Benefits of Rate-Sharing for Distributed Hypothesis Testing

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

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

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

***Abstract*—We study distributed binary hypothesis testing with a single sensor and two remote decision centers that are also equipped with local sensors. The communication between the sensor and the two decision centers takes place over three links: a shared link to both centers and an individual link to each of the two centers. All communication links are subject to *expected* rate constraints. This paper characterizes the optimal exponents region of the type-II error for given type-I error thresholds at the two decision centers and further simplifies the expressions in the special case of having only the single shared link. The exponents region illustrates a gain under expected rate constraints compared to equivalent maximum rate constraints. Moreover, it exhibits a tradeoff between the exponents achieved at the two centers.**

***Index Terms*—Broadcast channel, distributed hypothesis testing, error exponents, expected rate constraints, IoT, decision centers.**

I. Introduction

We address a distributed hypothesis testing problem where different decision centers have to decide on the same hypothesis based on their local sensing and the messages they receive from remote sensors over rate-limited communication links. Motivated by systems that share bandwidth among several applications with variable instantaneous bandwidth for each application, we consider *expected-rate constraints* that limit only the expected bandwidth for each application.

In our work, we focus on distributed binary hypothesis testing against independence. The decision centers have to decide between a *i) null hypothesis* (normal situation) indicating that the centers' and the sensors' observations are correlated, and an *ii) alternative hypothesis* (alert situation) where the observations are independent, for example because one of the systems fails. Two types of errors can be distinguished: the *type-I error* indicates a wrong decision under the null hypothesis and the *type-II error* occurs if a wrong decision is made under the alternative hypothesis. Since the alternative hypothesis corresponds to a more critical situation, we aim at maximizing the exponential decay of the type-II error probability, called *error exponent*, subject to a type-I error that stays below a given threshold. Such a setup has been studied in many previous works focusing mostly on *maximum-rate constraints* [1]–[16]. *Expected-rate constraints* were introduced in [17], where the maximum error exponent for single-sensor single-decision center setup was characterized in the special case of testing-against independence. Extensions of this work were first proposed for a multi-sensor scenario in [18], for a multi-hop scenario with multiple decision centers in [19], [20], for distributed sequential hypothesis testing with zero-rate [21], and most recently from a signal detection perspective in [22].

In this paper, we consider a single-sensor two-decision center scenario where the decision centers also have sensing capabilities. The communication takes place over three noise-free links: a common link to both decision centers and one private link to each decision center. For this one-to-many broadcast setup, we characterize the optimal exponents region under expected-rate constraints and we show that it improves over the exponents region under maximum-rate constraints, which we also establish in this paper. The optimal exponents region under expected-rate constraints illustrates two tradeoffs. The first tradeoff results from the shared link that has to serve both decision centers at the same time; this tradeoff is also present under maximum-rate constraints. The second tradeoff is particular to the setup with expected-rate constraints and stems from the rate-sharing between three different variants of the optimal coding scheme under maximum-rate constraints in [8], depending on the observations at the sensor. We show that two variants suffice when communication is only over a single shared link, leading to a significant reduction in the complexity of the optimal coding scheme.

Notation: We follow the notation in [23], [17]. In particular, we use sans serif font for bit-strings: e.g., m for a deterministic and M for a random bit-string, and we denote the length of m by $\operatorname{len}(\mathsf{m})$. In addition, $\mathcal{T}_\mu^{(n)}(P)$ denotes the strongly μ-typical set with respect to P as defined in [24, Definition 2.8].

II. System Model

Consider the distributed hypothesis testing problem in Figure 1 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:

$$\text{under } \mathcal{H}=0 : (Y_0^n, Y_1^n, Y_2^n)\text{i.i.d.} \sim P_{Y_0} \cdot P_{Y_1Y_2|Y_0}; \tag{1a}$$
$$\text{under } \mathcal{H}=1 : (Y_0^n, Y_1^n, Y_2^n)\text{i.i.d.} \sim P_{Y_0} \cdot P_{Y_1Y_2} \tag{1b}$$

for given probability mass functions (pmfs) P_{Y_0} and $P_{Y_1Y_2|Y_0}$ and where $P_{Y_1Y_2}$ denotes the marginal of the joint pmf $P_{Y_0Y_1Y_2} := P_{Y_0}P_{Y_1Y_2|Y_0}$.

The system consists of a transmitter T_{Y_0}, and two receivers $\mathsf{R}_{Y_1}, \mathsf{R}_{Y_2}$. Transmitter T_{Y_0} observes the source sequence Y_0^n and computes three bit-string messages $(\mathsf{M}_0, \mathsf{M}_1, \mathsf{M}_2) = \phi^{(n)}(Y_0^n)$,

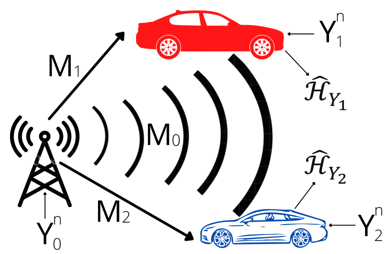

Fig. 1: Distributed hypothesis testing with a single sensor and two remote decision centers with integrated sensors.

where the encoding function is of the form $\phi^{(n)} : \mathcal{Y}_0^n \to \{0,1\}^\star \times \{0,1\}^\star \times \{0,1\}^\star$. Message M_0 is sent to both receivers $\mathrm{R}_{Y_1}, \mathrm{R}_{Y_2}$, while message M_1 only to receiver R_{Y_1} and message M_2 only to receiver R_{Y_2}. The messages have to satisfy the *expected-rate* constraints

$$\mathbb{E}\left[\mathrm{len}\left(\mathsf{M}_i\right)\right] \leq nR_i, \qquad i \in \{0,1,2\}. \tag{2}$$

