

# Distributed Hypothesis Testing with Concurrent Detections

Pierre Escamilla<sup>†\*</sup>   Michèle Wigger<sup>\*</sup>   Abdellatif Zaidi<sup>† ‡</sup>

<sup>†</sup> Paris Research Center, Huawei Technologies, Boulogne-Billancourt, 92100, France

<sup>‡</sup> Université Paris-Est, Champs-sur-Marne, 77454, France

<sup>\*</sup> LTCI, Télécom ParisTech, Université Paris-Saclay, 75013 Paris, France

{pierre.escamilla@huawei.com, abdellatif.zaidi@u-pem.fr}  
 {michele.wigger@telecom-paristech.fr}

**Abstract**—A detection system with a single sensor and  $K$  detectors is considered, where each of the terminals observes a memoryless source sequence and the sensor sends a common message to all the detectors. The communication of this message is assumed error-free but rate-limited. The joint probability mass function (pmf) of the source sequences observed at the terminals depends on an  $M$ -ary hypothesis ( $M \geq K$ ), and the goal of the communication is that each detector can guess the underlying hypothesis. Each detector  $k$  aims to maximize the error exponent under hypothesis  $k$ , while ensuring a small probability of error under all other hypotheses. This paper presents an achievable exponents region for the case of positive communication rate, and characterizes the optimal exponents region for the case of zero communication rate. All results extend also to a composite hypothesis testing scenario.

## I. INTRODUCTION

Consider the multiterminal hypothesis testing scenario in Figure 1 where an encoder observes a discrete memoryless source sequence  $X^n \triangleq (X_1, \dots, X_n)$  and communicates with multiple detectors over a *common* noise-free bit-pipe of rate  $R \geq 0$ . Here,  $n$  is a positive integer that denotes the blocklength. Each detector  $k \in \{1, \dots, K\}$  observes a memoryless source sequence  $Y_k^n \triangleq (Y_{k,1}, \dots, Y_{k,n})$ , where the sequence of observations  $\{(X_t, Y_{1,t}, Y_{2,t}, \dots, Y_{K,t})\}_{t=1}^n$  is independent and identically distributed (i.i.d) according to a joint probability mass function (pmf) that is determined by the hypothesis  $\mathcal{H} \in \{1, \dots, M\}$ . Under hypothesis  $\mathcal{H} = m$ :

$$\{(X_t, Y_{1,t}, Y_{2,t}, \dots, Y_{K,t})\}_{t=1}^n \text{ i.i.d. } \sim P_{XY_1 Y_2 \dots Y_K}^{(m)} \quad (1)$$

We assume  $M \geq K$ . Each detector  $k \in \{1, 2, \dots, K\}$  decides on a hypothesis,  $\hat{\mathcal{H}}_k \in \{1, 2, \dots, M\}$ , with the goal to maximize the exponential decrease of the probability of type-II error (i.e., of guessing  $\hat{\mathcal{H}}_k \neq k$  when  $\mathcal{H} = k$ ), while ensuring that the probabilities of type-I error (i.e., guessing  $\hat{\mathcal{H}}_k \neq m$  when  $\mathcal{H} = m$  for some  $m \neq k$ ) do not exceed a constant value  $\epsilon_k \in (0, 1)$  for all sufficiently large blocklengths  $n$ .

The described scenario models, e.g., a biometric system where  $K$  distinct accounts require authentication from users using biometric features. Each account  $k$  aims to grant access to a *different user* while denying access to all other users and guaranteeing a given probability  $\epsilon_k$  of false accept.

Problems of distributed hypothesis testing are strongly rooted in both statistics and information theory. In particular, the problem described above with only a single detector and

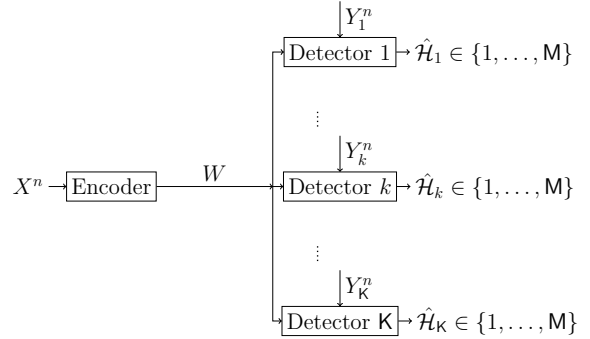


Fig. 1. Distributed hypothesis testing with multiple detectors.

two hypotheses ( $K = 1$  and  $M = 2$ ) was studied in [1]–[4]. The optimal exponent was derived for testing against conditional independence [4], but it remains open in general. Extensions to multiple terminals, multi-hop or interactive communication were presented in [3], [5], [6], [8], [9]. Moreover, [11] studies testing against conditional independence with two detectors that aim at maximizing the error exponent under the same hypothesis. The results in [11] report a tradeoff between the exponents achieved at the two detectors when the sensor sends a common message of positive rate  $R > 0$  to both sensors. As this paper shows, such a tradeoff does not necessarily arise when the two detectors aim at maximizing the error exponents under two different hypotheses. (See Remark 1.)

