(\tilde{R}_{\{\pi(1),\pi(3)\},1}\right) + \eta_2\left(\bar{R}_{\{\pi(1),\pi(3)\},2}\right) + \eta_1\left(\tilde{R}_{\{\pi(2),\pi(3)\},1}\right) + \eta_2\left(\bar{R}_{\{\pi(2),\pi(3)\},2}\right) \geq \sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right) + \sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \tag{354}$$

which combined with (352) implies (353).

Now that we established the existence of the desired value λ, we continue to show that for the choice in (349), Constraints (116a) remain valid. For $\theta_{\pi(1)}$ this can be verified through the following steps, where recall that $\pi(1) \leq 2$:

$$\min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\}$$

$$\leq \min\Bigg\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\Bigg\} \tag{355}$$

$$= \min\Bigg\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\Big(\eta_1\left(\tilde{R}_{\{\pi(1),\pi(3)\},1}\right) + \mathbb{1}\{\pi(1)=2\}\cdot\eta_2\left(\bar{R}_{\{\pi(1),\pi(3)\},2}\right)\Big) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\Bigg\} \tag{356}$$

$$\leq \min\left\{\sum_{\ell=1}^{\pi(1)}\eta_\ell\left(R'_{\{\pi(1),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(1)}\eta_\ell\left(R'_{\{1,2,3\},\ell}\right)\right\}, \tag{357}$$

where the second inequality holds since $\pi(1) \leq 2$, and by (348) and (352), and the last inequality holds by the definitions of the rates $\{R'_{\{1,2,3\},\ell}\}$, the choice of the λs, and the concavity and monotonicity of the functions $\{\eta_\ell(\cdot)\}$.

To verify that Constraint (116a) remains valid for $\theta_{\pi(2)}$, recall that $\pi(2) \geq 2$ and notice:

$$\min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\}$$

$$\leq \min\Bigg\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\Bigg\} \tag{358}$$

$$= \min\Bigg\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right), \frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{2}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right) + \frac{a}{\sigma'_{\{1,2,3\}}}\mathbb{1}\{\pi(2)=3\}\cdot\eta_3\left(\tilde{R}_{\{\pi(2),\pi(3)\},3}\right) + \frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\Bigg\} \tag{359}$$

$$= \min\Bigg\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right),$$

$$\frac{a}{\sigma'_{\{1,2,3\}}}\left(\eta_1\left(\tilde{R}_{\{\pi(1),\pi(3)\},1}\right)+\eta_2\left(\bar{R}_{\{\pi(1),\pi(3)\},2}\right)\right)$$
$$+\frac{a}{\sigma'_{\{1,2,3\}}}\mathbb{1}\left\{\pi(2)=3\right\}\cdot\eta_3\left(\tilde{R}_{\{\pi(2),\pi(3)\},3}\right)$$
$$\left.+\frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\} \tag{360}$$
$$\leq \min\left\{\sum_{\ell=1}^{\pi(2)}\eta_\ell\left(R'_{\{\pi(2),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(2)}\eta_\ell\left(R'_{\{1,2,3\},\ell}\right)\right\}. \tag{361}$$

Here, the second equality holds by (352).

Finally, to see that Constraint (116a) is also satisfied for $\theta_{\pi(3)}$, we distinguish two cases. If $\pi(3)=1$, the proof is similar to the proof of Lemma 3 because $\lambda_1=1$. For the proof in the case $\pi(3)\geq 2$, notice first:

$$\min\left\{\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\}$$
$$\leq \min\left\{\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right),\right.$$
$$\frac{a}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right)$$
$$\left.+\frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\} \tag{362}$$
$$=\min\left\{\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right),\right.$$
$$\frac{a}{\sigma'_{\{1,2,3\}}}\eta_1\left(\tilde{R}_{\{\pi(1),\pi(3)\},1}\right)$$
$$+\frac{a}{\sigma'_{\{1,2,3\}}}\eta_2\left(\bar{R}_{\{\pi(1),\pi(3)\},2}\right)$$
$$+\frac{a}{\sigma'_{\{1,2,3\}}}\mathbb{1}\{\pi(3)=3\}\ .\ \eta_\ell\left(\tilde{R}_{\{\pi(2),\pi(3)\},\ell}\right)$$
$$\left.+\frac{\tilde{\sigma}_{\{1,2,3\}}}{\sigma'_{\{1,2,3\}}}\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\} \tag{363}$$
$$\leq \min\left\{\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(R'_{\{\pi(2),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(3)}\eta_\ell\left(R'_{\{1,2,3\},\ell}\right)\right\}, \tag{364}$$

where the equality holds by Assumption (352).

