Distributed Hypothesis Testing: Cooperation and Concurrent Detection

Pierre Escamilla, Michèle Wigger, and Abdellatif Zaidi

Abstract—**A single-sensor two-detectors system is considered where the sensor communicates with both detectors and Detector 1 communicates with Detector 2, all over noise-free rate-limited links. The sensor and both detectors observe discrete memoryless source sequences whose joint probability mass function depends on a binary hypothesis. The goal at each detector is to guess the binary hypothesis in a way that, for increasing observation lengths, the probability of error under one of the hypotheses decays to zero with largest possible exponential decay, whereas the probability of error under the other hypothesis can decay to zero or to a small positive number arbitrarily slow. For the setting with positive communication rates from the sensor to the detectors and when both detectors are interested in maximizing the error exponent under the same hypothesis, we characterize the set of all possible exponents in a special case of testing against independence. In this case the cooperation link allows Detector 2 to increase its Type-II error exponent by an amount that is equal to the exponent attained at Detector 1. We also provide a general inner bound on the set of achievable error exponents that shows a tradeoff between the exponents at the two detectors in most cases. When the two detectors aim at maximizing the error exponent under different hypotheses and the distribution at the Sensor is different under the two hypotheses, then we show that such a tradeoff does not exist. We propose a general scheme that allows each detector to attain the same exponent as if it was the only detector in the system. For the setting with zero-rate communication on both links, we exactly characterize the set of possible exponents and the gain brought up by cooperation, in function of the number of bits that are sent over the two links. Notice that, for this setting, tradeoffs between the exponents achieved at the two detectors arise only in few particular cases. In all other cases, each detector achieves the same performance as if it were the only detector in the system.**

I. Introduction

Problems of distributed hypothesis testing are strongly rooted in both statistics and information theory. In particular, [1]–[3] considered a distributed hypothesis testing problem where a single sensor communicates with a single detector over a rate-limited but noise-free link. The goal of [1]–[3] was to determine the largest Type-II error exponent under a fixed(small) bound on the Type-I error exponent. Ahlswede and Csiszár in [1] presented a coding and testing scheme and the corresponding Type-II error exponent for this problem and established optimality of the exponent for the special case of *testing against independence*, i.e., when the distribution of the observations under the alternate hypothesis equals the product of the marginals distributions under the null hypothesis. For the general case, the scheme was subsequently improved by Han [2] and by Shimokawa, Han and Amari [3]. The latter scheme was shown to achieve the optimal exponent in the special case of *testing against conditional independence* by Rahman and Wagner [4]. This line of work has also been extended to networks with multiple sensors [2], [4]–[8], multiple detectors [9], interactive terminals [10]–[12], multi-hop networks [5], [13]–[16], noisy channels [17], [18], and scenarios with privacy constraints [19]–[22].

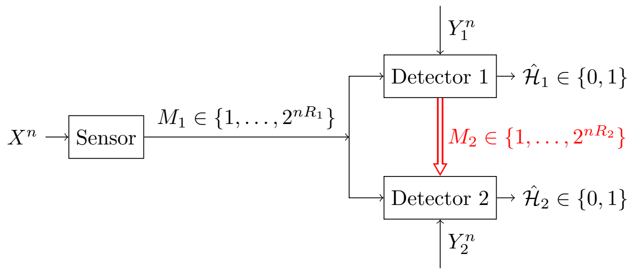

Fig. 1. A Heegard-Berger type source coding model with unidirectional conferencing for multiterminal hypothesis testing.

In this paper, we consider the single-sensor two-detectors system shown in Fig. 1 where Detector 1, after receiving a message from the Sensor, can send a message to Detector 2. This additional message allows the detectors to collaborate in their decision and one of the goals of the paper is to quantify the increase in the Type-II error exponents that is enabled by this collaboration. We show that even a single bit of communication between the detectors (the tentative guess about the hypothesis by the transmitting detector) can provide a large gain in the Type-II error exponent of the detector receiving the bit.

We consider two versions of binary hypothesis testing. In the first version, termed *coherent detection*, both detectors

P. Escamilla is with Paris Research Center, Huawei Technologies, Boulogne-Billancourt, 92100, France. Previously, he has also been with LTCI, Telecom Paris, IP Paris, 91120 Palaiseau, France (e-mail: pierre.escamilla1@huawei.com). M. Wigger is with LTCI, Telecom Paris, IP Paris, 91120 Palaiseau, France (email: michele.wigger@telecom-paris.fr). A. Zaidi is with Université Paris-Est, 77454 Champs-sur-Marne, France and with Paris Research Center, Huawei Technologies, 92100 Boulogne-Billancourt, France (e-mail: abdellatif.zaidi@u-pem.fr). The work of M. Wigger was supported by the European Union's Horizon 2020 Research and Innovation Programme, grant agreement no. 715111. The results in the current paper were presented in part at the 2018 IEEE International Symposium on Information Theory in Vail, (CO), USA, June 2018, and at the 2018 Annual Allerton Conference on Communication, Control, and Computing in Monticello (IL), USA.

are interested in maximizing the error exponent under the same hypothesis. In the second version, termed *concurrent detection*, one of the detectors wishes to maximize the error exponent under hypothesis $\mathcal{H} = 0$ and the other under $\mathcal{H} = 1$. In other words, in this second version the two detectors have concurring goals.

Decentralized detection systems are of major importance for various applications such as autonomous vehicles and complex monitoring systems. These systems use multiple detection pipelines that base their decisions on common or individual observations and often these decisions are combined at one or several central detectors. Our scenario can model both multiple detection pipelines and, through the cooperation link, the process of fusioning various decisions.

The works most closely related to the present manuscript are [9], [14], [15], [23]. The former two, [9], [14], fully characterize the set of possible Type-II error exponents in the special case of testing against independence and testing against conditional independence for a scenario with a single sensor and two non-cooperative detectors. Notice that the presence of a cooperation link between the detectors appears to make the problem of identifying the optimal Type-II exponents significantly more difficult. For example, without cooperation, the set of achievable exponents for testing against independence is achievable with a simple scheme that does not rely on binning. With cooperation, we managed to identify the optimal exponents only under the additional assumption that the observations at the two detectors are independent under both hypotheses and the cooperation rate is zero. In the general case, binning is necessary, and the optimal exponent is yet to be found. Notable exceptions where exponents achieved with binning are shown to be optimal can be found in [4], [6], [7], [17], [18], [24]–[26]. In [23] Zhao and Lai studied our setup in Figure 1 but where Detector 1 does not take any decision (i.e., Detector 1 does not produce $\hat{\mathcal{H}}_1$). As a consequence, in [23], there is no distinction between concurrent or coherent detection. The work [23] identifies the optimal Type-II error exponent for testing against independence when the joint distribution of the observations at the Sensor and Detector 1 is the same under both hypotheses (i.e. (X, Y_1) has the same distribution under both hypotheses $\mathcal{H} = 0$ and $\mathcal{H} = 1$). This special case is only of limited interest in our scenario because under both coherent and concurrent detection Detector 1 cannot achieve any positive Type-II error exponent. We will therefore consider a different testing against independence scenario where under the alternate hypothesis the distribution of the observations at the three terminals is given by the product of the three marginal distributions under the null hypothesis. Zhao and Lai also presented a lower bound on the optimal Type-II error exponent for arbitrary distributions under the two hypotheses. The lower bound is based on a simple scheme that does not use binning.

The third scenario closely related to ours is the multi-hop single-relay network considered in [15]. It differs from the setup of Figure 1 in that Detector 2 does not observe the message M_1 sent by the Sensor. If one wishes that information propagates from the Sensor to Detector 2, then in the multi-hop setup, Detector 1 has to forward this information. The work in [15] presented general inner bounds on the optimal exponents region of the multi-hop setup based on schemes that employ binning. Only coherent detection was considered where both detectors aim at maximizing the error exponents under the same hypothesis. The proposed schemes in [15] apply also to our setup, except that there is no need for Detector 1 to relay the information from the Sensor to Detector 2, because Detector 2 directly observes the Sensor's message M_1. We present the performance of this modified scheme for the simpler version of [15] without binning. The use of binning leads to exponents whose expressions are rather involved; and, for this reason, we sometimes omit them for simplicity and because this is not the main focus of this paper.

The results discussed so far all pertain to positive-rate communication scenarios. Another important line of work assumes zero-rate communication. The single-sensor single-detector version of this problem was addressed in [2] and [27], where Han identified the optimal exponent for the case where only a single bit is communicated and Shalaby and Papamarcou proved that this exponent is also optimal when communication comprises a sublinear number of bits. The finite length regime was investigated in [28]. The optimal Type-II error exponents of zero-rate hypothesis testing in an interactive setup and in a cascaded-encoders network were presented respectively in [29] and [13], [16], [23], [30]. In particular, in our previous work [16], we proved such a result for the single-sensor two-detectors setup. In the current manuscript we extend this result to a cooperative setup.

A. Main Contributions and Organization

In this paper we consider both scenarios of *coherent detection* and *concurrent detection* on the single-sensor two-detectors system in Figure 1. The exponents region can significantly differ under the two, in particular when based on its own observation the sensor can guess the hypothesis (with probability of error $< 1/2$), communicate this guess to the detectors, and adapt the communication to this guess. With this general strategy, the exponents region is a rectangle under concurrent detection, which means that each detector's exponent is the same as in a setup where the other detector is not present. Under coherent detection or concurrent detection when the sensor observations have the same marginal distribution under both hypotheses, the exponents region achieved by our scheme shows a tradeoff between the two exponents. In particular, we can show that the proposed scheme is optimal for a case of testing against independence under coherent detection and when the cooperation message from Detector 1 to Detector 2 has zero rate. This shows that the described tradeoff of the two exponents is not an artifact of our scheme but intrinsic to the problem – it reflects the fact that the message sent by the Sensor has to serve both detectors simulatenously.

We also consider the case with fixed-length communication. Under coherent detection or under concurrent detection when the sensor can send more than a single bit or the sensor observations have the same marginal distributions under the two hypotheses, the exponents region is a rectangle. In these cases, each detector achieves the same exponent as if it were the only one in the system. In contrast, a tradeoff arises under concurrent detection if the sensor can distinguish the two hypotheses but can only send a single bit to the detectors. A comparison with the optimal exponents regions without cooperation [16], allows us to exactly quantify the benefits of detector cooperation in this setup with fixed communication alphabets' sizes. All results summarized in this paragraph remain valid when the alphabets' sizes are not fixed but grow sublinearly in the length of the observed sequences. They also generalize to an arbitrary number of hypotheses. Whereas for two detectors a tradeoff between the exponents arises only when the sensor sends a single bit to the detectors, in a multi-hypothesis testing scenario with $\mathsf{H} \geq 3$ hypotheses such a trade-off can arise whenever the number of communicated bits does not exceed $\log_2 \mathsf{H}$.