In this paper, we propose a coding and testing scheme for a general number of hypotheses  $M$  and detectors  $K$  and positive rate  $R > 0$ . In our scheme, the sensor makes a tentative guess on the hypothesis and then applies the coding and testing scheme that is adapted to the guessed hypothesis. Based on this scheme, we describe a set of achievable error exponents for given rate  $R > 0$ .

Furthermore, we exactly characterize the set of achievable error exponents for communication rate  $R = 0$ . That means when the sensor can send only a sublinear number of bits to the  $K$  detectors. If the sensor sends at most  $\log(M)$  bits, a tradeoff arises between their error exponents. In contrast, if it sends more than  $\log(M)$  bits, each detector can achieve the same error exponent as if the other detectors were not present.

At last, we extend our results to a *composite hypothesis testing* scenario where each receiver merges a set of hypotheses and aims at maximizing the error exponent under this set of

hypotheses. This scenario models situations where a detector wishes to raise an alarm in multiple events, without the need to specify which of the events occurred. It can also model uncertainty about the distribution implied by the hypotheses. Error exponents for zero-rate single-detector composite hypothesis testing problems have previously been presented in [12].

*Notation:* The set of pmfs over an alphabet  $\mathcal{X}$  is denoted  $\mathcal{P}(\mathcal{X})$  and the  $n$ -fold product of a pmf  $P_X$  by  $P_X^{\otimes n}$ , and the type of a sequence  $x^n$  is denoted by  $\text{tp}(x^n)$ . We write  $\mathcal{T}_\mu(P_X)$  for the set of  $\mu$ -typical sequences with respect to  $P_X$ . Both  $D(P_X \| P_{\bar{X}})$  and  $D(X \| \bar{X})$  denote the Kullback-Leiber divergence between two pmfs  $P_X$  and  $P_{\bar{X}}$  over the same alphabet  $\mathcal{X}$ . For an integer number  $I$ , we abbreviate  $\{1, \dots, I\}$  by  $[1 : I]$ . Finally,  $H(\cdot)$  and  $I(\cdot; \cdot)$  denote entropy and mutual information. Where  $I_P(A; B)$  indicates that  $I(A; B)$  is computed according to  $P$ .

## II. FORMAL PROBLEM STATEMENT

Let  $(X^n, Y_1^n, \dots, Y_K^n)$  be i.i.d. according to one of the pmfs  $P_{XY_1 \dots Y_K}^{(1)}, \dots, P_{XY_1 \dots Y_K}^{(M)}$  as described in (1). For clarity of exposition, assume that for each  $k \in \{1, \dots, K\}$ , the pmfs  $P_{XY_k}^{(1)}, \dots, P_{XY_k}^{(M)}$  are all different and  $P_{XY_k}^{(m)} \ll P_{XY_k}^{(k)}$ , i.e.,  $P_{XY_k}^{(k)}$  is positive for all elements where  $P_{XY_k}^{(m)}(x, y_k)$  is positive. The encoder observes a source sequence  $X^n$ . It applies an encoding function

$$f_n: \mathcal{X}^n \rightarrow \mathcal{W} \triangleq [1 : W_n] \quad (2)$$

to  $X^n$ :

$$W = f_n(X^n), \quad (3)$$

and sends the resulting message  $W$  to all detectors. Detector  $k$  obtains  $W$  and observes  $Y_k^n$ . It applies a decision function:

$$g_{k,n}: \mathcal{W} \times \mathcal{Y}_k^n \rightarrow [1 : M]. \quad (4)$$

to decide on the hypothesis

$$\hat{\mathcal{H}}_k \triangleq g_{k,n}(W, Y_k^n). \quad (5)$$

Type-I and type-II error probabilities at detector  $k$  are defined as:

$$\alpha_{k,m,n} \triangleq \Pr\{\hat{\mathcal{H}}_k \neq m | \mathcal{H} = m\}, \quad m \in [1 : M] \setminus \{k\}, \quad (6)$$

$$\beta_{k,n} \triangleq \Pr\{\hat{\mathcal{H}}_k \neq k | \mathcal{H} = k\}. \quad (7)$$

**Definition 1.** Given rate  $R \geq 0$  and vector  $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_K)$  in  $(0, 1)^K$ , an error exponent vector  $\theta = (\theta_1, \dots, \theta_K)$  is said achievable, if for each blocklength  $n$  there exist functions  $f_n, g_{1,n}, g_{2,n}, \dots, g_{K,n}$  as in (2) and (4), so that the following limits hold for all  $k \in [1 : K]$ :

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{k,m,n} \leq \epsilon_k, \quad m \in [1 : M] \setminus \{k\}, \quad (8)$$

$$\underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_{k,n} \geq \theta_k, \quad (9)$$

and

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log W_n \leq R. \quad (10)$$

For rate  $R = 0$  we make a finer distinction. Constraint  $W_n \leq W$ , for some constant value  $W$ , is referred to as  $R =$

$0_W$ . The main interest of this document is on the set of all achievable error exponent vectors.

**Definition 2.** Given  $R > 0$  or  $R = 0_W$  and vector  $\epsilon \in (0, 1)^K$ , the closure of the set of all achievable exponent vector  $\theta = (\theta_1, \theta_2, \dots, \theta_K)$  is called the error exponents region  $\mathcal{E}(R, \epsilon)$ .