Notice further that

$$\min\left\{\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(3)\},\ell}\right)\right\}$$
$$\leq \min\left\{\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right),\right.$$
$$\frac{a}{\sigma'_{\{\pi(3)\}}}\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right)$$
$$\left.+\frac{\tilde{\sigma}_{\{\pi(3)\}}}{\sigma'_{\{\pi(3)\}}}\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(3)\},\ell}\right)\right\} \tag{365}$$
$$\leq \min\left\{\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{\pi(1),\pi(3)\},\ell}\right),\right.$$
$$\frac{a}{\sigma'_{\{\pi(3)\}}}\eta_1\left(\tilde{R}_{\{\pi(2),\pi(3)\},1}\right)$$
$$+\frac{a}{\sigma'_{\{\pi(3)\}}}\eta_2\left(\bar{R}_{\{\pi(2),\pi(3)\},2}\right)$$
$$+\frac{a}{\sigma'_{\{\pi(3)\}}}\mathbb{1}\{\pi(3)=3\}\ .\ \eta_3\left(\tilde{R}_{\{\pi(1),\pi(3)\},3}\right)$$
$$\left.+\frac{\tilde{\sigma}_{\{\pi(3)\}}}{\sigma'_{\{\pi(3)\}}}\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(\tilde{R}_{\{1,2,3\},\ell}\right)\right\} \tag{366}$$
$$\leq \min\left\{\sum_{\ell=1}^{\pi(3)}\eta_\ell\left(R'_{\{\pi(1),\pi(3)\},\ell}\right),\ \sum_{\ell=1}^{\pi(3)}\eta_\ell\left(R'_{\{\pi(3)\},\ell}\right)\right\}, \tag{367}$$

where the second inequality holds by (353).

Appendix G
Converse Prove to Proposition 7

We start with two auxiliary lemmas.

Lemma 5: Let $K=3$. In Theorem 5 it suffices to consider values $\{\sigma_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{P}(3)}$ so that

$$\sigma_{\{1,2,3\}}+\sigma_{\{\pi(1),\pi(2)\}}+\sigma_{\{\pi(1),\pi(3)\}}+\sigma_{\{\pi(1)\}}=1-\epsilon_{\pi(1)} \tag{368}$$
$$\sigma_{\{\pi(1),\pi(2)\}}+\sigma_{\{\pi(2)\}}\geq\sigma_{\{\pi(1),\pi(3)\}} \tag{369}$$

Proof: See Appendix F. ■

We thus continue with nonnegative numbers $\{\sigma_{\mathcal{I}}\}_{\mathcal{I}\in\mathcal{P}(3)}$, and $\{R_{\mathcal{I},1},\ldots,R_{\mathcal{I},\ell^*_{\mathcal{I}}}\}_{\mathcal{I}\in\mathcal{P}(3)}$ satisfying (116) for $K=3$ as well as (369). The proof of the desired proposition follows by the next lemma (which holds for any positive integer K) and by an appropriate choice of parameters $\{c_{\mathcal{J}}\}$, see (390) ahead.

Lemma 6: Let

$$\{c_{\mathcal{J}}\colon\ \mathcal{J}\in\mathcal{P}(K)\}, \tag{370}$$
$$\{\delta_{\mathcal{I},\mathcal{J}}\colon\ \mathcal{I},\mathcal{J}\in\mathcal{P}(K)\text{ and }\mathcal{I}\cap\mathcal{J}\neq\emptyset\} \tag{371}$$

be sets of nonnegative integers satisfying

$$\sum_{\substack{\mathcal{J}\in\mathcal{P}(K):\\ \mathcal{I}\cap\mathcal{J}\neq\emptyset}}\delta_{\mathcal{I},\mathcal{J}}\leq\sigma_{\mathcal{I}},\quad \mathcal{I}\in\mathcal{P}(K), \tag{372}$$

and

$$\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}}\delta_{\mathcal{I},\mathcal{J}}\geq c_{\mathcal{J}},\qquad \forall k\in\mathcal{J},\ \mathcal{J}\in\mathcal{P}(K). \tag{373}$$