The following two tables summarize our main results for the setup with positive communication rates and for fixed communication alphabets:

	Positive Rate
Concurrent Detection	Inner Bound: Prop. 4–5 in Subsec. III-A.
Coherent Detection	Inner Bound: Prop. 6. in Subsec III-B.

and

	Fixed Alphabets
Concurrent Detection	Optimal Region: Prop. 9–10 and Thm. 11 in Subsec. IV-A.
Coherent Detection	Optimal Region: Prop. 12 in Sec. IV-A.

In addition to these main results, the paper also presents a general characterization of the optimal exponents region $\mathcal{E}(R_1, R_2)$ in terms of the optimal exponents of previously studied hypothesis testing problems in the special case of concurrent detection and when $P_X \neq \bar{P}_X$ (Theorem 3 in Section III). Moreover, a computable single-letter characterization of the optimal exponents region $\mathcal{E}(R_1, R_2)$ under coherent detection is given for a case of testing against independence and with zero cooperation rate (Theorem 7 in Subsection III-B). Our result shows that Detector 2 achieves a Type-II error exponent which is given by the summation of the Type-II exponent of Detector 1 with its own Type-II error exponent that it achieves without cooperation.

The remainder of this paper is organized as follows. Section II describes the system model. Sections III and Section IV describe our main results: Section III focuses on communications of positive rates and Section IV on fixed communication alphabets. Technical proofs are referred to appendices. The paper is concluded in Section V.

B. Notation

Throughout, we use the following notation. Sets are denoted by script symbols, e.g., $\mathcal{X}$, random variables by capital letters and their realizations by lower case letters, e.g., X and x. The n-fold product of a set $\mathcal{X}$ is abbreviated as $\mathcal{X}^n$, and a random or deterministic indexed n-tuple $X_1, \ldots, X_n$ or $x_1, \ldots, x_n$ is abbreviated as X^n or as x^n. We typically write P_X or $\bar{P}_X$ for the probability mass function (pmf) of a random variable. The set of all pmfs over an alphabet $\mathcal{X}$ is denoted $\mathcal{P}(\mathcal{X})$.

The *type* of a tuple x^n (i.e., its empirical distribution) [31] is denoted $\text{tp}(x^n)$. We write $\mathcal{P}^n(\mathcal{X})$ for the set of all possible types of n-length sequences over an alphabet $\mathcal{X}$. For $\mu > 0$, the set of sequences x^n that are μ-typical with respect to the probability mass function (pmf) P_X is denoted $\mathcal{T}^n_\mu(P_X)$ [31].

For random variables X and $\bar{X}$ over the same alphabet $\mathcal{X}$ with pmfs P_X and $\bar{P}_X$ satisfying $P_X \ll \bar{P}_X$ (i.e., for every $x_0 \in \mathcal{X}$, if $\bar{P}_X(x_0) = 0$ then also $P_X(x_0) = 0$), both $D(P_X \| \bar{P}_X)$ and $D(X \| \bar{X})$ denote the Kullback-Leiber divergence between X and $\bar{X}$. Finally, $H(X)$ and $I(X;Y)$ denote entropy and mutual information of random variables X and Y. When the joint pmf of these random variables is not clear from the context, we write $I_P(X;Y)$ to indicate that mutual information is meant with respect to the joint pmf P.

II. Formal Problem Statement

Consider a three-terminal problem with a Sensor observing the sequence X^n, a Detector 1 observing Y_1^n, and a Detector 2 observing Y_2^n. The joint pmf of the tuple (X^n, Y_1^n, Y_2^n) depends on one of two hypotheses. Under hypothesis

$$\mathcal{H} = 0\colon \qquad \{(X_t, Y_{1,t}, Y_{2,t})\}_{t=1}^n \text{ i.i.d. } P_{XY_1Y_2} \tag{1}$$

and under hypothesis

$$\mathcal{H} = 1\colon \qquad \{(X_t, Y_{1,t}, Y_{2,t})\}_{t=1}^n \text{ i.i.d. } \bar{P}_{XY_1Y_2} \tag{2}$$

The Sensor applies an encoding function[1]

$$\phi_{1,n}\colon \mathcal{X}^n \to \mathcal{M}_1 := \{0, 1, \ldots, \mathsf{W}_{1,n} - 1\} \tag{3}$$

[1] For convenience, we sometimes also write $M_1 = (a, b)$ for positive integers a, b. We then assume that the numbers satisfy $0 \leq a \leq A-1$ and $0 \leq b \leq B-1$ for some positive integers A and B satisfying $AB = \mathsf{W}_{1,n}$ and there is thus a bijective mapping between the pairs (a, b) and the values $\{0, 1, \ldots, \mathsf{W}_{1,n}\}$. The writing $M_1 = (a, b)$ is then meant as a shorthand notation for $M_1 = aB + b$. The values of A and B will be clear from the context.

to its observed source sequence X^n and sends the resulting index

$$M_1 = \phi_{1,n}(X^n) \tag{4}$$

to both detectors. Detector 1 then applies two functions to the pair (M_1, Y_1^n): an encoding function

$$\phi_{2,n} \colon \mathcal{M}_1 \times \mathcal{Y}_1^n \to \mathcal{M}_2 := \{0, 1, \ldots, \mathsf{W}_{2,n} - 1\}, \tag{5}$$

and a decision function

$$\psi_{1,n} \colon \mathcal{M}_1 \times \mathcal{Y}_1^n \to \{0, 1\}. \tag{6}$$

It sends the index[2]

$$M_2 = \phi_{2,n}(M_1, Y_1^n) \tag{7}$$

to Detector 2, and decides on the hypothesis

$$\hat{\mathcal{H}}_1 := \psi_{1,n}(M_1, Y_1^n). \tag{8}$$

Detector 2 applies a decision function

$$\psi_{2,n} \colon \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{Y}_2^n \to \{0, 1\} \tag{9}$$

to the triple (M_1, M_2, Y_2^n) to produce the decision

$$\hat{\mathcal{H}}_2 := \psi_{2,n}(M_1, M_2, Y_2^n). \tag{10}$$

Both Detectors are required to have vanishing probabilities of error under both hypotheses. Moreover, for Detector 2, we require that the probability of error under $\mathcal{H} = 1$ decays exponentially fast with the largest possible exponent. For Detector 1, we consider two scenarios: *coherent detection* and *concurrent detection*. Under coherent detection, Detector 1 wishes to maximize the exponential decay of the probability of error under $\mathcal{H} = 1$. Under concurrent detection, Detector 1 wishes to maximize the exponential decay of the probability of error under $\mathcal{H} = 0$. In an unifying manner, we define the following error probabilities

$$\alpha_{1,n} := \Pr\left\{\hat{\mathcal{H}}_1 = \bar{h}_1 \middle| \mathcal{H} = h_1\right\}, \tag{11}$$

$$\beta_{1,n} := \Pr\left\{\hat{\mathcal{H}}_1 = h_1 \middle| \mathcal{H} = \bar{h}_1\right\}, \tag{12}$$

$$\alpha_{2,n} := \Pr\left\{\hat{\mathcal{H}}_2 = 1 \middle| \mathcal{H} = 0\right\}, \tag{13}$$

$$\beta_{2,n} := \Pr\left\{\hat{\mathcal{H}}_2 = 0 \middle| \mathcal{H} = 1\right\}, \tag{14}$$

where under coherent detection $h_1 = 0$ and under concurrent detection $h_1 = 1$ and in both cases $\bar{h}_1 := 1 - h_1$.

***Definition** 1 (Achievability under Rate-Constraints):* Given $h_1 \in \{0, 1\}$ and rates $R_1, R_2 \geq 0$, an error-exponents pair (θ_1, θ_2) is said achievable if for all blocklengths n there exist functions $\phi_{1,n}$, $\phi_{2,n}$, $\psi_{1,n}$ and $\psi_{2,n}$ as in (3), (5), (6), and (9) so that the following limits hold:

$$\lim_{n\to\infty} \alpha_{1,n} = 0, \quad \lim_{n\to\infty} \alpha_{2,n} = 0, \tag{15}$$

$$\theta_1 \leq \varliminf_{n\to\infty} -\frac{1}{n} \log \beta_{1,n}, \quad \theta_2 \leq \varliminf_{n\to\infty} -\frac{1}{n} \log \beta_{2,n}, \tag{16}$$

and

$$\varlimsup_{n\to\infty} \frac{1}{n} \log \mathsf{W}_{1,n} \leq R_1, \quad \varlimsup_{n\to\infty} \frac{1}{n} \log \mathsf{W}_{2,n} \leq R_2. \tag{17}$$

***Definition** 2 (Error-Exponents Region under Rate-Constraints):* For any $h_1 \in \{0, 1\}$ and rates $R_1, R_2 \geq 0$ the closure of the set of all achievable exponent pairs (θ_1, θ_2) is called the *error-exponents region* $\mathcal{E}(R_1, R_2)$.

When both rates are zero,

$$R_1 = R_2 = 0, \tag{18}$$

we are also interested in finding the exponents region with fixed communication alphabets of sizes:

$$\mathsf{W}_{1,n} = \mathsf{W}_1 \geq 2, \tag{19a}$$

$$\mathsf{W}_{2,n} = \mathsf{W}_2 \geq 2. \tag{19b}$$

***Definition** 3 (Achievability with Fixed Communication Alphabets):* For any $h_1 \in \{0, 1\}$, $(\epsilon_1, \epsilon_2) \in (0, 1)^2$ and communication alphabet sizes $\mathsf{W}_1, \mathsf{W}_2 \geq 0$, an error-exponents pair (θ_1, θ_2) is said achievable if for all blocklengths n there exist functions $\phi_{1,n}$, $\phi_{2,n}$, $\psi_{1,n}$ and $\psi_{2,n}$ as in (3), (5), (6), and (9) so that

$$\varlimsup_{n\to\infty} \alpha_{1,n} \leq \epsilon_1, \quad \varlimsup_{n\to\infty} \alpha_{2,n} \leq \epsilon_2, \tag{20}$$

and (16) and (19) hold.

***Definition** 4 (Error-Exponents Region for Fixed Communication Alphabets):* For fixed $h_1 \in \{0, 1\}$, $(\epsilon_1, \epsilon_2) \in (0, 1)^2$ and communication alphabet sizes $\mathsf{W}_1, \mathsf{W}_2 \geq 0$, the closure of the set of all achievable exponent pairs (θ_1, θ_2) is called the *error-exponents region* $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$.