**Remark 1.** The error exponents region only depends on the sets of joint marginals  $\{P_{XY_1}^{(m)}\}_{m=1}^M, \{P_{XY_2}^{(m)}\}_{m=1}^M, \dots, \{P_{XY_K}^{(m)}\}_{m=1}^M$ . In our setup, each detector  $k$  aims at maximizing the error exponent under hypothesis  $\mathcal{H} = k$ . When  $P_X^{(1)} = P_X^{(2)} = \dots = P_X^{(K)}$ , by relabelling above joint marginals, the setup is equivalent to a setup where all detectors aim at maximizing the error exponent under hypothesis  $\mathcal{H} = 1$ . In the special case of testing against conditional independence for  $K = 2$  detectors and  $M = 2$  hypotheses, this latter setup has been solved in [11]. This optimal exponent is also achieved by our first main result, Theorem 1.

## III. MAIN RESULTS

The following Theorem 1 characterizes an inner bound to  $\mathcal{E}(R, \epsilon)$ . As we will see, it does not depend on  $\epsilon$ .

Let for each positive rate  $R > 0$ ,  $\mathcal{U}(R)$  be the set of tuples  $(P_{U|X}^{(1)}, \dots, P_{U|X}^{(M)})$  satisfying

$$I_{P^{(m)}}(U; X | Y_k) \leq R, \quad k \in [1 : K], \quad m \in [1 : M] \setminus \{k\}, \quad (11)$$

and

$$P_{U|X}^{(m)} = P_{U|X}^{(m')} \quad \text{if} \quad P_X^{(m)} = P_X^{(m')}. \quad (12)$$

**Theorem 1 (Achievability under Positive Rate).** Given  $R > 0$  and  $\epsilon$  in  $(0, 1)^K$ , region  $\mathcal{E}(R, \epsilon)$  contains all nonnegative vectors  $\theta = (\theta_1, \theta_2, \dots, \theta_K)$  that satisfy the following two conditions for a tuple  $(P_{U|X}^{(1)}, \dots, P_{U|X}^{(M)}) \in \mathcal{U}(R)$ :

$$\theta_k \leq \min_{\substack{m \in [1: M] \\ m \neq k}} \min_{\substack{\pi_{UXY_k}: \\ \pi_{UX} = P_{UX}^{(m)} \\ \pi_{Y_k} = P_{UY_k}^{(m)}}} D(\pi_{UXY_k} \| P_{U|X}^{(m)} P_{XY_k}^{(k)}) \quad (13)$$

and

$$\theta_k \leq \min_{\substack{m \in [1: M] \\ m \neq k}} \min_{\substack{\pi_{UXY_k}: \\ \pi_{UX} = P_{UX}^{(m)} \\ \pi_{Y_k} = P_{Y_k}^{(m)} \\ H_{P^{(m)}}(U | Y_k) \leq H_{\pi}(U | Y_k)}} D(\pi_{UXY_k} \| P_{U|X}^{(m)} P_{XY_k}^{(k)}) + R - I_{P^{(m)}}(U; X | Y_k). \quad (14)$$

*Proof:* See Section V. ■

**Corollary 1.** Assume that  $K = M = 2$  and that  $P_X^{(1)} \neq P_X^{(2)}$ . In this case, for each  $k \in \{1, 2\}$ , detector  $k$ 's exponent  $\theta_k$  coincides with the Shimokawa-Han-Amari exponent [2] achieved in a single-detector system with only detector  $k$ .

The scheme leading to Theorem 1 applies binning as used in the Shimokawa-Han-Amari scheme [2]. A better performance could be achieved in general if the scheme was replaced by the Heegard-Berger type-scheme in [13].

We now consider the case of zero rate, i.e., where the encoder's message takes value in  $[1 : W]$ , for some fixed finite  $W > 0$ . Assume for the rest of this section that for

each  $k \in [1 : K]$ , each pair of pmfs  $P_{XY_k}^{(m)}$  and  $P_{XY_k}^{(m')}$  differs in at least one of the two marginals  $P_X$  or  $P_{Y_k}$ . Further, assume

$$P_{XY_k}^{(k)} > 0, \quad \forall k \in [1 : K], \quad (x, y_k) \in \mathcal{X} \times \mathcal{Y}_k. \quad (15)$$

Let  $L$  be the number of different pmfs in  $\{P_X^{(1)}, \dots, P_X^{(M)}\}$ . So,  $1 \leq L \leq M$ , where  $L = 1$  means that pmfs  $P_X^{(1)}, \dots, P_X^{(M)}$  all coincide. Let  $P_{X,1}, \dots, P_{X,L}$  be these  $L$  distinct pmfs. The error exponent region depends on  $L$  and on  $W$ . If  $W > L$ , the exponent region is a  $K$ -dimensional cube and in each dimension the exponent can be as large as in a system where only the corresponding detector is present.

**Proposition 1** (Exponents Region when  $W > L$ ). *For any  $\epsilon \in (0, 1)^K$  and  $W > L$ , the exponents region  $\mathcal{E}(0_W, \epsilon)$  coincides with the set of all nonnegative exponent vectors  $\theta$  satisfying:*

$$\theta_k \leq \min_{\substack{m \in [1:M] \\ m \neq k}} \min_{\substack{\pi_{XY_k} : \\ \pi_X = P_X^{(m)} \\ \pi_{Y_k} = P_{Y_k}^{(m)}}} D(\pi_{XY_k} \| P_{XY_k}^{(k)}) \quad (16)$$

*Proof:* The converse holds by [12, Theorem 5] and because any detector cannot have a larger type-II error exponent as in a setup where it is the only detector and the transmitter can send a message of size  $W$  to this single detector. Achievability follows by analyzing the following scheme. (Details of the analysis are omitted.) Fix a small  $\mu > 0$ .