Then, the rates

$$\tilde{R}_{\mathcal{J},k} := \max_{\substack{j\in\mathcal{J}:\\ j\geq k}} \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ j\in\mathcal{I}}} \frac{\delta_{\mathcal{I},\mathcal{J}}}{c_{\mathcal{J}}} R_{\mathcal{I},k}, \qquad k\leq \ell^*_{\mathcal{J}}, \mathcal{J}\in\mathcal{P}(K). \quad (374)$$

satisfy the following inequalities:

$$\theta_k \leq \min_{\substack{\mathcal{J}\in\mathcal{P}(K):\\ k\in\mathcal{J}}} \sum_{\ell=1}^{k} \eta_\ell\left(\tilde{R}_{\mathcal{J},\ell}\right), \qquad k\in\{1,\ldots,K\}, \quad (375)$$

and

$$R_k \geq \sum_{\substack{\mathcal{J}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{J}}}} c_{\mathcal{J}}\cdot \tilde{R}_{\mathcal{J},k}, \qquad k\in\{1,\ldots,K\}. \quad (376)$$

Proof: We start by proving (375). By (116a), for any $k\in\{1,\ldots,K\}$ and any set $\mathcal{J}\subseteq\mathcal{P}(K)$ containing index k:

$$\begin{aligned}
\theta_k &\leq \min_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}} \sum_{\ell=1}^{k}\eta_\ell(R_{\mathcal{I},\ell}) && (377)\\
&\leq \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}} \frac{\delta_{\mathcal{I},\mathcal{J}}}{\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}}\delta_{\mathcal{I},\mathcal{J}}}\cdot\sum_{\ell=1}^{k}\eta_\ell(R_{\mathcal{I},\ell}) && (378)\\
&\leq \sum_{\ell=1}^{k}\eta_\ell\left(\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}} \frac{\delta_{\mathcal{I},\mathcal{J}}}{\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}}\delta_{\mathcal{I},\mathcal{J}}}\cdot R_{\mathcal{I},\ell}\right) && (379)\\
&\leq \sum_{\ell=1}^{k}\eta_\ell\left(\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}} \frac{\delta_{\mathcal{I},\mathcal{J}}}{c_{\mathcal{J}}}\cdot R_{\mathcal{I},\ell}\right) && (380)\\
&\leq \sum_{\ell=1}^{k}\eta_\ell\left(\tilde{R}_{\mathcal{J},\ell}\right), && (381)
\end{aligned}$$

where (378) holds because the minimum of a set of numbers is never larger than any convex combination of these numbers; (379) holds by the concavity of the functions $\eta_1(\cdot),\ldots,\eta_k(\cdot)$; (380) holds by assumption (373) and by the monotonicity of the functions $\eta_1(\cdot),\ldots,\eta_k(\cdot)$; and (381) holds by the definition of $\tilde{R}_{\mathcal{J},k}$ in (374) because $k\geq\ell$ and $k\in\mathcal{J}$ thus $\ell\leq\ell^*_{\mathcal{J}}$.

To prove (376), fix $k\in\{1,\ldots,K\}$ and for each subset $\mathcal{J}\subseteq\mathcal{P}(K)$ with $\ell^*_{\mathcal{J}}\geq k$ pick an index $j_{\mathcal{J}}\in\mathcal{J}$ so that $j_{\mathcal{J}}\geq k$. Then, by (116b):