In the setup with rate-constraints we require that the type-I error probabilities vanish asymptotically, whereas in the setup with fixed communication alphabets their limits only need to be bounded by the given positive values $\epsilon_1, \epsilon_2 > 0$. All the achievability results in this paper hold with vanishing type-I error probabilities. But the converse results for the positive-rate setup are only proved under the assumption of vanishing type-I error probabilities, and are thus weak converses. In contrast, the converse results for fixed communication alphabets hold also when the type-I error probabilities tend to the positive numbers $\epsilon_1, \epsilon_2 \in (0, 1)$ and are thus strong converses.

In this article we will pay particular attention to the testing against independence scenario under coherent detection, where $h_1 = 0$ and

$$\bar{P}_{XY_1Y_2} = P_X P_{Y_1Y_2}. \tag{21}$$

***Remark** 1:* In some special cases, the described setup degenerates and the error-exponents region is the same as in a setup without cooperation or in a setup with a single detector.

For example, when the Markov chain $X -\!\!\circ\!\!- Y_2 -\!\!\circ\!\!- Y_1$ holds under both hypotheses with identical law $P_{Y_1|Y_2} = \bar{P}_{Y_1|Y_2}$:

$$P_{XY_1Y_2} = P_{XY_2} P_{Y_1|Y_2}, \tag{22a}$$

[2]Similarly to M_1 we sometimes write $M_2 = (a, b)$ for convenience. The meaning is the same as for M_1 and is described in the preceeding footnote.

$$\bar{P}_{XY_1Y_2}=\bar{P}_{XY_2}P_{Y_1|Y_2}, \tag{22b}$$

then the exponents regions $\mathcal{E}(R_1,R_2)$ and $\mathcal{E}_0(\mathsf{W}_1,\mathsf{W}_2,\epsilon_1,\epsilon_2)$ are the same as without cooperation. On the other hand, when the Markov chain $X \multimap Y_1 \multimap Y_2$ holds under both hypotheses with identical law $P_{Y_2|Y_1} = \bar{P}_{Y_2|Y_1}$, i.e.,

$$P_{XY_1Y_2} = P_{XY_1}P_{Y_2|Y_1}, \tag{23a}$$
$$\bar{P}_{XY_1Y_2} = \bar{P}_{XY_1}P_{Y_2|Y_1}, \tag{23b}$$

the exponents regions $\mathcal{E}(R_1,R_2)$ and $\mathcal{E}_0(\mathsf{W}_1,\mathsf{W}_2,\epsilon_1,\epsilon_2)$ correspond to the exponents region of a centralized setup where a single detector observes both Y_1^n and Y_2^n and takes both decisions $\hat{\mathcal{H}}_1$ and $\hat{\mathcal{H}}_2$.

When the communication rates $R_1,R_2 \geq 0$ are sufficiently large, $R_1 \geq H(X)$ and $R_2 \geq H(Y_1|X)$ where entropies are meant according to the pmf P_{XY_1} under $\mathcal{H} = 0$, then the exponents region $\mathcal{E}(R_1,R_2)$ coincides with the exponents region of a fully centralized setup where a single detector observes all X^n, Y_1^n, and Y_2^n and takes both decisions $\hat{\mathcal{H}}_1$ and $\hat{\mathcal{H}}_2$.

A. Related previous results

The model considered in this paper is closely related to the models in [23] and [15]. The main difference of the model studied here to [23] is that in [23] only Detector 2 guesses the binary hypothesis but not Detector 1. In [23] there is thus in particular no distinction between concurrent and coherent detection. The main difference of the model studied here to [15] is that in [15] message M_1 is only observed at Detector 1 but not Detector 2. Moreover, in [15] only coherent detection is considered but not concurrent detection.

The projection of the region $\mathcal{E}(R_1,R_2)$ onto the θ_1-axis characterizes the set of all achievable exponents in a point-to-point (P2P) hypothesis testing problem [2], [3] consisting only of the Sensor and Detector 1. We will denote this largest possible error exponent $\theta^\star_{\text{P2P}}(R_1)$ in the case of coherent detection and $\theta^\star_{\text{P2P,Ex}}(R_1)$ in the case of concurrent detection. So, if $h_1 = 0$ we define

$$\theta^\star_{\text{P2P}}(R_1) := \max\{\theta_1\colon (\theta_1,\theta_2) \in \mathcal{E}(R_1,R_2) \text{ for some } \theta_2 \geq 0\} \tag{24}$$

and if $h_1 = 1$ we define

$$\theta^\star_{\text{P2P,Ex}}(R_1) := \max\{\theta_1\colon (\theta_1,\theta_2) \in \mathcal{E}(R_1,R_2) \text{ for some } \theta_2 \geq 0\}. \tag{25}$$

The projection of the region $\mathcal{E}(R_1,R_2)$ onto the θ_2-axis characterizes the set of all achievable exponents in the cascaded-encoders (CE) hypothesis testing problem as considered in [23]. (This holds for both coherent and concurrent detection.) We denote this exponent by $\theta^\star_{\text{CE}}(R_1,R_2)$:

$$\theta^\star_{\text{CE}}(R_1,R_2) := \max\{\theta_2\colon (\theta_1,\theta_2) \in \mathcal{E}(R_1,R_2) \text{ for some } \theta_1 \geq 0\}. \tag{26}$$

The exponents $\theta^\star_{\text{P2P}}(R_1), \theta^\star_{\text{P2P,Ex}}(R_1)$, and $\theta^\star_{\text{CE}}(R_1,R_2)$ are known only in some special cases. The best known lower bound to $\theta^\star_{\text{P2P}}(R_1)$ was proposed by Shimokawa-Han-Amari [3]. By swapping P_{XY_1} and $\bar{P}_{XY_1}$ their result also provides a lower bound to $\theta^\star_{\text{P2P,Ex}}(R_1)$, which we present next.

Theorem *1 (Obtained from Theorem 1 in [3]):*

$$\theta_{\text{SHA,Ex}}(R_1) \leq \theta^\star_{\text{P2P,Ex}}(R_1), \tag{27}$$

where

$$\theta_{\text{SHA,Ex}}(R_1) := \max_{\bar{P}_{U_1|X}\colon R_1 \geq I_{\bar{P}}(U_1;X|Y_1)} \min\{\eta_1,\eta_2\}, \tag{28}$$

and the mutual information $I_{\bar{P}}(U_1;X|Y_1)$ is computed according to the joint pmf $\bar{P}_{U_1XY_1} := \bar{P}_{U_1|X}\bar{P}_{XY_1}$ and where

$$\eta_1 := \min_{\substack{\tilde{P}_{U_1XY_1}\colon \\ \tilde{P}_{U_1X}=\bar{P}_{U_1X} \\ \tilde{P}_{U_1Y_1}=\bar{P}_{U_1Y_1}}} D(\tilde{P}_{U_1XY_1}\|\bar{P}_{U_1|X}P_{XY_1}), \tag{29}$$

and if $R_1 \geq I_{\bar{P}}(U_1;X)$ then $\eta_2 := \infty$ and otherwise

$$\eta_2 := \min_{\substack{\tilde{P}_{U_1XY_1}\colon \\ \tilde{P}_{U_1X}=\bar{P}_{U_1X} \\ \tilde{P}_{Y_1}=\bar{P}_{Y_1} \\ H_{\tilde{P}}(U_1|Y_1)\leq H_{\bar{P}}(U_1|Y_1)}} D(\tilde{P}_{U_1XY_1}\|\bar{P}_{U_1|X}P_{XY_1}) + R_1 - I_{\bar{P}}(U_1;X|Y_1). \tag{30}$$

Zhao and Lai proposed [23] the following lower bound $\theta_{\text{ZL}}(R_1,R_2)$ for the optimal exponent $\theta^\star_{\text{CE}}(R_1,R_2)$. It can be improved with binning. This would however lead to an expression with more than 10 competing exponents and is omitted for simplicity. For given rates $R_1 \geq 0$ and $R_2 \geq 0$, define the following set of auxiliary random variables:

$$\mathcal{S}(R_1,R_2) := \left\{ (U,V) : \begin{array}{c} U \multimap X \multimap (Y_1,Y_2) \\ V \multimap (Y_1,U) \multimap (Y_2,X) \\ I(U;X) \leq R_1 \\ I(V;Y_1|U) \leq R_2 \end{array} \right\}. \tag{31}$$

Further, define for each $(U,V) \in \mathcal{S}(R_1,R_2)$, the set

$$\mathcal{L}_2(UV) := \left\{ \tilde{P}_{UVXY_1Y_2}\colon \begin{array}{c} P_{\tilde{U}\tilde{X}} = P_{UX} \\ P_{\tilde{U}\tilde{V}\tilde{Y}_1} = P_{UVY_1} \\ P_{\tilde{U}\tilde{V}\tilde{Y}_2} = P_{UVY_2} \end{array} \right\}, \tag{32}$$

Theorem *2 (Obtained from Theorem 5 in [23]):*

$$\theta_{\text{ZL}}(R_1,R_2) \leq \theta^\star_{\text{CE}}(R_1,R_2), \tag{33}$$

where

$$\theta_{\text{ZL}}(R_1,R_2) := \max_{(U,V)\in\mathcal{S}(R_1,R_2)} \min_{\tilde{P}_{UVXY_1Y_2}\in\mathcal{L}_2(UV)} D\left(\tilde{P}_{UVXY_1Y_2}\|P_{V|Y_1U}P_{U|X}\bar{P}_{XY_1Y_2}\right). \tag{34}$$

III. Results for positive rates $R_1 > 0$

In this section, we assume that $R_1 > 0$, whereas the cooperation rate R_2 is 0 or larger. Our results in this section show that the Type-II error exponents is largest under concurrent detection and when $P_X \neq \bar{P}_X$, in which case the exponents region is a rectangle and each detector can achieve the optimal exponent as if it was the only detector in the system. The reason is that in this case, the Sensor can determine with vanishing probability of error whether $\mathcal{H} = 0$ or $\mathcal{H} = 1$, and as a consequence it can adapt its coding scheme to the detected hypothesis and to the detector that wishes to maximize its error exponent under this hypothesis. This strategy is not possible or not useful when $P_X = \bar{P}_X$ or under coherent detection, and our achievable Type-II exponents regions show a tradeoff between the two exponents, because the common communication link from the Sensor has to serve both detectors simultaneously.

A. Concurrent Detection

We first consider concurrent detection. We have the following general result when $P_X \neq \bar{P}_X$.