Receiver $\mathrm{R}_{Y_i}, i \in \{1,2\}$, observes the source sequence Y_i^n and with messages $\mathsf{M}_0, \mathsf{M}_i$ received from T_{Y_0}, it produces a guess $\hat{\mathcal{H}}_{Y_i}$ of the hypothesis $\mathcal{H}$ using a decision function $g_i^{(n)} : \mathcal{Y}_i^n \times \{0,1\}^\star \times \{0,1\}^\star \to \{0,1\}$:

$$\hat{\mathcal{H}}_{Y_i} = g_i^{(n)}\left(Y_i^n, \mathsf{M}_0, \mathsf{M}_i\right) \in \{0,1\}, \qquad i \in \{1,2\}. \tag{3}$$

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

$$\alpha_{i,n} \triangleq \Pr[\hat{\mathcal{H}}_{Y_i} = 1 | \mathcal{H} = 0], \qquad i \in \{1,2\}, \tag{4}$$

stay below given thresholds $\epsilon_i > 0, i \in \{1,2\}$, and the type-II error probabilities

$$\beta_{i,n} \triangleq \Pr[\hat{\mathcal{H}}_{Y_i} = 0 | \mathcal{H} = 1] \tag{5}$$

decay to 0 with largest possible exponential decay.

Definition 1: Fix maximum type-I error probabilities $\epsilon_1, \epsilon_2 \in [0,1]$ and rates $R_1, R_2 \geq 0$. The exponent pair (θ_1, θ_2) is called (ϵ_1, ϵ_2)-*achievable* if there exists a sequence of encoding and decision functions $\{\phi^{(n)}, g_1^{(n)}, g_2^{(n)}\}_{n \geq 1}$ satisfying:

$$\mathbb{E}[\mathrm{len}(\mathsf{M}_i)] \leq nR_i, \quad i \in \{0,1,2\} \tag{6a}$$

$$\overline{\lim_{n \to \infty}}\, \alpha_{i,n} \leq \epsilon_i, \qquad i \in \{1,2\} \tag{6b}$$

$$\underline{\lim_{n \to \infty}}\, \frac{1}{n} \log \frac{1}{\beta_{i,n}} \geq \theta_i, \qquad i \in \{1,2\}. \tag{6c}$$

Definition 2: The closure of the set of all (ϵ_1, ϵ_2)-achievable exponent pairs (θ_1, θ_2) is called the (ϵ_1, ϵ_2)-*exponents region* and is denoted $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$.

III. Main Results

Our main results are a complete characterization of the exponents region $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ under the *expected-rate* constraints in (2) as well as a strong converse under analogous *maximum-rate* constraints. A simplified expression is provided for $\mathcal{E}^*(R_0, 0, 0, \epsilon_1, \epsilon_2)$.

A. Individual and Common Communication Links

Theorem 1: The (ϵ_1, ϵ_2)-exponents region $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ *is the set of* all (θ_1, θ_2) pairs satisfying

$$\theta_i \leq \min\left\{I\left(U_0^0 U_i^0; Y_i\right), I\left(U_0^i U_i^i; Y_i\right)\right\}, \qquad i \in \{1,2\} \tag{7a}$$

for some non-negative numbers $\sigma_0, \sigma_1, \sigma_2$ with sum ≤ 1 and conditional pmfs $P_{U_0^0|Y_0}$, $P_{U_0^1|Y_0}$, $P_{U_0^2|Y_0}$, $P_{U_1^0|U_0^0 Y_0}$, $P_{U_1^1|U_0^1 Y_0}$, $P_{U_2^0|U_0^0 Y_0}$, $P_{U_2^2|U_0^2 Y_0}$ satisfying

$$R_0 \geq \sigma_0 I(U_0^0; Y_0) + \sigma_1 I(U_0^1; Y_0) + \sigma_2 I(U_0^2; Y_0), \tag{7b}$$

$$R_i \geq \sigma_0 I(U_i^0; Y_0 | U_0^0) + \sigma_i I(U_i^i; Y_0 | U_0^i), \quad i \in \{1,2\}, \tag{7c}$$

$$\text{and} \quad \sigma_0 + \sigma_i \geq 1 - \epsilon_i, \quad i \in \{1,2\}, \tag{7d}$$

$$\sigma_0 \geq 1 - \epsilon_1 - \epsilon_2, \tag{7e}$$

and where the mutual information quantities are calculated according to the joint pmfs

$$P_{Y_0 Y_1 Y_2 U_0^0 U_1^0 U_2^0} \triangleq P_{Y_0 Y_1 Y_2} P_{U_0^0|Y_0} P_{U_1^0 U_2^0|U_0^0 Y_0} \tag{8}$$

$$P_{Y_0 Y_1 Y_2 U_0^i U_i^i} \triangleq P_{Y_0 Y_1 Y_2} P_{U_0^i|Y_0} P_{U_i^i|U_0^i Y_0}, \quad i \in \{1,2\}. \tag{9}$$