$$\begin{aligned}
R_k &\geq \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{I}}}} \sigma_{\mathcal{I}}\cdot R_{\mathcal{I},k}, && (382)\\
&\geq \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{I}}}} \sum_{\substack{\mathcal{J}\in\mathcal{P}(K):\\ \mathcal{I}\cap\mathcal{J}\neq\emptyset}} \delta_{\mathcal{I},\mathcal{J}}\cdot R_{\mathcal{I},k} && (383)\\
&= \sum_{\mathcal{J}\in\mathcal{P}(K)} \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{I}}\\ \mathcal{I}\cap\mathcal{J}\neq\emptyset}} \delta_{\mathcal{I},\mathcal{J}}\cdot R_{\mathcal{I},k} && (384)\\
&\geq \sum_{\substack{\mathcal{J}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{J}}}} \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{I}}\\ \mathcal{I}\cap\mathcal{J}\neq\emptyset}} \delta_{\mathcal{I},\mathcal{J}}\cdot R_{\mathcal{I},k} && (385)\\
&\geq \sum_{\substack{\mathcal{J}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{J}}}} \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{I}}\\ j_{\mathcal{J}}\in\mathcal{I}}} \delta_{\mathcal{I},\mathcal{J}} R_{\mathcal{I},k} && (386)\\
&= \sum_{\substack{\mathcal{J}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{J}}}} \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ j_{\mathcal{J}}\in\mathcal{I}}} \delta_{\mathcal{I},\mathcal{J}} R_{\mathcal{I},k} && (387)\\
&= \sum_{\substack{\mathcal{J}\in\mathcal{P}(K):\\ k\leq\ell^*_{\mathcal{J}}}} c_{\mathcal{J}}\cdot\frac{\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ j_{\mathcal{J}}\in\mathcal{I}}}\delta_{\mathcal{I},\mathcal{J}}R_{\mathcal{I},k}}{c_{\mathcal{J}}}, && (388)
\end{aligned}$$

where (383) holds by Assumption (372); inequalities (385) and (386) hold because we consider less summands and each summand is nonnegative (recall that $j_{\mathcal{J}}\in\mathcal{J}$); and finally (g) holds because the two conditions $j_{\mathcal{J}}\geq k$ and $j_{\mathcal{J}}\in\mathcal{I}$ imply that $\ell^*_{\mathcal{I}}\geq k$.

The proof of the lemma is concluded by recalling the definition of rate $\tilde{R}_{\mathcal{J},k}$ in (374) and noting that Inequality (381) holds for any set $\mathcal{J}$ containing k whereas Inequality (388) holds for any index $j_{\mathcal{J}}\in\mathcal{J}$ larger than k. ■

To obtain the desired simplification in Proposition 7 from Theorem 5, define the subsets

$$\mathcal{J}_k := \{\pi(k),\ldots,\pi(K)\}, \quad k\in\{1,\ldots,K\}, \quad (389)$$

and the values $\pi(0):=0$ and $\epsilon_0:=1$. Applying above Lemma 6 to the choice

$$c_{\mathcal{J}} := \begin{cases}\epsilon_{\pi(k-1)}-\epsilon_{\pi(k)}, & \mathcal{J}=\mathcal{J}_k,\\ 0, & \text{otherwise},\end{cases} \quad (390)$$

establishes the converse to Conjecture 6 for general values of K, if one renames rates $\tilde{R}_{\mathcal{J}_k,\ell}$ as $\tilde{R}_{k,\ell}$. The proof is concluded by showing that above parameter choice is permissible, i.e., that there exist nonnegative numbers $\{\delta_{\mathcal{I},\mathcal{J}}\}$ satisfying conditions (372) and (373) for $\{c_{\mathcal{J}}\}$ in (390). For general values of K this seems cumbersome.

For $K=3$, this can be achieved by means of the Fourier-Motzkin Elimination algorithm [45], which shows the existence of nonnegative numbers $\{\delta_{\mathcal{I},\mathcal{J}}\}$ satisfying conditions (372) and (373) for $\{c_{\mathcal{J}}\}$ in (390), whenever (redundant conditions are omitted)

$$\sigma_{\{1,2,3\}}+\sigma_{\{\pi(1),\pi(2)\}}+\sigma_{\{\pi(1),\pi(3)\}}+\sigma_{\{\pi(1)\}}\geq 1-\epsilon_{\pi(1)} \quad (391a)$$

$$\sigma_{\{1,2,3\}}+\sigma_{\{\pi(1),\pi(2)\}}+\sigma_{\{\pi(2),\pi(3)\}}+\sigma_{\{\pi(2)\}}\geq 1-\epsilon_{\pi(2)} \quad (391b)$$

$$\sigma_{\{1,2,3\}}+\sigma_{\{\pi(1),\pi(3)\}}+\sigma_{\{\pi(2),\pi(3)\}}+\sigma_{\{\pi(3)\}}\geq 1-\epsilon_{\pi(3)} \quad (391c)$$

and

$$\begin{aligned}
&2\sigma_{\{1,2,3\}}+2\sigma_{\{\pi(1),\pi(2)\}}+\sigma_{\{\pi(1),\pi(3)\}}+\sigma_{\{\pi(2),\pi(3)\}}\\
&\quad+\sigma_{\{\pi(1)\}}+\sigma_{\{\pi(2)\}}+\sigma_{\{\pi(3)\}}\geq 1-\epsilon_{\pi(1)}+1-\epsilon_{\pi(3)}.
\end{aligned} \quad (391d)$$

Since Conditions (391a)–(391c) are satisfied by Assumption (116c) and Condition (391d) is implied by (391a), (391b), and (369), this concludes the proof for $K=3$ and thus establishes Proposition 7.