*Encoder:* If the encoder observes  $X^n = x^n$  and  $x^n \in \mathcal{T}_\mu(P_{X,\ell})$  for some index  $\ell \in \{1, \dots, L\}$ , then it sends  $W = \ell$ . Otherwise, it sends  $W = L + 1$ . (Notice that since pmfs  $P_{X,1}, \dots, P_{X,L}$  are all distinct, for sufficiently small  $\mu$ , every  $x^n \in \mathcal{X}^n$  belongs to only one typical set  $\mathcal{T}_\mu(P_{X,\ell})$ .)

*Detector  $k$ :* If the received message  $W \leq L$  and for some  $m \in [1 : M]$  the following two conditions hold:

$$P_X^{(m)} = P_{X,W} \quad (17)$$

$$Y_k^n \in \mathcal{T}_\mu(P_{Y_k}^{(m)}), \quad (18)$$

detector  $k$  sets  $\hat{\mathcal{H}}_k = m$ . Otherwise, it sets  $\hat{\mathcal{H}}_k = k$ . ■

For each vector  $\mathbf{r} := (r_1, \dots, r_{K-1}) \in \mathbb{R}^{K-1}$  and mapping  $b: [1 : M] \rightarrow [1 : W]$ , define a partition  $\phi_{b,1}(\mathbf{r}), \dots, \phi_{b,W}(\mathbf{r})$  of  $\mathcal{P}(\mathcal{X})$  so that  $\phi_{b,i}(\mathbf{r})$  contains  $P_X^{(m)}$  if and only if  $b(m) = i$ , and  $\phi_{b,i}(\mathbf{r})$  contains any other type  $\tilde{P}_X$  only if

$$i = \operatorname{argmax}_{j \in [1:W]} \min_{\kappa \in [1:K]} \left[ \min_{\substack{m \in [1:M] \\ b(m)=j \\ m \neq \kappa}} \min_{\substack{\pi_{XY_\kappa} : \\ \pi_X = \tilde{P}_X \\ \pi_{Y_\kappa} = P_{Y_\kappa}^{(m)}}} D(\pi_{XY_\kappa} \| P_{XY_\kappa}^{(\kappa)}) + \sum_{l=\kappa}^K r_l \right]. \quad (19)$$

**Theorem 2** (Exponents Region when  $W \leq L$ ). *For any  $\epsilon \in (0, 1)^K$ , the exponents region  $\mathcal{E}(0_W, \epsilon)$  coincides with the set of nonnegative exponent vectors  $\theta = (\theta_1, \dots, \theta_K)$  that for some  $\mathbf{r} \in \mathbb{R}^{K-1}$  and mapping  $b: [1 : M] \rightarrow [1 : W]$  satisfy*

$$\theta_k \leq \min_{\substack{m \in [1:M] \\ m \neq k}} \min_{\substack{\pi_X \in \phi_{b,b(m)}(\mathbf{r}) \\ \pi_{Y_k} = P_{Y_k}^{(m)}}} D(\pi_{XY_k} \| P_{XY_k}^{(k)}), \quad k \in [1 : K]. \quad (20)$$

*Proof:* See Section VI. ■

Given the mapping  $b$ , the theorem describes an optimal choice of the partition  $\phi_{b,1}(\mathbf{r}), \dots, \phi_{b,W}(\mathbf{r})$ . It remains to optimize over the choice  $b$ . A similar phenomenon is encountered in the characterization of the optimal error exponent of the single-detector composite hypothesis testing problem in [12] and in the characterization of the optimal set of type-I and type-II error exponents in [14].

**Example 1.** Consider a setup where  $K = 2$ ,  $M = 3$  and  $W = 2$  and where  $X, Y_1, Y_2$  are binary with pmfs

$$\begin{cases} P_{XY_k}^{(1)}(0, 0) = 0.30 & P_{XY_k}^{(1)}(0, 1) = 0.23 \\ P_{XY_k}^{(1)}(1, 0) = 0.27 & P_{XY_k}^{(1)}(1, 1) = 0.20 \end{cases} \quad (21)$$

$$\begin{cases} P_{XY_k}^{(2)}(0, 0) = 0.14 & P_{XY_k}^{(2)}(0, 1) = 0.29 \\ P_{XY_k}^{(2)}(1, 0) = 0.31 & P_{XY_k}^{(2)}(0, 1) = 0.26 \end{cases} \quad (22)$$

$$\begin{cases} P_{XY_k}^{(3)}(0, 0) = 0.52 & P_{XY_k}^{(3)}(0, 1) = 0.18 \\ P_{XY_k}^{(3)}(1, 0) = 0.23 & P_{XY_k}^{(3)}(1, 1) = 0.07. \end{cases} \quad (23)$$

The exponent region corresponding to this example is depicted in Fig. 2. It is non-convex. (Notice that time-sharing arguments cannot be applied to convexify the region.)