Proof: The converse is proved in Section IV. To prove achievability, define three sets $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2 \subseteq \mathcal{Y}_0^n$ with probabilities (under $P_{Y_0}^n$) equal to $\sigma_0, \sigma_1, \sigma_2$, respectively. For each set $\mathcal{D}_i$ ($i \in \{0,1,2\}$), we apply the optimal coding scheme under maximum rate constraints in [8], but for each set $\mathcal{D}_i$ we construct different codebooks and use different auxiliaries U_0^i, U_1^i, U_2^i. In particular, we choose U_1^2 and U_2^1 constants, indicating that when $Y_0^n \in \mathcal{D}_i$ then only messages $(\mathsf{M}_0, \mathsf{M}_i)$ are sent, for $i \in \{1,2\}$. When $Y_0^n \in \mathcal{D}_0$, then all three messages $\mathsf{M}_0, \mathsf{M}_1, \mathsf{M}_2$ are sent. For a detailed analysis, see [25]. ∎

Theorem 1 shows a tradeoff between the two achievable exponents θ_1 and θ_2. (Figure 2 ahead illustrates this tradeoff at hand of a numerical example in the special case $R_1 = R_2 = 0$.) The tradeoff stems from the common random variable U_0^0 that is included in the exponent constraint (7a) for both $i \in \{1,2\}$, and from the rate-sharing of the coding scheme in [8] for three different choices of $(\sigma_i, U_0^i, U_1^i, U_2^i)$, for $i = 0,1,2$.

To see the effect of the expected-rate constraint in (2), we compare above exponents region $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ with the exponents region $\mathcal{E}_{\text{fix}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ under more stringent maximum-length constraints

$$\mathrm{len}\left(\mathsf{M}_i\right) \leq nR_i, \qquad i \in \{0,1,2\}. \tag{10}$$

In the limit $\epsilon_1, \epsilon_2 \downarrow 0$, the exponents region $\mathcal{E}_{\text{fix}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ was determined in [8]. Here, we strengthen this result by providing a strong converse, whose proof follows similar steps (but with the expected rate replaced by the maximum rate) as the converse to Theorem 1.

Theorem 2: Under the maximum rate constraints (10), the exponents region $\mathcal{E}_{\text{fix}}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ *is independent of* (ϵ_1, ϵ_2) and equals the set of (θ_1, θ_2) pairs satisfying:

$$\theta_i \leq I(U_0 U_i; Y_i), \quad i \in \{1,2\}, \tag{11a}$$

for some conditional pmfs $P_{U_0|Y_0}$, $P_{U_i|Y_0}$ satisfying

$$R_0 \geq I(U_0; Y_0), \tag{11b}$$

$$R_i \geq I(U_i; Y_0|U_0), \quad i \in \{1,2\}. \tag{11c}$$

Proof: Achievability is proved in [8]. The converse is proved in [25]. ∎

Notice that (11) is obtained from (7) by setting $\sigma_0 = 1$ and $U_0^0, U_0^1, U_0^2, U_2^2$ constants. Moreover, $\mathcal{E}^*_{\text{fix}}(R_0, R_1, R_2, \epsilon_1, \epsilon_2) = \mathcal{E}^*(R_0, R_1, R_2, 0, 0)$. Since $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ is generally increasing in (ϵ_1, ϵ_2), expected-rate constraints allow to boost the exponents region compared to maximum-rate constraints.

B. Only a Common Communication Link

For $R_1 = R_2 = 0$, i.e., without individual communication links, we can simplify the expression for $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$.

Definition 3: Define the two functions

$$\eta_i\left(R_0^i\right) := \max_{\substack{P_{U_0^i|Y_0}: \\ R_0^i \geq I\left(U_0^i; Y_0\right)}} I\left(U_0^i; Y_i\right), \qquad i \in \{1,2\}, \tag{12}$$

where the mutual information quantities are calculated with respect to the joint pmf $P_{U_0^i Y_0 Y_1 Y_2} \triangleq P_{U_0^i|Y_0} P_{Y_0 Y_1 Y_2}$.

Corollary 1: Let $\pi\colon \{1,2\} \to \{1,2\}$ be a permutation ordering the ϵ-values in decreasing order:

$$\epsilon_{\pi(1)} \geq \epsilon_{\pi(2)}. \tag{13}$$

Then $\mathcal{E}^*(R_0, 0, 0, \epsilon_1, \epsilon_2)$ is the set of all (θ_1, θ_2) pairs satisfying

$$\theta_{\pi(1)} \leq I\left(U_0; Y_{\pi(1)}\right), \tag{14a}$$

$$\theta_{\pi(2)} \leq \min\left\{I\left(U_0; Y_{\pi(2)}\right), \eta_{\pi(2)}\left(R_0^{\pi(2)}\right)\right\}, \tag{14b}$$

for some conditional pmf $P_{U_0|Y_0}$ and rate $R_0^{\pi(2)}$ satisfying

$$R_0 \geq \left(1 - \epsilon_{\pi(1)}\right) I(U_0; Y_0) + \left(\epsilon_{\pi(1)} - \epsilon_{\pi(2)}\right) R_0^{\pi(2)}. \tag{14c}$$

Proof: See Appendix A. ∎

The following example illustrates the benefits of expected-rate constraints versus maximum-rate constraints, and the trade-off between the two exponents when $R_1 = R_2 = 0$.