***Theorem** 3:* If

$$h_1 = 1 \qquad \text{and} \qquad P_X \neq \bar{P}_X, \tag{35}$$

then the exponents region $\mathcal{E}(R_1, R_2)$ is the set of all non-negative pairs (θ_1, θ_2) that satisfy

$$\theta_1 \leq \theta^\star_{\text{P2P,Ex}}(R_1), \tag{36}$$
$$\theta_2 \leq \theta^\star_{\text{CE}}(R_1, R_2), \tag{37}$$

where recall that $\theta^\star_{\text{P2P,Ex}}(R_1)$ is the optimal exponent in the point-to-point hypothesis testing setup in [2], [3] including only the Sensor and Detector 1 but with exchanged roles for the two distributions P_{XY_1} and $\bar{P}_{XY_1}$ (or equivalently for the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$), and recall that $\theta^\star_{\text{CE}}(R_1, R_2)$ is the optimal exponent in the CE hypothesis testing setup introduced by Zhao and Lai [23].

Proof: The converse follows because the exponent at a given detector can only decrease if one imposes an additional constraint on the decision performed by another detector.

Achievability follows from the following scheme. Fix a positive number $\mu > 0$ sufficiently small such that the sets $\mathcal{T}^n_\mu(P_X)$ and $\mathcal{T}^n_\mu(\bar{P}_X)$ do not intersect. The Sensor checks whether

$$X^n \in \mathcal{T}^n_\mu(\bar{P}_X). \tag{38}$$

If this is the case, it applies an optimal encoding for the P2P hypothesis testing problem that includes only the Sensor and Detector 1 and minimizes Detector 1's error exponent under hypothesis $\mathcal{H} = 0$. Letting $\tilde{M}_{\text{P2P}}$ denote the message produced by this optimal encoding, the Sensor sends the message $M_1 = [1, \tilde{M}_{\text{P2P}}]$ to both detectors. After receiving a message of this form, Detector 1 decides as in an optimal P2P scheme that minimizes the error exponent under $\mathcal{H} = 0$. It does not send any message to Detector 2. Detector 2 produces $\hat{\mathcal{H}}_2 = 1$.

If the test in (38) fails, the Sensor checks whether

$$X^n \in \mathcal{T}^n_\mu(P_X). \tag{39}$$

If this condition is fulfilled, all three terminals act as in an optimal scheme for the Zhao-Lai CE hypothesis testing problem [23]. (Recall that the Zhao-Lai setup includes all three terminals and their communication links, but Detector 1 does not produce a decision.) Letting $\tilde{M}_{\text{CE}}$ denote the message produced by this optimal CE encoding, the Sensor sends the message $M_1 = [2, \tilde{M}_{\text{CE}}]$, where the "2" indicates to both detectors to act as in the mentioned optimal CE hypothesis testing scheme. Detector 1 produces the decision $\hat{\mathcal{H}}_1 = 0$. Detector 2 takes the same decision $\mathcal{H}^\star_2$ as the only detector in the optimal CE scheme.

If both (38) and (39) are violated, the Sensor sends $M_1 = 0$ and the two detectors produce $\hat{\mathcal{H}}_1 = 0$ and $\hat{\mathcal{H}}_2 = 1$.

We analyze the described coding scheme. The probability of Type-I error at Detector 2 (so under $\mathcal{H} = 0$) satisfies:

$$\begin{aligned}
\alpha_{2,n} &= \Pr\left[\hat{\mathcal{H}}_2 = 1 \middle| \mathcal{H} = 0\right] \\
&= \Pr\left[\hat{\mathcal{H}}^\star_2 = 1 \text{ or } X^n \notin \mathcal{T}_\mu(P_X) \middle| \mathcal{H} = 0\right] \\
&\leq \Pr\left[\hat{\mathcal{H}}^\star_2 = 1 \middle| \hat{\mathcal{H}}_2 = 0\right] + \Pr\left[X^n \notin \mathcal{T}_\mu(P_X) \middle| \hat{\mathcal{H}}_2 = 0\right] \\
&= \alpha^\star_{2,n} + \epsilon_n,
\end{aligned} \tag{40}$$

where $\alpha^\star_{2,n}$ denotes the Type-I error probability in the optimal CE hypothesis testing scheme and $\epsilon_n \to 0$ when $n \to \infty$. Since $\lim_{n\to\infty} \alpha^\star_{2,n} = 0$, we conclude that $\lim_{n\to\infty} \alpha_{2,n} = 0$. The probability of Type-II error at Detector 2 satisfies:

$$\begin{aligned}
\beta_{2,n} &= Pr\left[\hat{\mathcal{H}}_2 = 0 \middle| \mathcal{H} = 1\right] \\
&= Pr\left[\hat{\mathcal{H}}^\star_2 = 0 \text{ and } X^n \in \mathcal{T}_\mu(P_X) \middle| \mathcal{H} = 1\right] \\
&\leq Pr\left[\hat{\mathcal{H}}^\star_2 = 0 \middle| \mathcal{H} = 1\right] \\
&= \beta^\star_{2,n},
\end{aligned} \tag{41}$$

where $\beta^\star_{2,n}$ is the Type-II error probability of the optimal CE scheme. This proves achievablity of all exponents $\theta_2 \leq \theta^\star_{\text{CE}}(R_1, R_2)$.

Achievability of all exponents $\theta_1 \leq \theta^\star_{\text{P2P}}(R_1)$ can be proved in a similar way. ■

The exponents region in Theorem 3 is rectangular and both exponents can be maximized simultaneously without any tradeoff between the two exponents. As we will see, this is different for the inner bounds we propose for concurrent detection with $P_X = \bar{P}_X$ or for coherent detection, see Propositions 5 and 6 that will follow.

Similarly to the proof of Theorem 3, it can be shown that without cooperation, the optimal exponents region is the set of all non-negative pairs (θ_1, θ_2) that satisfy

$$\theta_1 \leq \theta^\star_{\text{P2P,Ex}}(R_1), \tag{42}$$
$$\theta_2 \leq \theta^\star_{\text{P2P}}(R_1). \tag{43}$$

The benefit of cooperation under coherent detection and when $P_X \neq \bar{P}_X$ is thus that the exponent θ_2 is increased by

$$\theta^\star_{\text{CE}}(R_1, R_2) - \theta^\star_{\text{P2P}}(R_1). \tag{44}$$

In spite of Theorem 3, determining the optimal exponents region $\mathcal{E}(R_1, R_2)$ explicitely remains a difficult problem because the exponents $\theta^\star_{\text{P2P,Ex}}(R_1)$ and

$\theta^\star_{\text{CE}}(R_1, R_2)$ are in general unknown. By Theorem 3 however all explicit lower bounds to these two exponents directly lead to an explicit inner bound on the exponents region $\mathcal{E}(R_1, R_2)$. In the following proposition we present such an explicit inner bound using the lower bound on $\theta^\star_{\text{P2P,Ex}}(R_1)$ in Theorem 1 and the lower bound on $\theta^\star_{\text{CE}}(R_1, R_2)$ obtained in [23, Theorem 5].

***Corollary** 4 (Concurrent Detection and $P_X \neq \bar{P}_X$):* If

$$h_1 = 1 \qquad \text{and} \qquad P_X \neq \bar{P}_X,$$

then the exponents region $\mathcal{E}(R_1, R_2)$ contains all nonnegative pairs (θ_1, θ_2) that for some $(U, V) \in \mathcal{S}(R_1, R_2)$ satisfy:

$$\theta_1 \leq \theta_{\text{SHA,Ex}}(R_1), \tag{45}$$

$$\theta_2 \leq \theta_{\text{ZL}}(R_1, R_2). \tag{46}$$

We next consider concurrent detection but where $P_X = \bar{P}_X$. An inner bound on the achievable error exponent region can be obtained by using a scheme similar to the one described in [23, Section III-A] where Detector 1 uses the information sent by the encoder to produce a typicality test under $\mathcal{H} = 1$.

***Proposition** 5 (Concurrent Detection with $P_X = \bar{P}_X$):* If

$$h_1 = 1 \text{ and } P_X = \bar{P}_X,$$

then the exponents region $\mathcal{E}(R_1, R_2)$ contains all nonnegative pairs (θ_1, θ_2) that for some $(U, V) \in \mathcal{S}(R_1, R_2)$ satisfy:

$$\theta_1 \leq \min_{\substack{\tilde{P}_{UXY_1}: \\ \tilde{P}_{UX} = \bar{P}_{UX} \\ \tilde{P}_{UY_1} = \bar{P}_{UY_1}}} D\left(\tilde{P}_{UXY_1} \| P_{UXY_1}\right), \tag{47a}$$

$$\theta_2 \leq \min_{\tilde{P}_{UVXY_1Y_2} \in \mathcal{L}_2(UV)} D\left(\tilde{P}_{UVXY_1Y_2} \| P_{V|Y_1U} P_{U|X} \bar{P}_{XY_1Y_2}\right), \tag{47b}$$

where $\bar{P}_{UX}$ and $\bar{P}_{UY_1}$ are the marginals of the joint pmf $P_{U|X}\bar{P}_{XY_1}$.

Proof: Similar to the analysis in [23, Theorem 5] and omitted. ∎

In the proposition, U is a random variable that represents a common description of the source X^n which is sent by the Sensor to both detectors, and V is the random variable related to the description of Y_1^n that is sent by Detector 1 to Detector 2. This latter description is superimposed on the common description U.

Notice that while in Corollary 4 each of the two exponents θ_1 and θ_2 is optimized independently over the auxiliary parameters (conditional pmfs and rates), this is not the case in above Proposition 5. In this latter proposition, generally a tradeoff occurs between the two exponents.

B. Coherent Detection

We finally consider coherent detection $h_1 = 0$. The following achievable error exponents region is obtained by modifying the scheme in [23, Section III-A] so that Detector 1 uses the information sent by the Sensor to guess the hypothesis. Specificallly, it declares $\mathcal{H} = 0$ if obtained codeword and its own source Y_1^n are jointly typical according to P_{UY_1}.

***Proposition** 6 (Coherent Detection):* If

$$h_1 = 0,$$

the exponents region $\mathcal{E}(R_1, R_2)$ contains all nonnegative pairs (θ_1, θ_2) that satisfy:

$$\theta_1 \leq \min_{\substack{\tilde{P}_{UXY_1}: \\ \tilde{P}_{UX} = P_{UX} \\ \tilde{P}_{UY_1} = P_{UY_1}}} D(\tilde{P}_{UXY_1} \| P_{U|X} \bar{P}_{XY_1}), \tag{48a}$$

$$\theta_2 \leq \min_{\tilde{P}_{UVXY_1Y_2} \in \mathcal{L}_2(UV)} D\left(\tilde{P}_{UVXY_1Y_2} \| P_{V|Y_1U} P_{U|X} \bar{P}_{XY_1Y_2}\right). \tag{48b}$$

for some $(U, V) \in \mathcal{S}(R_1, R_2)$ (see (31) and (32)).