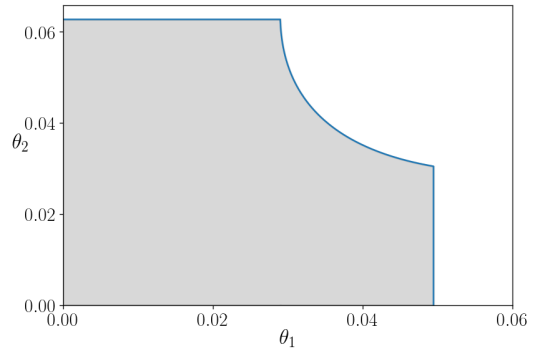


Fig. 2. Exponent region of Example 1, see [15] for implementation details.

#### IV. COMPOSITE HYPOTHESIS TESTING

Consider now the related *composite testing* setup in which each detector aims at maximizing the error exponent that is associated to a set of hypothesis. For each  $k \in [1 : K]$ , let  $\mathcal{S}_k$  be a subset of  $[1 : M]$ . Detector  $k$  declares  $\hat{\mathcal{H}}_k = 0$  to indicate that  $\mathcal{H} \in \mathcal{S}_k$ , and it declares  $\hat{\mathcal{H}}_k = m$  to indicate that  $\mathcal{H} = m$  for any hypothesis  $m \in [1 : M] \setminus \mathcal{S}_k$ . Specifically, it produces  $\hat{\mathcal{H}}_k$  as in (5) but using a decision function of the form

$$g_{k,n}: [1 : W] \times \mathcal{Y}^n \rightarrow \{0\} \cup ([1 : M] \setminus \mathcal{S}_k). \quad (24)$$

Type-I error probabilities are defined as in the previous section. The type-II error probability at detector  $k$  is defined as

$$\gamma_{k,n} := \min_{\xi \in \mathcal{S}_k} \Pr\{\hat{\mathcal{H}}_k \neq 0 | \mathcal{H} = \xi\}. \quad (25)$$

The exponents region is defined as in Definitions 1 and 2, except that (8) and (9) needs to be replaced by:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{k,m,n} \leq \epsilon_k, \quad \forall m \in [1 : M] \setminus \mathcal{S}_k, \quad (26)$$

$$\eta_k \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \gamma_{k,n}, \quad (27)$$

and symbols  $\theta$  and  $\boldsymbol{\theta}$  have to be replaced by  $\eta$  and  $\boldsymbol{\eta} := (\eta_1, \dots, \eta_K)$ .

In this section, we allow for any number of hypotheses  $M \geq 2$ , even  $M < K$ . The result from the previous section, can be extended to this composite hypothesis testing scenario. More specifically, the achievability parts of the following Theorem 3 and Proposition 2 rely on the coding and testing schemes described previously, but where each detector simply declares  $\hat{\mathcal{H}}_k = 0$  instead of  $\hat{\mathcal{H}}_k = m$  when  $m \in \mathcal{S}_k$ . The achievability proof of Theorem 4 necessitates further changes.

**Theorem 3** (Achievability under Positive Rate). *Given  $R > 0$  and  $\epsilon$  in  $(0, 1)^K$ , region  $\mathcal{E}(R, \epsilon)$  contains all nonnegative vectors  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K)$  that satisfy the following two conditions for a tuple  $(P_{U|X}^{(1)}, \dots, P_{U|X}^{(M)}) \in \mathcal{U}(R)$ :*

$$\eta_k \leq \min_{\substack{(\xi, m): \\ \xi \in \mathcal{S}_k \\ m \in [1:M] \setminus \mathcal{S}_k}} \min_{\substack{\pi_{UXY_k}: \\ \pi_{UX} = P_{UX}^{(m)} \\ \pi_{UY_k} = P_{UY_k}^{(m)}}} D\left(\pi_{UXY_k} \| P_{U|X}^{(m)} P_{XY_k}^{(\xi)}\right) \quad (28)$$

and

$$\eta_k \leq \min_{\substack{(\xi, m): \\ \xi \in \mathcal{S}_k \\ m \in [1:M] \setminus \mathcal{S}_k}} \min_{\substack{\pi_{UXY_k}: \\ \pi_{UX} = P_{UX}^{(m)} \\ \pi_{Y_k} = P_{Y_k}^{(m)} \\ H_{P^{(m)}}(U|Y_k) \leq H_{\pi}(U|Y_k)}} \left[ D\left(\pi_{UXY_k} \| P_{U|X}^{(m)} P_{XY_k}^{(\xi)}\right) + R - I_{P^{(m)}}(U; X|Y_k) \right]. \quad (29)$$

*Proof:* Omitted. ■

This result differs from its simple-hypothesis testing counterpart in Theorem 1 only in the first minimization of (28) and (29). A similar statement applies to the following result.

Consider zero-rate communication. Assume (15) and that any two  $P_{XY_k}^{(m)}$  and  $P_{XY_k}^{(m')}$  differ in at least one of the marginals.