Appendix H
Converse for the case $\epsilon_1 = \cdots = \epsilon_K$

Define

$$\phi_k \triangleq \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}} \sigma_{\mathcal{I}} \tag{392}$$

and notice that by (116c):

$$\phi_k \geq 1-\epsilon_k = 1-\epsilon. \tag{393}$$

By (116a) we have for any $k \in \{1,\ldots,K\}$:

$$\theta_k \leq \min_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}} \sum_{\ell=1}^{k} \eta_\ell(R_{\mathcal{I},\ell}), \tag{394}$$

$$\leq \sum_{\ell=1}^{k} \sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}} \frac{\sigma_I}{\phi_k}\eta_\ell(R_{\mathcal{I},\ell}) \tag{395}$$

$$\leq \sum_{\ell=1}^{k} \eta_\ell\left(\sum_{\substack{\mathcal{I}\in\mathcal{P}(K):\\ k\in\mathcal{I}}} \frac{\sigma_I}{\phi_k} R_{\mathcal{I},\ell}\right) \tag{396}$$

$$\leq \sum_{\ell=1}^{k} \eta_\ell\left(\frac{R_\ell}{\phi_k}\right) \tag{397}$$

$$\leq \sum_{\ell=1}^{k} \eta_\ell\left(\frac{R_\ell}{1-\epsilon}\right), \tag{398}$$

where (395) holds because the minimum of a set of numbers is smaller than any convex combination thereof; (396) holds by the concavity of the functions $\eta_\ell(\cdot)$; (397) holds by the rate constraints (116b); and (398) holds by (393). This establishes the desired converse. Achievability follows by setting $\sigma_{\{1,\ldots,K\}} = 1-\epsilon$ and all other $\sigma_{\mathcal{I}} = 0$ in (116a).

Mustapha Hamad (S'14–M'22) received the B.E. degree in Computer Engineering and the M.Sc. degree in Computer and Communication Engineering, with distinction, from the Lebanese American University (LAU), Byblos, Lebanon, in 2015 and 2018, respectively, and received the Ph.D. degree in Networks, Information, and Communications from Télécom Paris, Institut Polytechnique de Paris, France, in 2022. Between 2016 and 2019, he was a Systems and Security Engineer at Computer Business Machines (CBM), a partner company of IBM in Lebanon, as well as a Research Assistant at LAU. Since November 2022, he is a Senior Research Engineer at Huawei Mathematical and Algorithmic Sciences Lab, Paris Research Center, France. His research interests are in information theory, mainly distributed hypothesis testing, and in communication theory, semantic communications, and explainable machine learning.

Mireille Sarkiss (S'05–M'09) received the engineering degree in Telecommunications and Computer Science from the Lebanese University, Faculty of Engineering, Lebanon, in 2003 and the M.Sc. and Ph.D. degrees in Communications and Electronics from Télécom Paris, France, in 2004 and 2009, respectively. She was a doctoral researcher at Orange Labs from 2004 to 2007 and a post-doctoral researcher at Télécom Paris, from 2009 to 2010. From 2010 to 2018, she was a Researcher at CEA LIST, Communicating Systems Laboratory, France. Since December 2018, she is an Associate Professor at Télécom SudParis, Institut Polytechnique de Paris, France. Her research interests are in communication theory and information theory, including resource allocation, optimization and reinforcement learning for mobile-edge computing and energy-efficient wireless networks, physical layer security, and distributed hypothesis testing.

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 Télécom Paris, France, where she is currently a full professor. She has held visiting professor appointments at the Technion–Israel Institute of Technology, at the University of Zurich, and at ETH Zurich. Michèle Wigger has previously served as an Associate Editor of the IEEE Communication Letters and is currently serving 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. Her research interests are in multi-terminal information theory, in particular in distributed source coding and in capacities of networks, and in the limits of distributed hypothesis testing.