Proof: Similar to the analysis in [23] and omitted. ∎

As in the previous Proposition 5, U and V relate to the superpositioned compression codewords sent from the Sensor to both detectors and from Detector 1 to Detector 2. The difference between the two propositions are the constraints in the minimizations describing θ_1, see (47a) and (48a). The reason is that under concurrent detection, Detector 1 aims to maximize the error exponent under $\mathcal{H} = 0$ and therefore checks typicality with respect to the law under $\mathcal{H} = 1$, whereas under coherent detection, it aims to maximize the error exponent under $\mathcal{H} = 1$ and therefore checks typicality with respect to the law under $\mathcal{H} = 0$. If the checks fail, Detector 1 declares $\mathcal{H} = 0$ under concurrent detection and $\mathcal{H} = 1$ under coherent detection.

We investigate the special case of "testing-against-independence" scenario under coherent detection, $h_1 = 0$, where

$$P_{XY_1Y_2} = P_{X|Y_1Y_2} P_{Y_1} P_{Y_2}, \tag{49}$$

$$\bar{P}_{XY_1Y_2} = P_X P_{Y_1} P_{Y_2}. \tag{50}$$

This setup differs from the testing-against independence scenario in [23] where $P_{XY_1} = \bar{P}_{XY_1}$ and Detector 1 cannot obtain a positive error exponent.

We assume a cooperation rate $R_2 = 0$, which means that Detector 1 can send a message M_2 to Detector 2 that is described by a sublinear number of bits.

***Theorem** 7 (Testing Against Independence):* Assume $h_1 = 0$ and (49). Then, $\mathcal{E}(R_1, 0)$ is the set of all nonnegative exponent pairs (θ_1, θ_2) for which

$$\theta_1 \leq I(U; Y_1), \tag{51a}$$

$$\theta_2 \leq I(U; Y_1) + I(U; Y_2), \tag{51b}$$

for some U satisfying the Markov chain $U \multimap X \multimap (Y_1, Y_2)$ and the rate constraint $R_1 \geq I(U; X)$.

Proof: The achievability follows by specializing and evaluating Proposition 6 for this setup. In particular no binning is required. The converse is proved in Appendix A. ∎

***Lemma** 8 (Cardinality bound):* Theorem 7 remains valid if we impose the cardinality bound $|\mathcal{U}| = |\mathcal{X}| + 2$.

Proof: Similar to the proof of [2, Theorem 3]. ∎

The performance in Theorem 7 is obtained by letting the Sensor send a compression codeword U^n to both detectors and each detector checks whether the obtained codeword and its own source observation Y_1^n or Y_2^n are jointly typical with respect to the distribution under $\mathcal{H} = 0$. Moreover, Detector 1 sends the outcome of this test to Detector 2. It thus sends only a single bit over the cooperation link (even though by the definition of zero rate it would be allowed to send a sublinear number of bits). Detector 1 decides on $\mathcal{H} = 0$ if and only if its own test was successful, and Detector 2 decides on $\mathcal{H} = 0$ if and only if both tests were successful. Detector 1 thus achieves the exponent $I(U; Y_1)$ as in Han's single-detector point-to-point system, and Detector 2 achieves the sum of the two mutual information terms $I(U; Y_1)$ and $I(U; Y_2)$. Here, the former stems from the test at Detector 1 (whose outcome is conveyed to Detector 2) and the latter from the test at Detector 2.
Without cooperation, Detector 2 only achieves an exponent equal to $I(U; Y_2)$ [9]. Hence, the benefit of a single cooperation bit from Detector 1 to Detector 2 is quantified by $I(U; Y_1)$.

***Remark** 2:* The problem of evaluating the rate-exponents region in Theorem 7, and so the optimal test channel $P^\star_{U|X}$ that exhausts this region, is generally non-convex. In fact, even the single-detector version in which one does not care about θ_2 is non-convex. More precisely, for that setting the problem is convex on each of the distributions P_U, $P_{U|X}$ and $P_{Y_1|U}$ independently but is not convex on the joint distribution P_{UXY_1}. Iterative Blahut-Arimoto type algorithms whose convergence is guaranteed but not necessarily to the optimal solution are possible - the reader may refer to [32] and [8] where algorithms are developed for a similar setting. (The results of those algorithms were observed numerically to coincide with those of exhaustive search for all the examples considered therein.)

In the following example 1, the source alphabets are small. For this reason, and also accounting for the above, we resorted to exhaustive search to obtain our numerical results.

We illustrate the benefit of cooperation for testing against independence at hand of the following example.

***Example** 1:* Consider a setup with coherent detection, $h_1 = 0$, where X, Y_1, Y_2 are ternary and under $\mathcal{H} = 0$:

$$\begin{cases} P_{XY_1Y_2}(0,0,0) = 0.05 & P_{XY_1Y_2}(0,0,1) = 0.05 \\ P_{XY_1Y_2}(0,1,0) = 0.15 & P_{XY_1Y_2}(0,1,1) = 0.083325 \\ P_{XY_1Y_2}(1,0,0) = 0.05 & P_{XY_1Y_2}(1,0,1) = 0.15 \\ P_{XY_1Y_2}(1,1,0) = 0.05 & P_{XY_1Y_2}(1,1,1) = 0.08335 \\ P_{XY_1Y_2}(2,0,0) = 0.15 & P_{XY_1Y_2}(2,0,1) = 0.05 \\ P_{XY_1Y_2}(2,1,0) = 0.05 & P_{XY_1Y_2}(2,1,1) = 0.083325 \end{cases} \tag{52}$$

whereas under $\mathcal{H} = 1$ they are independent with same marginals as under $\mathcal{H} = 0$.

Fig. 2 illustrates the error-exponents regions for various cooperation scenarios when the communication rate from the Sensor to the detectors is $R_1 = 0.1$ bits. The dark blue curve shows the optimal exponents region if there is no cooperation among the two detectors, see [15, Theorem 1]. The light blue curve shows the optimal exponents region $\mathcal{E}(0.1, 0)$ under zero-rate cooperation, see Theorem 7. As explained after Lemma 8, this exponent can be achieved by sending only a single cooperation bit indicating Detector 1's decision. The observations of the detectors and the sensor being independent under the alternative hypothesis, the decision of Detector 1 is complementary to that of Detector 2 and will strongly enhance its decision. The light grey region shows the achievable error-exponents region of Proposition 6 for $R_2 = 1$. All three regions show some trade-off between the two exponents θ_1 and θ_2. Moreover, from the figure we observe that even a single bit of cooperation can enlarge the exponents region significantly.

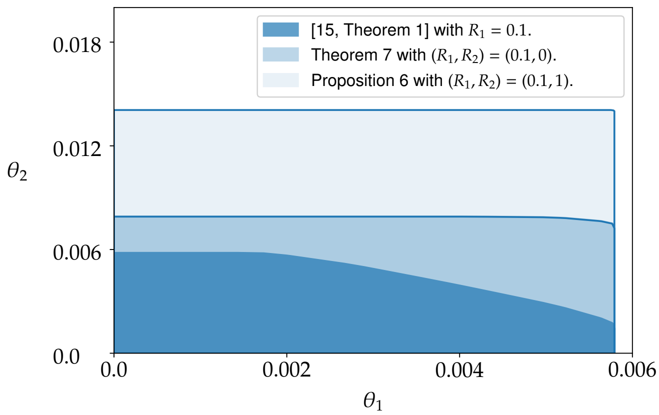

Fig. 2. Comparison of error-exponents regions for different cooperation rates as studied in Example 1.

IV. Results for fixed communication alphabets

We now present our results for the fixed-alphabets case, so we assume (19) and are interested in the error-exponents region $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$. For simplicity, we assume that $P_{XY_1}(x, y_1) > 0$ and $\bar{P}_{XY_1Y_2}(x, y_1, y_2) > 0$ for all $(x, y_1, y_2) \in \mathcal{X}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_2$.[3]

A. Optimal Exponents Regions

The optimal exponents region $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$ for coherent detection can be characterized given straightforward arguments developed in [27, Section IV]. We state them here without proof. Our main finding in this section is the exact characterization of the optimal exponents region $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$ under concurrent detection. Examining the converse proofs, it can be verified

[3]These assumptions are technicalities and ensure that all terms used in the following are finite. Similar conditions were also present in [27] and subsequently relaxed in [33]. Our requirement ensures that all expressions are finite. The condition that $\bar{P}_{XY_1Y_2}$ be positive is required since (both under coherent and concurrent detection) Detector 2 aims at maximizing the error exponent under hypothesis $\mathcal{H} = 1$, and the conditions that $\bar{P}_{XY_1}$ and P_{XY_1} be positive are required because Detector 1 aims at maximizing the error exponent under any of the two hypotheses, depending on whether we consider coherent or concurrent detection.

that our results remain valid when the alphabet sizes are not fixed but grow sublinearly in the blocklength n, i.e.,

$$\lim_{n\to\infty} \frac{\mathsf{W}_{i,n}}{n} = 0, \qquad i \in \{1,2\}. \tag{53}$$

Consider first the case of concurrent detection with $P_X = \bar{P}_X$.

***Proposition** 9 (Concurrent Detection with $P_X = \bar{P}_X$):* Under concurrent detection, i.e., $h_1 = 1$, for all $(\epsilon_1, \epsilon_2) \in (0,1)^2$, and when $P_X = \bar{P}_X$, then for all values $\mathsf{W}_1 \geq 2$ and $\mathsf{W}_2 \geq 2$, the exponents region $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$ is the set of all nonnegative rate pairs (θ_1, θ_2) satisfying

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}:\ \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big), \tag{54}$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}:\ \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1},\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big). \tag{55}$$

Proof: The converse holds because (54) characterizes the largest exponent that Detector 1 could achieve if it was the only detector in the systems and (55) characterizes the largest exponent that Detector 2 could achieve if it was the only detector in the systems. (For this latter setup, Detector 1 is still present but does not take a decision.) Achievability follows by combining two instances of the scheme described in [27]: one instance aims at achieving exponent (54) at Detector 1 and one instance aims at achieving exponent (55) at Detector 2. (Two instances are necessary because the scheme in [27] only considers coherent detection.) It can be shown that with $\mathsf{W}_1 = 2$ and $\mathsf{W}_2 = 2$ it is possible to simultaneously describe both schemes. ■

We now focus on the case where $P_X \neq \bar{P}_X$. Here the optimal exponents region depends on whether the alphabet size W_1 equals 2 or is larger. We first assume

$$\mathsf{W}_1 \geq 3 \qquad \text{and} \qquad \mathsf{W}_2 \geq 2. \tag{56}$$

***Proposition** 10 (Concurrent Detection, $P_X \neq \bar{P}_X$ and $\mathsf{W}_1 \geq 3$):* Under concurrent detection, for all $(\epsilon_1, \epsilon_2) \in (0,1)^2$ and for all values $\mathsf{W}_1 \geq 3$ and $\mathsf{W}_2 \geq 2$, the exponents region $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$ is the set of all nonnegative rate pairs (θ_1, θ_2) satisfying

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}:\ \tilde{P}_X = \bar{P}_X \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big), \tag{57}$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}:\ \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1},\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big). \tag{58}$$

Proof: Similar to the proof of Proposition 9. The only difference is that since $P_X \neq \bar{P}_X$ two different values are required to indicate that a sequence is typical according to P_X or that it is typical according to $\bar{P}_X$. Therefore $\mathsf{W}_1 = 3$ and $\mathsf{W}_2 = 2$ are required to simultaneously describe the communication for both instances of the scheme in [27] that target Detectors 1 and 2. ■

The exponents region $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$ in these first two Propositions 9–10 is rectangular, and each of the detectors can simultaneously achieve the optimal exponent as if it were the only detector in the system. As we see in the following, this is not always the case.