**Proposition 2** (Exponents Region when  $W > L$ ). *For any  $\epsilon \in (0, 1)^K$  and  $W > L$ , the exponents region  $\mathcal{E}(0_W, \epsilon)$  coincides with the set of all nonnegative exponent vectors  $\boldsymbol{\theta}$  satisfying:*

$$\eta_k \leq \min_{\substack{(\xi, m): \\ \xi \in \mathcal{S}_k \\ m \in [1:M] \setminus \mathcal{S}_k}} \min_{\substack{\pi_{XY_k}: \\ \pi_X = P_X^{(m)} \\ \pi_{Y_k} = P_{Y_k}^{(m)}}} D\left(\pi_{XY_k} \| P_{XY_k}^{(\xi)}\right) \quad (30)$$

*Proof:* Omitted. ■

**Theorem 4** (Exponents Region when  $W \leq L$ ). *For any  $\epsilon \in (0, 1)^K$  the exponents region  $\mathcal{E}(0_W, \epsilon)$  coincides with the set of nonnegative exponent vectors  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$  that for a partition  $\psi_1, \dots, \psi_W$  of  $\mathcal{P}(\mathcal{X})$  and a mapping  $b: [1:M] \rightarrow [1:W]$  satisfy*

$$\eta_k \leq \min_{\substack{(\xi, m): \\ \xi \in \mathcal{S}_k \\ m \in [1:M] \setminus \mathcal{S}_k}} \min_{\substack{\pi_{XY_k}: \\ \pi_X \in \psi_{b(m)} \\ \pi_{Y_k} = P_{Y_k}^{(m)}}} D\left(\pi_{XY_k} \| P_{XY_k}^{(\xi)}\right), \quad k \in [1:K].$$

*Proof:* Omitted. ■

The optimal partition  $\psi_1, \dots, \psi_W$  of  $\mathcal{P}(\mathcal{X})$  in above Theorem 4 is in general different from the partitions  $\phi_{b,1}(\mathbf{r}), \dots, \phi_{b,W}(\mathbf{r})$  introduced for the simple hypothesis testing setup.

This shows suboptimality of the naive strategy of applying the coding and testing scheme for the simple hypothesis testing problem and then declaring  $\hat{\mathcal{H}}_k = 0$  at a detector  $k$  if the outcome of the test lies in  $\mathcal{S}_k$  and declaring the produced hypothesis otherwise.

## V. PROOF OF THEOREM 1

Fix  $\mu > 0$ , a sufficiently large blocklength  $n$  and conditional pmfs  $P_{U|X}^{(1)}, \dots, P_{U|X}^{(M)}$  so that (11) and (12) are satisfied. Define the joint pmfs  $P_{UXY}^{(m)} = P_{XY}^{(m)} P_{U|X}^{(m)}$  and the nonnegative rate  $R'$  such that for each  $m \in [1:M]$ :

$$R + R' = I_{P^{(m)}}(X; U) + \mu, \quad (31)$$

$$R' < I_{P^{(m)}}(Y_k; U), \quad k \in [1:K] \setminus \{m\}. \quad (32)$$

**Code Construction:** For each  $m \in [1:M]$ , construct a random codebook

$$\mathcal{C}_{U,m} = \{u_m^n(\tilde{w}, \ell): \tilde{w} \in \{1, \dots, \lfloor e^{nR} \rfloor\}, \ell \in \{1, \dots, \lfloor e^{nR'} \rfloor\}\},$$

by drawing all entries i.i.d. according to  $P_U^{(m)}$ .

**Sensor:** Given that it observes the sequence  $X^n = x^n$ , the sensor looks for indices  $(m, \tilde{w}, \ell)$  such that

$$(u_m^n(\tilde{w}, \ell), x^n) \in \mathcal{T}_{\mu/2}(P_{UX}^{(m)}). \quad (33)$$

If successful, it picks one of these indices uniformly at random and sends the pair  $(m, \tilde{w})$  over the noise-free bit pipe. Otherwise, it sends  $W = (0, 1)$ . (Notice that sending index  $m$  takes only  $\lceil \log M \rceil$  bits and is thus of zero rate.)

**Detector  $k$ :** Assume that detector  $k$  observes  $Y_k^n = y_k^n$ . If it obtains message  $W = (0, 1)$ , then it declares  $\hat{\mathcal{H}}_k = k$ . Otherwise, it parses message  $W$  into the pair  $(m, \tilde{w})$ , and picks an index  $\ell' \in \{1, \dots, \lfloor e^{nR'} \rfloor\}$  such that for all  $\tilde{\ell} \in \{1, \dots, \lfloor e^{nR'} \rfloor\}$ :

$$H_{\text{tp}(u_m^n(\tilde{w}, \ell'), y_k^n)}(U|Y_k) \leq H_{\text{tp}(u_m^n(\tilde{w}, \tilde{\ell}), y_k^n)}(U|Y_k). \quad (34)$$

It then checks whether

$$(u_m^n(\tilde{w}, \ell'), y_k^n) \in \mathcal{T}_{\mu}(P_{UY}^{(m)}). \quad (35)$$

If this test is successful, detector  $k$  declares  $\hat{\mathcal{H}}_k = m$ . Otherwise,  $\hat{\mathcal{H}}_k = k$ .

**Error analysis:** By the same technical arguments as in the single decision center setup, see [10, Proof of Thm. 4, App. H].