Example 1: Consider the following joint pmf $P_{Y_0 Y_1 Y_2}$:

	(Y_1, Y_2)			
Y_0	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
0	0.05	0.05	0.15	0.083325
1	0.05	0.15	0.05	0.08335
2	0.15	0.05	0.05	0.083325

For this pmf, Figure 2 shows the optimal exponents regions under maximum- and expected-rate constraints when $R_0 = 0.1$ and $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$. The figure illustrates the boost in the exponents region due to the *expected*-rate constraints. It also emphasizes the benefits of sharing the rate in (14c) between two summands, which relate to the fact that depending on the observation Y_0^n we use two variants of the coding scheme in [8], one with auxiliary U_0 and the other with an auxiliary $U_0^{\pi(2)}$ that satisfies $I(U_0^{\pi(2)}; Y_0) \leq R_0^{\pi(2)}$ and $I(U_0^{\pi(2)}; Y_1) = \eta_{\pi(2)}(R_0^{\pi(2)})$. Restricting to a single auxiliary U_0 in (14) (i.e., setting $R_0^{\pi(2)} = I(U_0; Y_0)$) results in an exponents region, denoted $\mathcal{E}_{\text{no-RS}}\left(R_0, 0, 0, \epsilon_1, \epsilon_2\right)$ which coincides with $\mathcal{E}^*(R_0, 0, 0, \epsilon_2, \epsilon_2)$ and $\mathcal{E}^*_{\text{fix}}\left((1-\epsilon_2)^{-1} R_0, 0, 0, \epsilon_1, \epsilon_2\right)$.

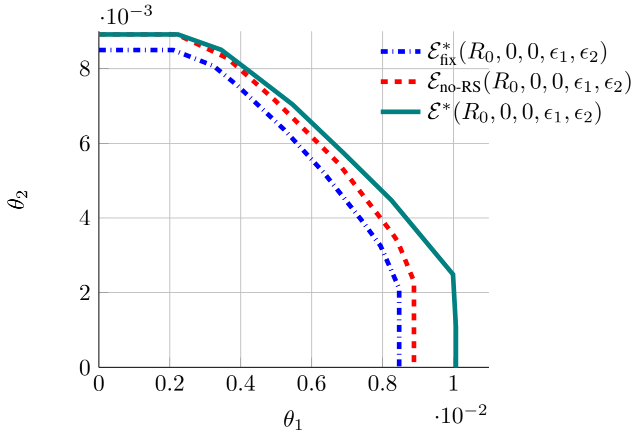

Fig. 2: Optimal error exponents regions under expected- and maximum-rate constraints for $R_0 = 0.1, \epsilon_1 = 0.15, \epsilon_2 = 0.05$.

IV. Converse Proof to Theorem 1

Fix an exponent pair in $\mathcal{E}^*(R_0, R_1, R_2, \epsilon_1, \epsilon_2)$ and a sequence (in n) of encoding and decision functions $\{(\phi^{(n)}, g_1^{(n)}, g_2^{(n)})\}$ satisfying the constraints on the rate and the error probabilities in (6). Our proof relies on the following lemma:

Lemma 1: Fix a blocklength n and a set $\mathcal{D} \subseteq \mathcal{Y}_0^n$ of positive probability, and let the tuple $(\tilde{\mathsf{M}}_0, \tilde{\mathsf{M}}_1, \tilde{\mathsf{M}}_2, \tilde{Y}_0^n, \tilde{Y}_1^n, \tilde{Y}_2^n)$ follow the pmf

$$\begin{aligned} P_{\tilde{\mathsf{M}}_0 \tilde{\mathsf{M}}_1 \tilde{\mathsf{M}}_2 \tilde{Y}_0^n \tilde{Y}_1^n \tilde{Y}_2^n}(\mathsf{m}_0, \mathsf{m}_1, \mathsf{m}_2, y_0^n, y_1^n, y_2^n) &\triangleq \\ P_{Y_0^n Y_1^n Y_2^n}(y_0^n, y_1^n, y_2^n) \cdot \frac{\mathbb{1}\{y_0^n \in \mathcal{D}\}}{P_{Y_0^n}(\mathcal{D})} & \\ \cdot \mathbb{1}\{\phi^{(n)}(y_0^n) = (\mathsf{m}_0, \mathsf{m}_1, \mathsf{m}_2)\}. & \end{aligned} \tag{15}$$

Further, define $U_0 \triangleq (\tilde{\mathsf{M}}_0, \tilde{Y}_0^{T-1}, T)$, $U_1 \triangleq \tilde{\mathsf{M}}_1$, $U_2 \triangleq \tilde{\mathsf{M}}_2$, $\tilde{Y}_i \triangleq \tilde{Y}_{i,T}$ (for $i \in \{0,1,2\}$), where T is uniform over $\{1, \ldots, n\}$ and independent of all other random variables. Notice the Markov chain $(U_0, U_1, U_2) \to \tilde{Y}_0 \to (\tilde{Y}_1, \tilde{Y}_2)$. Then the following inequalities hold:

$$H(\tilde{\mathsf{M}}_0) \geq nI(U_0; \tilde{Y}_0) + \log P_{Y_0^n}(\mathcal{D}), \tag{16}$$

$$H(\tilde{\mathsf{M}}_i) \geq nI(U_i; \tilde{Y}_0|U_0), \quad i \in \{1,2\}. \tag{17}$$

Let $\eta > 0$ be arbitrary. For $i \in \{1,2\}$, if

$$\Pr[\hat{\mathcal{H}}_{Y_i} = 0 | \mathcal{H} = 0, Y_0^n = y_0^n] \geq \eta, \ \ \forall y_0^n \in \mathcal{D}, \tag{18}$$

then

$$-\frac{1}{n} \log \beta_{i,n} \leq I(U_0 U_i; \tilde{Y}_i) + \o_i(n), \tag{19}$$

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

Proof: Similar to the proof of [20, Lemma 1]. For details, see Appendix B in [25]. ∎