For any real number r and function $\mathsf{b}\colon \{0,1\} \to \{0,1\}$ that is either

$$\mathsf{b}(0) = \mathsf{b}(1) = 0, \tag{59}$$

or

$$\mathsf{b}(0) = 0 \qquad \text{and} \qquad \mathsf{b}(1) = 1, \tag{60}$$

define the sets of pmfs $\Gamma_0^{\mathsf{b},r}, \Gamma_1^{\mathsf{b},r} \in \mathcal{P}(\mathcal{X})$ as follows. If $\mathsf{b}(0) = \mathsf{b}(1) = 0$, then the pmfs P_X and $\bar{P}_X$ are assigned to $\Gamma_0^{\mathsf{b},r}$ and all other pmfs to $\Gamma_1^{\mathsf{b},r}$. Otherwise, if $\mathsf{b}(0) = 0$ and $\mathsf{b}(1) = 1$, then

$$\pi \in \Gamma_1^{\mathsf{b},r}$$
$$\Longleftrightarrow$$
$$\min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X = \pi \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big) + r \geq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X = \pi \\ \tilde{P}_{Y_1} = P_{Y_1} \\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big), \tag{61}$$

and $\pi \in \Gamma_0^{\mathsf{b},r}$ otherwise. In particular, in this case $P_X \in \Gamma_0^{\mathsf{b},r}$ and $\bar{P}_X \in \Gamma_1^{\mathsf{b},r}$.

***Theorem** 11 (Concurrent Detection, $P_X \neq \bar{P}_X$, and $\mathsf{W}_1 = 2$):* Under concurrent detection, for all $(\epsilon_1, \epsilon_2) \in (0,1)^2$ and for all values $\mathsf{W}_1 = 2$ and $\mathsf{W}_2 \geq 2$, the exponents region $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$ is the set of all non-negative rate pairs (θ_1, θ_2) that satisfy

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_{\mathsf{b}(1)}^{\mathsf{b},r} \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big), \tag{62}$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_{\mathsf{b}(0)}^{\mathsf{b},r}, \\ \tilde{P}_{Y_1} = P_{Y_1},\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big), \tag{63}$$

for some real number r, a mapping $\mathsf{b}\colon \{0,1\} \to \{0,1\}$ that is either

$$\mathsf{b}(0) = \mathsf{b}(1) = 0 \tag{64}$$

or

$$\mathsf{b}(0) = 0 \qquad \text{and} \qquad \mathsf{b}(1) = 1, \tag{65}$$

and where the sets of pmfs $\Gamma_{\mathsf{b}(0)}^{\mathsf{b},r}$ and $\Gamma_{\mathsf{b}(1)}^{\mathsf{b},r}$ are defined ahead of the theorem.

Proof: See Appendix B. ■

Notice that the coding scheme presented in Appendix B differs from previous known schemes for fixed communication alphabets such as [2], [27], [34].

Consider now the case of coherent detection (both with $P_X = \bar{P}_X$ and $P_X \neq \bar{P}_X$)

***Proposition** 12 (Coherent Detection): Under coherent detection, $h_1 = 0$, for all $(\epsilon_1, \epsilon_2) \in (0,1)^2$, and for all values* $\mathsf{W}_1 \geq 2$

and $\mathsf{W}_2 \geq 2$, the exponents region $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2, \epsilon_1, \epsilon_2)$ is the set of all nonnegative rate pairs (θ_1, θ_2) satisfying

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}:\ \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| \bar{P}_{XY_1}\Big), \tag{66}$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}:\ \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1},\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big). \tag{67}$$

Proof: *Analogous to the proofs in [27]. Consider in particular the extensions in [27, Section IV].* ■

Remark *3 (Extension to many hypotheses):* Most of the results in this section can be extended in a straightforward manner to a scenario with more than two hypotheses. For $\mathsf{H} = 2$ hypotheses the exponents region showed a tradeoff in the exponents under concurrent detection only when $\mathsf{W}_1 = \mathsf{W}_2 = 2$. In contrast, for $\mathsf{H} \geq 3$ hypotheses, a tradeoff arises for a variety of pairs $\mathsf{W}_1, \mathsf{W}_2$. In general, the minimum required values for W_1 and W_2 leading to a rectangular exponents region coincides respectively with the number of hypotheses which have distinct X-marginals and the number of hypotheses which have distinct Y_1-marginals.

B. Benefits of Cooperation

To discuss the benefits of cooperation, we quickly state the optimal exponents region without cooperation determined in [16]. Assume thus:[4]

$$\mathsf{W}_2 = 0. \tag{68}$$

Under coherent detection or under concurrent detection with $P_X = \bar{P}_X$ or $\mathsf{W}_1 \geq 3$, the exponents regions $\mathcal{E}_0(\mathsf{W}_1, \mathsf{W}_2 = 0)$ are similar to Propositions 9–10 but with a modified constraint on θ_2. More precisely, Propositions 9–10 remain valid for $\mathsf{W}_2 = 0$ if the constraints on θ_2, (55) and (58), are replaced by

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_2}: \\ \tilde{P}_X = P_X, \\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_2} \| \bar{P}_{XY_2}\Big). \tag{69}$$

So, in these scenarios, the exponents region is a rectangle both in the case with and without cooperation, and when cooperation is possible the θ_2-side of the rectangle is increased by the quantity

$$\min_{\substack{\tilde{P}_{XY_1Y_2}:\ \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1},\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big) - \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X = P_X; \\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_2} \| \bar{P}_{XY_2}\Big). \tag{70}$$

Under concurrent detection when $P_X \neq \bar{P}_X$ and $\mathsf{W}_1 = 2$, the exponents region is not a rectangle, but there is a tradeoff between the two exponents. In this case, it seems difficult to quantify the cooperation benefit in general.

[4]Equivalently, the no cooperation setup could be parametrized as $\mathsf{W}_2 = 1$.

C. Numerical Example

We now present an example with $P_X \neq \bar{P}_X$.

Example *2:* Consider a setup where

$$\begin{cases} P_{XY_1Y_2}(0,0,0) = 0.1 & P_{XY_1Y_2}(0,0,1) = 0.1125 \\ P_{XY_1Y_2}(0,1,0) = 0.0875 & P_{XY_1Y_2}(0,1,1) = 0.0825 \\ P_{XY_1Y_2}(1,0,0) = 0.1675 & P_{XY_1Y_2}(1,0,1) = 0.1625 \\ P_{XY_1Y_2}(1,1,0) = 0.1375 & P_{XY_1Y_2}(1,1,1) = 0.15 \end{cases} \tag{71}$$

and

$$\begin{cases} \bar{P}_{XY_1Y_2}(0,0,0) = 0.15 & \bar{P}_{XY_1Y_2}(0,0,1) = 0.1375 \\ \bar{P}_{XY_1Y_2}(0,1,0) = 0.1625 & \bar{P}_{XY_1Y_2}(0,1,1) = 0.1675 \\ \bar{P}_{XY_1Y_2}(1,0,0) = 0.0825 & \bar{P}_{XY_1Y_2}(1,0,1) = 0.0875 \\ \bar{P}_{XY_1Y_2}(1,1,0) = 0.1125 & \bar{P}_{XY_1Y_2}(1,1,1) = 0.1 \end{cases} \tag{72}$$

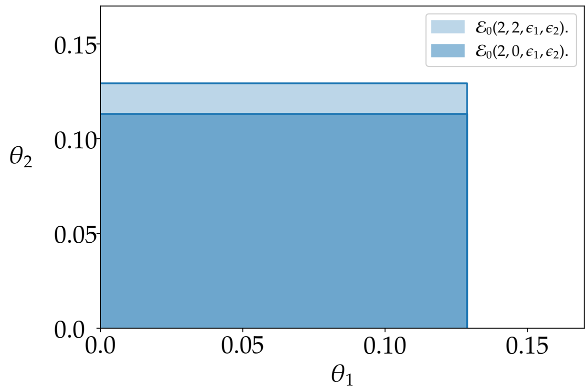

Fig. 3. Exponents regions of Example 2 under coherent detection, without cooperation and with a single-bit cooperation message.

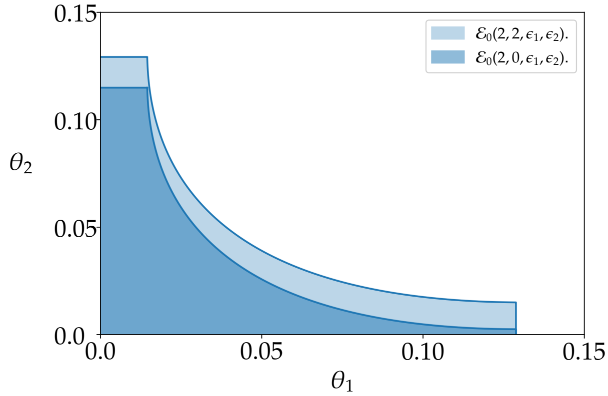

Fig. 4. Exponents regions of Example 2 under coherent detection, without cooperation and with a single-bit cooperation message.

The exponents regions for this example are depicted in Fig. 3 (for coherent detection) and Fig. 4 (for concurrent detection) assuming that the Sensor sends a single bit to the two detectors, i.e., $\mathsf{W}_1 = 2$. The figures illustrate the exponents regions for the scenario without cooperation and the scenario with a single-bit cooperation message. From these figures one observes an almost-uniform cooperation benefit over all achievable θ_1-values. Under concurrent detection one further observes a tradeoff between the achievable exponents θ_1 and θ_2, both in the case with and without cooperation.