## VI. PROOF OF THEOREM 2

**Achievability:** Fix  $\mu > 0$  and a function  $b: [1:M] \rightarrow [1:W]$ . Fix also a vector  $\mathbf{r} = (r_1, \dots, r_{K-1}) \in \mathbb{R}^{K-1}$  and define the sets  $\phi_{b,1,\mu}(\mathbf{r}), \dots, \phi_{b,W,\mu}(\mathbf{r})$  so that they partition  $\mathcal{P}^n(\mathcal{X})$ , so that the types of the sequences in  $\mathcal{T}_{\mu}(P_X^{(m)})$  belong to subset  $\phi_{b,b(m),\mu}(\mathbf{r})$ , and so that any other type  $P_X$  is in  $\phi_{b,b(m),\mu}(\mathbf{r})$  only if (19) holds.

**Sensor:** Given that it observes the sequence  $X^n = x^n$ , the sensor sends  $W = i$  if  $\text{tp}(X^n) \in \phi_{b,i,\mu}(\mathbf{r})$ .

**Detector  $k$ :** Given that it observes  $Y_k^n = y_k^n$  and receives message  $W = w$  from the sensor, detector  $k$  checks whether for some  $m \in [1:M]$  the following two conditions hold:

$$b(m) = w \quad \text{and} \quad y_k^n \in \mathcal{T}_{\mu}(P_{Y_k}^{(m)}). \quad (36)$$

If such an index  $m$  exists, Detector  $k$  declares  $\hat{\mathcal{H}}_k = m$ . If none or multiple exist, it declares  $\hat{\mathcal{H}}_k = k$ .

*Error analysis:* Notice first that because all pmfs  $P_{XY_k}^{(1)}, \dots, P_{XY_k}^{(M)}$  differ in at least one of the two marginals, for sufficiently small  $\mu$ , when  $X^n \in \mathcal{T}_\mu(P_X^{(m)})$  and  $Y_k^n \in \mathcal{T}_\mu(P_{Y_k}^{(m)})$ , then Detector  $k$  declares  $\hat{\mathcal{H}}_k = m$ . Thus, by the weak law of large numbers, for each  $k \in [1 : K]$  and sufficiently large  $n$ :

$$\alpha_{k,m,n} \leq 1 - \Pr \left[ (X^n, Y_k^n) \in \mathcal{T}_\mu(P_X^{(m)}) \times \mathcal{T}_\mu(P_{Y_k}^{(m)}) \right] \leq \epsilon_k.$$

Define now for each  $k \in [1 : K]$ ,  $m \in [1 : M]$ , and  $\mathbf{r} \in \mathbb{R}^{K-1}$ :

$$\mathcal{A}_{k,m,\mu}(\mathbf{r}) := \left\{ (x^n, y_k^n) : \begin{array}{l} \text{tp}(x^n) \in \phi_{b,b(m),\mu}(\mathbf{r}), \\ \text{and } y_k^n \in \mathcal{T}_\mu(P_{Y_k}^{(m)}) \end{array} \right\}. \quad (37)$$

The type-II error probability at detector  $k$  satisfies:

$$\begin{aligned} \beta_{k,n} &\leq \Pr \left[ (X^n, Y_k^n) \in \bigcup_{\substack{m \in [1:M] \\ m \neq k}} \mathcal{A}_{k,m,\mu} \mid \mathcal{H} = k \right] \\ &\leq \min_{\substack{m \in [1:M] \\ m \neq k}} \min_{\substack{\pi_{XY_k} \\ \pi_X \in \phi_{b,b(m),\mu}(\mathbf{r}) \\ |\pi_{Y_k} - P_{Y_k}^{(m)}| \leq \mu}} e^{-n(D(\pi_{XY_k} \| P_{XY_k}^{(k)}) - \mu)}, \quad (38) \end{aligned}$$

where the last inequality holds for sufficiently large values of  $n$  and by Sanov's theorem. Taking  $\mu \rightarrow 0$  and  $n \rightarrow \infty$  establishes the desired achievability result.

*Converse:* Fix an integer number  $n$  and define  $\epsilon_{\min}$  to be the minimum component of  $\epsilon$ . Since the sensor sends an index in  $[1 : W]$ , for each  $m \in [1 : M]$  there exists a partition  $\mathcal{C}_1, \dots, \mathcal{C}_W$  of  $\mathcal{X}^n$  and subsets  $\mathcal{F}_1^{k,m}, \dots, \mathcal{F}_W^{k,m} \subseteq \mathcal{Y}_k^n$  so that:

$$\hat{\mathcal{H}}_k = m \iff (X^n, Y_k^n) \in \left( \bigcup_{i=1}^W \mathcal{C}_i \times \mathcal{F}_i^{k,m} \right). \quad (39)$$

For each  $i \in [1 : W]$ , define the set

$$\phi_{i,n} := \left\{ \tilde{P}_X \in \mathcal{P}(\mathcal{X}) : \tilde{P}_X^{\otimes n}[X^n \in \mathcal{C}_i] \geq \frac{1 - \epsilon_{\min}}{W} \right\}. \quad (40)$$