We now proceed to prove the converse to Theorem 1. Fix a positive $\eta > 0$. Denote for each blocklength n, the set of strongly typical sequences in $\mathcal{Y}_0^n$ by $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0})$. Set $\mu_n = n^{-1/3}$ and define for $i \in \{1,2\}$, the sets

$$\mathcal{B}_i(\eta) \triangleq \{y_0^n \in \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0}):$$

$$\Pr[\hat{\mathcal{H}}_{Y_i}=0|Y_0^n=y_0^n,\mathcal{H}=0]\geq\eta\},\ i\in\{1,2\}, \tag{20}$$

$$\mathcal{D}_0(\eta)\triangleq\mathcal{B}_1(\eta)\cap\mathcal{B}_2(\eta), \tag{21}$$

$$\mathcal{D}_i(\eta)\triangleq\mathcal{B}_i(\eta)\backslash\mathcal{D}_0(\eta). \tag{22}$$

Further define for each n the probabilities

$$\Delta_j\triangleq P_{Y_0^n}(\mathcal{D}_j(\eta)),\quad j\in\{0,1,2\}, \tag{23}$$

and notice that by the laws of probability

$$\Delta_0+\Delta_i=P_{Y_0^n}(\mathcal{B}_i(\eta)),\quad i\in\{1,2\}, \tag{24}$$

$$\Delta_0\geq P_{Y_0^n}(\mathcal{B}_1(\eta))+P_{Y_0^n}(\mathcal{B}_2(\eta))-1. \tag{25}$$

By (6b), it can be shown that

$$1-\epsilon_i\leq\eta(1-P_{Y_0^n}(\mathcal{B}_i(\eta)))+P_{Y_0^n}(\mathcal{B}_i(\eta))+P_{Y_0}^n(\overline{\mathcal{T}}_{\mu_n}^{(n)}). \tag{26}$$

Thus, by (26) and [24, Lemma 2.12]:

$$P_{Y_0^n}(\mathcal{B}_i(\eta))\geq\frac{1-\epsilon_i-\eta}{1-\eta}-\frac{|\mathcal{Y}_0|}{(1-\eta)4\mu_n^2n},\quad i\in\{1,2\}, \tag{27}$$

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

$$\lim_{\eta\downarrow 0}\lim_{n\to\infty}(\Delta_0+\Delta_i)\geq 1-\epsilon_i,\quad i\in\{1,2\} \tag{28a}$$

$$\lim_{\eta\downarrow 0}\lim_{n\to\infty}\Delta_0\geq 1-\epsilon_1-\epsilon_2 \tag{28b}$$

$$\lim_{\eta\downarrow 0}\lim_{n\to\infty}\sum_{j=0}^{2}\Delta_j\leq 1. \tag{28c}$$

We proceed by applying Lemma 1 to the set $\mathcal{D}_j$ for any $j\in\{0,1,2\}$ with $\Delta_j>0$, and conclude that for any $j\in\{0,1,2\}$ with $\Delta_j>0$ there is a tuple (U_0^j,U_1^j,U_2^j) satisfying

$$H(\tilde{\mathsf{M}}_0^j)\geq nI(U_0^j;\tilde{Y}_0^j)+\log P_{Y_0^n}(\mathcal{D}_j),\quad j\in\{0,1,2\}, \tag{29}$$

$$H(\tilde{\mathsf{M}}_i^j)\geq nI(U_i^j;\tilde{Y}_0^j|U_0^j),\qquad i\in\{1,2\},\ j\in\{0,i\}, \tag{30}$$

and for $i\in\{1,2\},\ j\in\{0,i\}$:

$$-\frac{1}{n}\log\beta_{i,n}\leq I(U_0^jU_i^j;\tilde{Y}_i^j)+\o_i^j(n), \tag{31}$$

where for each pair (i,j), the function $\o_i^j(n)\to 0$ as $n\to\infty$ and the random variables $\tilde{Y}_0^j,\tilde{Y}_i^j,\tilde{\mathsf{M}}_0^j,\tilde{\mathsf{M}}_i^j$ are defined as in the lemma applied to the subset $\mathcal{D}_j$.

To summarize:

$$-\frac{1}{n}\log\beta_{i,n}\leq\min\{I(U_0^0U_i^0;\tilde{Y}_i^0);I(U_0^iU_i^i;\tilde{Y}_i^i)\}+\o_i(n), \tag{32}$$

where $\o_i(n)$ is a function tending to 0 as $n\to\infty$.

Define the following random variables for $i\in\{1,2\}$ and $j\in\{0,1,2\}$

$$\tilde{L}_{i,j}\triangleq\mathrm{len}(\tilde{\mathsf{M}}_i^j). \tag{33}$$

By the rate constraints (2), and the definition of the random variables $\tilde{\mathsf{M}}_i^j$, we obtain by the total law of expectations

$$nR_0\geq\mathbb{E}[L_0]\geq\sum_{j\in\{0,1,2\}}\mathbb{E}[\tilde{L}_{0,j}]\Delta_j. \tag{34}$$