V. Summary and conclusion

In this paper we investigated the role of cooperation under both coherent and concurrent detection in a two-detector hypothesis testing system. For the general positive-rate scenario, we proposed a simple scheme in which Detector 1 uses the cooperation link to inform Detector 2 about its guess and a compressed version of its observations. Under concurrent detection with unequal marginals $P_X \neq \bar{P}_X$, in our scheme the Sensor makes a tentative guess of whether $\mathcal{H}=0$ or $\mathcal{H}=1$. Depending on the outcome of this test, it decides to target its communication only for the decision taken at Detector 1 or at Detector 2. This strategy at the sensor is shown to be optimal and to lead to a rectangular exponents region where each detector achieves the optimal performance as if it was the only detector in the system. We further present the optimal exponents region for both coherent and concurrent detection when the communication alphabets are fixed or of zero rate. In most cases this optimal exponents region is rectangular and no tradeoff arises between the two decisions. However, when $P_X \neq \bar{P}_X$ and the sensor can communicate only a single bit, then under concurrent detection there is a tradeoff in the exponents achieved at the two detectors.

Appendix A
Converse to Theorem 7

Let $R_2 = 0$. Fix a rate $R_1 \geq 0$ and a pair of exponents $(\theta_1, \theta_2) \in \mathcal{E}_0(R_1, 0)$. Then, choose an $\epsilon \in (0,1)$, a sufficiently large blocklength n, encoding and decision functions $\phi_{1,n}$, $\phi_{2,n}$, $\psi_{1,n}$, and $\psi_{2,n}$ that satisfy

$$\alpha_{1,n} \leq \epsilon, \tag{73}$$

$$\alpha_{2,n} \leq \epsilon, \tag{74}$$

and

$$-\frac{1}{n}\log \beta_{1,n} \geq \theta_1 - \epsilon, \tag{75}$$

$$-\frac{1}{n}\log \beta_{2,n} \geq \theta_2 - \epsilon. \tag{76}$$

Notice first that for each $i \in \{1,2\}$ [11]:

$$D\Big(P_{\hat{\mathcal{H}}_i|\mathcal{H}=0} \| P_{\hat{\mathcal{H}}_i|\mathcal{H}=1}\Big) = -h_2(\alpha_{i,n}) - (1-\alpha_{i,n})\log(\beta_{i,n}) - \alpha_{i,n}\log(1-\beta_{i,n}) \tag{77}$$

where $h_2(p)$ denotes the entropy of a Bernouilli-(p) memoryless source. Since $\alpha_{i,n} \leq \epsilon < 1$, for each $i \in \{1,2\}$, Inequality (77) yields:

$$-\frac{1}{n}\log(\beta_{i,n}) \leq \frac{1}{n(1-\epsilon)} D\Big(P_{\hat{\mathcal{H}}_i|\mathcal{H}=0} \| P_{\hat{\mathcal{H}}_i|\mathcal{H}=1}\Big) + \mu_n \tag{78}$$

with $\mu_n := \frac{1}{n(1-\epsilon)} h_2(\epsilon)$. Notice that $\mu_n \to 0$ as $n \to \infty$.

Consider now:

$$\begin{aligned}
\theta_1 - \epsilon &\leq -\frac{1}{n}\log(\beta_{1,n}) \\
&\leq \frac{1}{n(1-\epsilon)} D\Big(P_{\hat{\mathcal{H}}_1|\mathcal{H}=0} \| P_{\hat{\mathcal{H}}_1|\mathcal{H}=1}\Big) + \mu_n \\
&\stackrel{(a)}{\leq} \frac{1}{n(1-\epsilon)} D\Big(P_{Y_1^n M_1|\mathcal{H}=0} \| P_{Y_1^n M_1|\mathcal{H}=1}\Big) + \mu_n \\
&\stackrel{(b)}{=} \frac{1}{n(1-\epsilon)} I\Big(Y_1^n; M_1\Big) + \mu_n \\
&\stackrel{(c)}{=} \frac{1}{n(1-\epsilon)} \left(\sum_{k=1}^n H\Big(Y_{1k}\Big) - H\Big(Y_{1k}|M_1 Y_1^{k-1}\Big)\right) + \mu_n \\
&\stackrel{(d)}{\leq} \frac{1}{n(1-\epsilon)} \left(\sum_{k=1}^n H\Big(Y_{1k}\Big) - H\Big(Y_{1k}|M_1 Y_1^{k-1} X^{k-1}\Big)\right) + \mu_n \\
&\stackrel{(e)}{=} \frac{1}{n(1-\epsilon)} \left(\sum_{k=1}^n H\Big(Y_{1k}\Big) - H\Big(Y_{1k}|M_1 X^{k-1}\Big)\right) + \mu_n \\
&\stackrel{(f)}{=} \frac{1}{n(1-\epsilon)} \left(\sum_{k=1}^n I(Y_{1k}; U_k)\right) + \mu_n \\
&\stackrel{(g)}{=} \frac{1}{n(1-\epsilon)} I\Big(Y_{1Q}; U_Q \Big| Q\Big) + \mu_n \\
&\stackrel{(h)}{=} \frac{1}{1-\epsilon} I(Y_1(n); U(n)) + \mu_n,
\end{aligned} \tag{79}$$

where: (a) follows by the data processing inequality for relative entropy; (b) holds since M_1 and Y_1^n are independent under the alternative hypothesis $\mathcal{H}=1$; (c) is due to the chain rule for mutual information; (d) follows since conditioning reduces entropy; (e) is due to the Markov chain $Y_1^{k-1} -\!\!\circ\!\!- (M_1, X^{k-1}) -\!\!\circ\!\!- Y_{1k}$; ($f$) holds by defining $U_k := (M_1, X^{k-1})$; (g) is obtained by introducing a random variable Q that is uniform over the set $\{1, \cdots, n\}$ and independent of all previously defined random variables; and (h) holds by defining $U(n) := (U_Q, Q)$ and $Y_1(n) := Y_{1Q}$.

In a similar way, one obtains:

$$\begin{aligned}
&\theta_2 - \epsilon \\
&\leq -\frac{1}{n}\log(\beta_{2,n}) \\
&\stackrel{(i)}{\leq} \frac{1}{n(1-\epsilon)} D\Big(P_{Y_2^n M_1 M_2|\mathcal{H}=0} \| P_{Y_2^n M_1 M_2|\mathcal{H}=1}\Big) + \mu_n \\
&\stackrel{(j)}{=} \frac{1}{n(1-\epsilon)} \Big(I\Big(Y_2^n; M_1 M_2\Big) + D(P_{M_1 M_2|\mathcal{H}=0} \| P_{M_1 M_2|\mathcal{H}=1})\Big) + \mu_n \\
&\stackrel{(k)}{\leq} \frac{1}{n(1-\epsilon)} \Big(I\Big(Y_2^n; M_1\Big) + I\Big(Y_2^n; M_2|M_1\Big) \\
&\qquad + D(P_{Y_1^n M_1|\mathcal{H}=0} \| P_{Y_1^n M_1|\mathcal{H}=1})\Big) + \mu_n \\
&\stackrel{(\ell)}{\leq} \frac{1}{n(1-\epsilon)} \Big(I(Y_2^n; M_1) + \log \mathsf{W}_{2,n} + D(P_{Y_1^n M_1|\mathcal{H}=0} \| P_{Y_1^n M_1|\mathcal{H}=1})\Big) \\
&\quad + \mu_n \\
&\stackrel{(m)}{=} \frac{1}{n(1-\epsilon)} \Big(I\Big(Y_2^n; M_1\Big) + I(Y^n{}_1; M_1)\Big) + \tilde{\mu}_n \\
&\stackrel{(o)}{\leq} \frac{1}{1-\epsilon} (I(Y_2(n); U(n)) + I(Y_1(n); U(n))) + \tilde{\mu}_n,
\end{aligned} \tag{80}$$

where (i) follows by the data processing inequality for relative entropy; (j) holds by the independence of the pair (M, M_2) with Y_2^n under the alternative hypothesis $\mathcal{H}=1$; (k) by the data processing inequality for relative entropy; (ℓ) holds because $I(Y_2^n; M_2|M_1) \leq H(M_2) \leq \log \mathsf{W}_{2,n}$; ($o$) follows by proceeding along the steps (b)

to (*h*) above; and (*m*) holds by defining $\tilde{\mu}_n := n^{-1}(1-\epsilon)^{-1}\log \mathsf{W}_{2,n} + \mu_n$.

Notice that by the assumption $R_2 = 0$, the term $1/n \log \mathsf{W}_{2,n} \to 0$ as $n \to \infty$. Thus, also $\tilde{\mu}_n \to 0$ as $n \to \infty$. We next lower bound the rate R_1:

$$\begin{aligned} nR_1 &\geq H(M_1) \\ &= H(M_1) - H(M|X^n) \\ &= I(M_1; X^n) \\ &= \sum_{k=1}^{n} I\left(M_1; X_k | X^{k-1}\right) \\ &= \sum_{k=1}^{n} I(X_k; U_k) \\ &= nI\left(X_Q; U_Q | Q\right) \\ &= nI(U(n); X(n)). \end{aligned} \tag{81}$$

For any blocklength n, the newly defined random variables satisfy $X(n), Y_1(n), Y_2(n) \sim P_{XY_1Y_2}$ and $U(n) -\!\!\circ\!\!- X(n) -\!\!\circ\!\!- (Y_1(n), Y_2(n))$. Notice that in the derived inequalities (79), (80), and (81), only the probability masses matter but not the specific values of $U(n)$. By standard applications of Carathéodory's theorem, it can therefore be shown that there exists a finite set $\mathcal{U}$ and for each blocklength n random variables $\tilde{U}(n), \tilde{X}(n), \tilde{Y}_1(n), \tilde{Y}_2(n)$ over $\mathcal{U} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$ satisfying the three inequalities (79), (80), and (81) (when the tuple $U(n), X(n), Y_1(n), Y_2(n)$ is replaced by the new tuple $\tilde{U}(n), \tilde{X}(n), \tilde{Y}_1(n), \tilde{Y}_2(n)$) and the properties $\tilde{U}_n -\!\!\circ\!\!- \tilde{X}(n) -\!\!\circ\!\!- (\tilde{Y}_1(n), \tilde{Y}_2(n))$ and $\tilde{X}(n), \tilde{Y}_1(n), \tilde{Y}_2(n) \sim P_{XY_1Y_2}$. Since $\mathcal{U}$ is fixed, we can employ the Bolzano-Weierstrass theorem to deduce that there exists an increasing sequence of positive integers $n_1, n_2, \ldots$ such that the subsequence of distributions $\{P_{U(n_k)X(n_k)Y_1(n_k)Y_2(n_k)}\}_{k=1}^{\infty}$ converges to a limiting distribution $P^{\star}_{UXY_1Y_2}$. By standard continuity arguments, $P^{\star}_{UXY_1Y_2}$ must satisfy the Markov chain $U -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1, Y_2)$ and $P^{\star}_{XY_1Y_2} = P_{XY_1Y_2}$. Letting first $k \to \infty$ and then $\epsilon \to 0$, the desired converse result follows by the continuity of mutual information and because μ_{n_k} and $\tilde{\mu}_{n_k}$ vanish as $k \to \infty$.