Since the sets  $\mathcal{C}_1, \dots, \mathcal{C}_W$  form a partition of  $\mathcal{X}^n$  and since for each  $\tilde{P}_X \in \mathcal{P}(\mathcal{X})$  it holds that  $\tilde{P}_X^{\otimes n}[X^n \in \mathcal{X}^n] = 1$ , the subsets  $\phi_{1,n}, \dots, \phi_{W,n}$  cover the set  $\mathcal{P}(\mathcal{X})$ . For a similar reason and by the constraint on the type-I error probabilities, for each  $k \in [1 : K]$  and  $m \in [1 : M]$  with  $m \neq k$ , there exists an index  $i(m, k) \in [1 : W]$  so that:

$$P_X^{(m)\otimes n}[X^n \in \mathcal{C}_{i(m,k)}] \geq \frac{1 - \epsilon_{\min}}{W}, \quad (41a)$$

$$P_{Y_k}^{(m)\otimes n}[Y_k^n \in \mathcal{F}_{i(m,k)}^{k,m}] \geq \frac{1 - \epsilon_{\min}}{W}. \quad (41b)$$

Notice that (41) implies that  $P_X^{(m)} \in \phi_{i(m,k)}$ . Combining (41) with the definition of  $\phi_{i(m,k),n}$  in (40) and with [12, Theorem 3] yields that for any  $\mu > 0$  and sufficiently large  $n$ :

$$\Pr[\hat{\mathcal{H}}_k = m \mid \mathcal{H} = k] \geq \max_{\substack{\pi_{XY_k} \\ \pi_X \in \phi_{i(m,k),n} \\ \pi_{Y_k} = P_{Y_k}^{(m)}}} e^{-n(D(\pi_{XY_k} \| P_{XY_k}^{(k)}) + \mu)}.$$

Thus,

$$\begin{aligned} \Pr[\hat{\mathcal{H}}_k \neq k \mid \mathcal{H} = k] \\ \geq \exp \left( -n \min_{\substack{m \in [1:M] \\ m \neq k}} \min_{\substack{\pi_{XY_k} \\ \pi_X \in \phi_{i(m,k),n} \\ \pi_{Y_k} = P_{Y_k}^{(m)}}} D(\pi_{XY_k} \| P_{XY_k}^{(k)}) - n\mu \right) \end{aligned}$$

Taking  $\mu \rightarrow 0$  and  $n \rightarrow \infty$ , by continuity we can conclude that if the exponent vector  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$  is achievable, then there exist subsets  $\phi_1, \dots, \phi_W$  so that

$$\theta_k \leq \min_{\substack{m \in [1:M] \\ m \neq k}} \min_{\substack{\pi_X \in \phi_{b(m)}, \\ \pi_{Y_k} = P_{Y_k}^{(m)}}} D(\pi_{XY_k} \| P_{XY_k}^{(k)}). \quad (42)$$

where  $b(m)$  is the index of the set  $\phi_i$  containing  $P_X^{(m)}$ . Since the upper bounds in (42) become looser when elements are removed from the sets  $\phi_i$ , the converse statement remains valid by imposing that  $\phi_1, \dots, \phi_W$  form a partition of  $\mathcal{P}(\mathcal{X})$ .

Fix now  $\mathbf{r} = (r_1, \dots, r_{K-1}) \in \mathbb{R}^{K-1}$  and consider only the exponent vectors  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$  that satisfy

$$\theta_k = \theta_{k+1} + r_k, \quad k \in \{1, \dots, K-1\}. \quad (43)$$

Then constraints (42) are loosest if  $\phi_i$  contains pmf  $\tilde{P}_X \in \mathcal{P}(\mathcal{X}) \setminus \{P_X^{(1)}, \dots, P_X^{(M)}\}$  only if condition (19) is satisfied.

#### ACKNOWLEDGEMENT

M. Wigger was supported by the ERC Grant CTO Com.

#### REFERENCES

- [1] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, July 1986.
- [2] H. Shimokawa, T. S. Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *Proc. IEEE ISIT'94*, Jun. 1994, p. 114.
- [3] T. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, November 1987.
- [4] M. Rahman and A. Wagner, "On the Optimality of Binning for Distributed Hypothesis Testing," in *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6282–6303, October 2012.
- [5] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *Proc. of Allerton Conference on Comm., Control, and Comp.*, Monticello (IL), USA, Oct. 2014, pp. 1246–1251.
- [6] —, "Distributed testing with zero-rate compression," in *Proc. IEEE ISIT'15*, Hong Kong, Jun. 2015, pp. 2792–2796.
- [7] J. Liao, L. Sankar, F. Calmon, and V. Tan, "Hypothesis testing under maximal leakage privacy constraints," *arXiv:1701.07099*, 2017.
- [8] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing," *arXiv:1604.01292*, 2016.
- [9] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing with communication constraints," in *Proc. of Allerton Conference on Comm., Control, and Comp.*, Monticello (IL), USA, Oct. 2012, pp. 1065–1072.
- [10] S. Salehkalaibar, M. A. Wigger, and L. Wang, "Hypothesis testing in multi-hop networks," *arXiv:1708.05198*, 2017.
- [11] S. Salehkalaibar, M. A. Wigger, and R. Timo, "On hypothesis testing against conditional independence with multiple decision centers," to appear in *IEEE Trans. on Comm.*
- [12] H. M. H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 254–267, Mar. 1992.
- [13] R. Timo, T. Chan, and A. Grant, "Rate distortion with side-information at many decoders," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5240–5257, Aug. 2011.
- [14] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 2–14, Jan 1989.
- [15] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detection," *Simulation available at https://perso.telecom-paristech.fr/wigger/simulation.ipynb*, 2017.