Moreover,

$$H(\tilde{\mathsf{M}}_0^j)=H(\tilde{\mathsf{M}}_0^j,\tilde{L}_{0,j}) \tag{35}$$

$$=\sum_{l_j}\Pr[\tilde{L}_{0,j}=l_j]H(\tilde{\mathsf{M}}_0^j|\tilde{L}_{0,j}=l_j)+H(\tilde{L}_{0,j}) \tag{36}$$

$$\leq\sum_{l_j}\Pr[\tilde{L}_{0,j}=l_j]l_j+H(\tilde{L}_{0,j}) \tag{37}$$

$$=\mathbb{E}[\tilde{L}_{0,j}]+H(\tilde{L}_{0,j}), \tag{38}$$

which combined with (34) establishes

$$\sum_{j\in\{0,1,2\}}\Delta_jH(\tilde{\mathsf{M}}_0^j)\leq\sum_{j\in\{0,1,2\}}\Delta_j\mathbb{E}[\tilde{L}_{0,j}]+\Delta_jH(\tilde{L}_{0,j}) \tag{39}$$

$$\leq nR_0\left(1+\sum_{j\in\{0,1,2\}}h_b\left(\frac{\Delta_j}{nR_0}\right)\right), \tag{40}$$

where (40) holds by (34) and because the entropy of a discrete and positive random variable $\tilde{L}_{0,j}$ of mean $\mathbb{E}[\tilde{L}_{0,j}]\leq\frac{nR_0}{\Delta_j}$ is bounded by $\frac{nR_0}{\Delta_j}\cdot h_b\left(\frac{\Delta_j}{nR_j}\right)$, see [26, Theorem 12.1.1].

In a similar way we obtain for $i\in\{1,2\}$

$$\sum_{j\in\{0,i\}}\Delta_jH(\tilde{\mathsf{M}}_i^j)\leq nR_i\left(1+\sum_{j\in\{0,i\}}h_b\left(\frac{\Delta_j}{nR_i}\right)\right). \tag{41}$$

Notice that when $\Delta_j=0$, the trivial choice $U_i^j=\tilde{Y}_i^j$ satisfies the inequalities (32), (40), and (41). Therefore, above conclusions hold for (U_0^j,U_1^j,U_2^j) for any $j\in\{0,1,2\}$.

Combining (40) and (41) with (29) and (30), noting (24) and (27), and considering also (32), we have proved so far that for all $n\geq 1$ there exist joint pmfs $P_{U_0^jU_1^jU_2^j\tilde{Y}_0^j\tilde{Y}_1^j\tilde{Y}_2^j}=P_{\tilde{Y}_0^j}P_{\tilde{Y}_1^j\tilde{Y}_2^j|\tilde{Y}_0^j}P_{U_0^jU_1^jU_2^j|\tilde{Y}_0^j}$ (abbreviated as $P_j^{(n)}$) for $j\in\{0,1,2\}$ so that the following conditions hold for $i\in\{1,2\}$ (where I_P indicates that the mutual information should be calculated according to a pmf P):

$$R_0\geq\sum_{j\in\{0,1,2\}}\left(I_{P_j^{(n)}}(U_0^j;\tilde{Y}_0^j)+g_{1,j}(n)\right)\cdot g_{2,j}(n,\eta), \tag{42a}$$

$$R_i\geq\sum_{j\in\{0,i\}}\left(I_{P_j^{(n)}}(U_i^j;\tilde{Y}_0^j|U_0^j)\right)\cdot g_{2,j}(n,\eta), \tag{42b}$$

$$\theta_i\leq\min\{I_{P_0^{(n)}}(U_0^0U_i^0;\tilde{Y}_i^0),I_{P_i^{(n)}}(U_0^iU_i^i;\tilde{Y}_i^i)\}+g_{3,i}(n), \tag{42c}$$

for some nonnegative functions $g_{1,j}(n),g_{2,j}(n,\eta),g_{3,i}(n)$ with the following asymptotic behaviors:

$$\lim_{n\to\infty}g_{1,j}(n)=0,\qquad\forall j\in\{0,1,2\}, \tag{43}$$

$$\lim_{n\to\infty}g_{3,i}(n)=0,\qquad\forall i\in\{1,2\}, \tag{44}$$

$$\lim_{n\to\infty}(g_{2,0}(n,\eta)+g_{2,i}(n,\eta))\geq\frac{1-\epsilon_i-\eta}{1-\eta},\qquad\forall i\in\{1,2\}. \tag{45}$$

By Carathéodory's theorem [23, Appendix C], there exist for each n, random variables $U_0^0,U_0^1,U_0^2,U_1^0,U_1^1,U_2^0,U_2^2$ satisfying (42) over alphabets of sizes

$$|\mathcal{U}_0^0|\leq|\mathcal{Y}_0|+3, \tag{46}$$

$$|\mathcal{U}_0^j|\leq|\mathcal{Y}_0|+2,\qquad j\in\{1,2\}, \tag{47}$$

$$|\mathcal{U}_i^j|\leq|\mathcal{U}_0^j|\cdot|\mathcal{Y}_0|+1,\quad i\in\{1,2\},j\in\{0,i\}. \tag{48}$$

Then we invoke the Bolzano-Weierstrass theorem and consider for each $j \in \{0,1,2\}$ a sub-sequence $P^{(n_k)}_{U_0^j U_1^j U_2^j \tilde{Y}_0^j \tilde{Y}_1^j \tilde{Y}_2^j}$ that converges to a limiting pmf $P^*_{U_0^j U_1^j U_2^j Y_0^j Y_1^j Y_2^j}$. For these limiting pmfs, which we abbreviate by P_j^*, we conclude by (42a)–(42c) and (28) that for all $i \in \{1,2\}$:

$$R_0 \geq \sigma_0 \cdot I_{P_0^*}(U_0^0;Y_0^0) + \sigma_1 \cdot I_{P_1^*}(U_0^1;Y_0^1) \\ +\sigma_2 \cdot I_{P_2^*}(U_0^2;Y_0^2), \tag{49}$$

$$R_i \geq \sigma_0 \cdot I_{P_0^*}(U_i^0;Y_0^0|U_0^0) + \sigma_i \cdot I_{P_i^*}(U_i^i;Y_0^i|U_0^i), \tag{50}$$

$$\theta_i \leq \min\{I_{P_0^*}(U_0^0U_i^0;Y_i^0), I_{P_i^*}(U_0^iU_i^i;Y_i^i)\}, \tag{51}$$

where numbers $\sigma_0,\sigma_1,\sigma_2 > 0$ satisfy $\sigma_0+\sigma_1+\sigma_2 \leq 1$ and

$$\sigma_0 + \sigma_i \geq 1-\epsilon_i, \qquad i \in \{1,2\}, \tag{52a}$$

$$\sigma_0 \geq 1-\epsilon_1-\epsilon_2. \tag{52b}$$

Notice further that since for any $j \in \{0,1,2\}$ and any k, the sequence $\tilde{Y}_0^{j,n_k}$ lies in the typical set $\mathcal{T}^{(n_k)}_{\mu_{n_k}}(P_{Y_0})$, we have for all $j \in \{0,1,2\}$, $|P_{\tilde{Y}_0^j} - P_{Y_0}| \leq \mu_{n_k}$ and thus the limiting pmf satisfies $P^*_{Y_0^j} = P_{Y_0}$. Moreover, since for each n_k the pair of random variables $(\tilde{Y}_1^j, \tilde{Y}_2^j)$ is drawn according to $P_{Y_1Y_2|Y_0}$ given $\tilde{Y}_0^j$, the limiting pmf also satisfies $P^*_{Y_1^jY_2^j|Y_0^j} = P_{Y_1Y_2|Y_0}$. We also notice for all $j \in \{0,1,2\}$ that under P_j^* the Markov chain $(U_0^j,U_1^j,U_2^j) \to Y_0 \to (Y_1,Y_2)$ holds. This concludes the converse proof.

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.

Appendix
Proof of Corollary 1

By Theorem 1, $\mathcal{E}^*(R_0,0,0,\epsilon_1,\epsilon_2)$ is the set of all (θ_1,θ_2) pairs satisfying

$$\theta_i \leq \min\left\{I(U_0;Y_i), \eta_i\big(R_0^i\big)\right\}, \qquad i \in \{1,2\}. \tag{53a}$$

for some non-negative numbers $\sigma_0,\sigma_1,\sigma_2$ with sum ≤ 1 and satisfying (7d) and (7e), a conditional pmf $P_{U_0|Y_0}$, and nonnegative rates R_0^1, R_0^2 such that

$$R_0 \geq \sigma_0 I(U_0;Y_0) + \sigma_1 R_0^1 + \sigma_2 R_0^2. \tag{53b}$$

Notice that without loss in optimality, in the evaluation of above region, we can restrict to tuples $\left(P_{U_0|Y_0}, R_0^1, R_0^2\right)$ satisfying

$$I(U_0;Y_i) \geq \eta_i\left(R_0^i\right), \tag{54}$$

which by the maximum in the definition of function η_i implies

$$I(U_0;Y_0) \geq R_0^i, \quad i \in \{1,2\}. \tag{55}$$

In fact, if (54) is violated, rates R_0^1 and/or R_0^2 can be reduced without changing (53a) and so that (54) holds.

We next show that any exponent pair (θ_1,θ_2) and tuple $(P_{U_0|Y_0}, R_0^1, R_0^2)$ satisfying (53), (54), and

$$I(U_0;Y_0) \leq R_0^{\pi(1)} + R_0^{\pi(2)} \tag{56}$$

also satisfies (14). The exponents' constraints (14a) and (14b) are easily verified. To verify (14c), notice that when $\sigma_0 > 1-\epsilon_{\pi(1)}$:

$$R_0 \geq \sigma_0 I(U_0;Y_0) + \sigma_{\pi(1)} R_0^{\pi(1)} + \sigma_{\pi(2)} R_0^{\pi(2)} \tag{57}$$

$$= (1-\epsilon_{\pi(1)}) I(U_0;Y_0) + \sigma_{\pi(1)} R_0^{\pi(1)} \\ +(\sigma_0 - 1 + \epsilon_{\pi(1)}) I(U_0;Y_0) + \sigma_{\pi(2)} R_0^{\pi(2)} \tag{58}$$

$$\geq (1-\epsilon_{\pi(1)}) I(U_0;Y_0) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)}) R_0^{\pi(2)} \tag{59}$$

where (59) holds because $\sigma_{\pi(1)} R_0^{\pi(1)} \geq 0$, because $I(U_0;Y_0) \geq R_0^{\pi(2)}$ by (55), and $\sigma_0 + \sigma_{\pi(2)} \geq 1-\epsilon_{\pi(2)}$ by (7d).

For $\sigma_0 \leq 1-\epsilon_{\pi(1)}$, (14c) can be verified as follows:

$$R_0 \geq \sigma_0 I(U_0;Y_0) + \sigma_{\pi(1)} R_0^{\pi(1)} + \sigma_{\pi(2)} R_0^{\pi(2)} \tag{60}$$

$$\geq \sigma_0 I(U_0;Y_0) + (1-\epsilon_{\pi(1)} - \sigma_0) R_0^{\pi(1)} + \sigma_{\pi(2)} R_0^{\pi(2)} \tag{61}$$

$$\geq \sigma_0 I(U_0;Y_0) + (1-\epsilon_{\pi(1)} - \sigma_0)\big(R_0^{\pi(1)} + R_0^{\pi(2)}\big) \\ +(\epsilon_{\pi(1)} - \epsilon_{\pi(2)}) R_0^{\pi(2)} \tag{62}$$

$$\geq (1-\epsilon_{\pi(1)}) I(U_0;Y_0) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)}) R_0^{\pi(2)} \tag{63}$$

where (61) holds by (7d), (62) holds because $\sigma_{\pi(2)} \geq 1-\epsilon_{\pi(2)} - \sigma_0$ by (7d), and (63) holds by (56) and $\sigma_0 \leq 1-\epsilon_{\pi(1)}$. This establishes that (53) holds under condition (56).

The proof is concluded by showing that for any tuple $(\theta_1,\theta_2,P_{U_0|Y_0},R_0^1,R_0^2)$ satisfying (53), (54), and

$$I(U_0;Y_0) > R_0^{\pi(1)} + R_0^{\pi(2)}, \tag{64}$$

we can find a pmf $P_{\tilde{U}_0|Y_0}$, satisfying (14) when U_0 is replaced by $\tilde{U}_0$. Choose a bivariate $\tilde{U}_0 = (\tilde{U}_0^1, \tilde{U}_0^2)$ such that $\tilde{U}_0^1 \to Y_0 \to \tilde{U}_0^2$ forms a Markov chain and for each $i \in \{1,2\}$ the new random-variable $\tilde{U}_0^i$ achieves $\eta_i\big(R_0^i\big)$, i.e.,

$$R_0^i \geq I\big(Y_0;\tilde{U}_0^i\big) \quad \text{and} \quad \eta_i\big(R_0^i\big) = I\big(\tilde{U}_0^i;Y_i\big). \tag{65}$$

Since for any $i \in \{1,2\}$ we have $I(\tilde{U}_0;Y_i) \geq I(\tilde{U}_0^i;Y_i) = \eta_i(R_0^i)$, the exponents satisfy

$$\theta_{\pi(1)} \leq \min\left\{I(U_0;Y_{\pi(1)}), \eta_{\pi(1)}\big(R_0^{\pi(1)}\big)\right\} = \eta_{\pi(1)}\big(R_0^{\pi(1)}\big) \tag{66}$$

$$\leq I(\tilde{U}_0;Y_{\pi(1)}), \tag{67}$$

$$\theta_{\pi(2)} \leq \min\left\{I(U_0;Y_{\pi(2)}), \eta_{\pi(2)}\big(R_0^{\pi(2)}\big)\right\} = \eta_{\pi(2)}\big(R_0^{\pi(2)}\big) \tag{68}$$

$$= \min\left\{I(\tilde{U}_0;Y_{\pi(2)}), \eta_{\pi(2)}\big(R_0^{\pi(2)}\big)\right\}, \tag{69}$$

where the inequalities in (66) and (68) hold by (54). Similarly,

$$R_0 \geq \sigma_0 I(U_0;Y_0) + \sigma_{\pi(1)} R_0^{\pi(1)} + \sigma_{\pi(2)} R_0^{\pi(2)} \tag{70}$$

$$> (1-\epsilon_{\pi(1)}) R_0^{\pi(1)} + (1-\epsilon_{\pi(2)}) R_0^{\pi(2)} \tag{71}$$

$$= (1-\epsilon_{\pi(1)}) I\big(\tilde{U}_0^{\pi(1)};Y_0\big) + (1-\epsilon_{\pi(2)}) I\big(\tilde{U}_0^{\pi(2)};Y_0\big) \tag{72}$$

$$\geq (1-\epsilon_{\pi(1)}) I\big(\tilde{U}_0^{\pi(1)};Y_0\big) + (1-\epsilon_{\pi(1)}) I\big(\tilde{U}_0^{\pi(2)};Y_0|\tilde{U}_0^{\pi(1)}\big) \\ +(\epsilon_{\pi(1)} - \epsilon_{\pi(2)}) I\big(\tilde{U}_0^{\pi(2)};Y_0\big) \tag{73}$$

$$= (1-\epsilon_{\pi(1)}) I\big(\tilde{U}_0;Y_0\big) + (\epsilon_{\pi(1)} - \epsilon_{\pi(2)}) I\big(\tilde{U}_0^{\pi(2)};Y_0\big) \tag{74}$$

where inequality (71) holds by the assumption that $I(U_0;Y_0) > R_0^1 + R_0^2$ and by condition (7d); equality (72) holds by (65); inequality (73) holds by the Markov chain $\tilde{U}_0^1 \to Y_0 \to \tilde{U}_0^2$; and (74) by the chain rule and the definition of $\tilde{U}_0$.

REFERENCES

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

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

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

[4] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, pp. 6282–6303, Oct. 2012.

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

[6] 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.

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

[8] 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.

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

[10] S. Watanabe, "Neyman–pearson test for zero-rate multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4923–4939, 2018.

[11] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via korner-marton coding," in *Proc. IEEE Info. Theory Work. (ITW)*, 2016.

[12] 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.

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

[14] 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.

[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] K. R. Sahasranand and H. Tyagi, "Communication complexity of distributed high dimensional correlation testing," *IEEE Transactions on Information Theory*, vol. 67, no. 9, pp. 6082–6095, 2021.

[17] 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.

[18] M. Hamad, M. Wigger, and M. Sarkiss, "Cooperative multi-sensor detection under variable-length coding," in *2020 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2021.

[19] 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.

[20] 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.

[21] 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.

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

[23] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

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

[25] 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.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed.* Wiley, 2006.