Appendix B
Proof of Theorem 11

A. Achievability

Pick a real number r, a small positive number $\mu > 0$ satisfying

$$\mathcal{T}_{\mu}^{n}(P_X) \cap \mathcal{T}_{\mu}^{n}(\bar{P}_X) = \emptyset, \tag{82}$$

and a function $\mathsf{b}\colon \{0,1\} \to \{0,1\}$ either as

$$\mathsf{b}(0) = \mathsf{b}(1) = 0 \tag{83}$$

or as

$$\mathsf{b}(0) = 0 \qquad \text{and} \qquad \mathsf{b}(1) = 1. \tag{84}$$

Further, depending on the parameter r and the chosen function b, partition the set of types $\mathcal{P}^n(\mathcal{X})$ into the sets Γ_0 and Γ_1 as follows. All types π satisfying

$$|\pi - P_X| \leq \mu \tag{85}$$

are assigned to Γ_0 and all types π satisfying

$$|\pi - \bar{P}_X| \leq \mu \tag{86}$$

are assigned to Γ_1. If the function b was chosen as in (83), then any other type π that neither satisfies (85) nor (86) is assigned to Γ_1. If the function b was chosen as in (84), then a type π that neither satisfies (85) nor (86) is assigned to Γ_1 if

$$\min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X = \pi \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\left(\tilde{P}_{XY_1} \| P_{XY_1}\right) + r \geq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X = \pi \\ \tilde{P}_{Y_1} = P_{Y_1} \\ \tilde{P}_{Y_2} = P_{Y_2}}} D\left(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\right), \tag{87}$$

and it is assigned to the set Γ_0 otherwise.

<u>Sensor:</u> Given that it observes $X^n = x^n$, it sends

$$M_1 = \begin{cases} 0 & \text{if } \operatorname{tp}(x^n) \in \Gamma_0 \\ 1 & \text{if } \operatorname{tp}(x^n) \in \Gamma_1. \end{cases} \tag{88}$$

<u>Detector 1:</u> Given that it observes $Y_1^n = y_1^n$ and $M_1 = m_1$, it decides

$$\hat{\mathcal{H}}_1 = \begin{cases} 1 & \text{if } m_1 = \mathsf{b}(1) \quad \text{and} \quad y_1^n \in \mathcal{T}_{\mu}^{n}(\bar{P}_{Y_1}) \\ 0 & \text{otherwise.} \end{cases} \tag{89}$$

It sends

$$M_2 = \begin{cases} 0 & \text{if } m_1 = \mathsf{b}(0) \quad \text{and} \quad y_1^n \in \mathcal{T}_{\mu}^{n}(P_{Y_1}) \\ 1 & \text{otherwise} \end{cases} \tag{90}$$

to Detector 2.

<u>Detector 2:</u> Given that it observes $Y_2^n = y_2^n$ and messages $M_1 = m_1$ and $M_2 = m_2$, it decides

$$\hat{\mathcal{H}}_2 = \begin{cases} 0 & \text{if } m_1 = \mathsf{b}(0) \text{ and } m_2 = 0 \text{ and } y_2^n \in \mathcal{T}_{\mu}^{n}(P_{Y_2}) \\ 1 & \text{otherwise.} \end{cases} \tag{91}$$

The described scheme achieves the optimal error-exponents region. The analysis of the described scheme follows from Sanov's theorem, and by noting that $\hat{\mathcal{H}}_1 = 1$ if, and only if,

$$\operatorname{tp}(x^n) \in \Gamma_{\mathsf{b}(1)}^{\mathsf{b},r} \text{ and } y_1^n \in \mathcal{T}_{\mu}^{n}(P_{Y_1}) \tag{92}$$

whereas $\hat{\mathcal{H}}_2 = 0$, if, and only if,

$$\operatorname{tp}(x^n) \in \Gamma_{\mathsf{b}(0)}^{\mathsf{b},r} \text{ and } y_1^n \in \mathcal{T}_{\mu}^{n}(P_{Y_1}) \text{ and } y_2^n \in \mathcal{T}_{\mu}^{n}(P_{Y_2}). \tag{93}$$

The result then follows by letting $n \to \infty$ and $\mu \to 0$.

B. Converse

This converse is inspired by [27]. Fix a real number r, $(\epsilon_1, \epsilon_2) \in (0,1)^2$ and an exponent pair $(\theta_1, \theta_2) \in \mathcal{E}_0(2, 2, \epsilon_1, \epsilon_2)$ satisfying

$$\theta_2 = \theta_1 + r. \tag{94}$$

Then fix a small number $\epsilon > 0$, a sufficiently large blocklength n, and encoding and decision functions $\phi_{1,n}, \phi_{2,n}, \psi_{1,n}, \psi_{2,n}$ satisfying

$$\alpha_{1,n} \leq \epsilon_1, \tag{95}$$

$$\alpha_{2,n} \leq \epsilon_2, \tag{96}$$

and

$$-\frac{1}{n}\log\beta_{1,n} \geq \theta_1 - \epsilon, \tag{97}$$

$$-\frac{1}{n}\log\beta_{2,n} \geq \theta_2 - \epsilon. \tag{98}$$

For the chosen encoding and decision functions, define for each $m_1 \in \{0,1\}$ and $m_2 \in \{0,1,\ldots,\mathsf{W}_2-1\}$, the subsets

$$\mathcal{C}_{m_1} := \{x^n \in \mathcal{X}^n : \phi_{1,n}(x^n) = m_1\}, \tag{99}$$

$$\mathcal{F}^1_{m_1} := \{y_1^n \in \mathcal{Y}_1^n : \psi_{1,n}(m_1, y_1^n) = 1\}, \tag{100}$$

$$\mathcal{G}_{m_1,m_2} := \{y_1^n \in \mathcal{Y}_1^n : \phi_{2,n}(m_1, y_1^n) = m_2\}, \tag{101}$$

$$\mathcal{F}^2_{m_1,m_2} := \{y_2^n \in \mathcal{Y}_2^n : \psi_{2,n}(m_1, m_2, y_2^n) = 0\}. \tag{102}$$

Notice that the sets $\mathcal{C}_0$ and $\mathcal{C}_1$ partition $\mathcal{X}^n$ and for each $m_1 \in \{0,1\}$ the sets $\mathcal{G}_{m_1,0},\ldots,\mathcal{G}_{m_1,\mathsf{W}_2-1}$ partition $\mathcal{Y}_1^n$. Moreover, the acceptance regions $\mathcal{A}_n^1$ and $\mathcal{A}_n^2$ at Detectors 1 and 2, defined through the relations

$$(X^n, Y_1^n) \in \mathcal{A}_n^1 \iff \hat{\mathcal{H}}_1 = 1, \tag{103}$$

$$(X^n, Y_1^n, Y_2^n) \in \mathcal{A}_n^2 \iff \hat{\mathcal{H}}_2 = 0, \tag{104}$$

can be expressed as

$$\mathcal{A}_n^1 = \mathcal{C}_0 \times \mathcal{F}_0^1 \cup \mathcal{C}_1 \times \mathcal{F}_1^1 \tag{105a}$$

and

$$\mathcal{A}_n^2 = \bigcup_{m_2=0}^{\mathsf{W}_2-1} \mathcal{C}_0 \times \mathcal{G}_{0,m_2} \times \mathcal{F}^2_{0,m_2} \cup \bigcup_{m_2=0}^{\mathsf{W}_2-1} \mathcal{C}_1 \times \mathcal{G}_{1,m_2} \times \mathcal{F}^2_{1,m_2}. \tag{105b}$$

Define now for each $m_1 \in \{0,1\}$ the set

$$\Gamma_{m_1,n} := \left\{\tilde{P}_X \in \mathcal{P}(\mathcal{X}) : \ \tilde{P}_X^{\otimes n}\Big[X^n \in \mathcal{C}_{m_1}\Big] \geq \frac{1-\epsilon}{2}\right\}, \tag{106}$$

and for each pair $(m_1, m_2) \in \{0,1\} \times \{0,\ldots,\mathsf{W}_2-1\}$ the set

$$\Delta_{m_1,m_2,n} := \left\{\tilde{P}_{Y_1} \in \mathcal{P}(\mathcal{Y}_1) : \ \tilde{P}_{Y_1}^{\otimes n}\Big[Y_1^n \in \mathcal{G}_{m_1,m_2}\Big] \geq \frac{1-\epsilon}{2\mathsf{W}_2}\right\}. \tag{107}$$

Since the sets $\mathcal{C}_0$ and $\mathcal{C}_1$ cover $\mathcal{X}^n$ and since for each $\tilde{P}_X \in \mathcal{P}(\mathcal{X})$, it holds that $\tilde{P}_X^{\otimes n}\Big[X^n \in \mathcal{X}^n\Big] = 1$, the subsets $\Gamma_{0,n}$ and $\Gamma_{1,n}$ cover the set $\mathcal{P}(\mathcal{X})$. Similarly, since for each $m_1 \in \{0,1\}$ the sets $\mathcal{G}_{m_1,0},\ldots,\mathcal{G}_{m_1,\mathsf{W}_2-1}$ cover $\mathcal{Y}_1^n$, the subsets $\Delta_{m_1,0,n},\ldots,\Delta_{m_1,\mathsf{W}_2-1,n}$ cover the set $\mathcal{P}(\mathcal{Y}_1)$. Moreover, by the constraints on the Type-I error probability at Detectors 1 and 2, (95) and (96):

$$\bar{P}_{XY_1}^{\otimes n}\left[\Big(X^n, Y_1^n\Big) \in \bigcup_{m_1=0}^{1} \mathcal{C}_{m_1} \times \mathcal{F}^1_{m_1}\right] \geq 1 - \epsilon_1 \tag{108}$$

and

$$P_{XY_1Y_2}^{\otimes n}\left[\Big(X^n, Y_1^n, Y_2^n\Big) \in \bigcup_{m_1=0}^{1}\bigcup_{m_2=0}^{\mathsf{W}_2-1} \mathcal{C}_{m_1} \times \mathcal{G}_{m_1,m_2} \times \mathcal{F}^2_{m_1,m_2}\right] \geq 1 - \epsilon_2. \tag{109}$$

By the union bound, there exist thus an index $\tilde{m}_1 \in \{0,1\}$ and an index pair $(m_1^*, m_2^*) \in \{0,1\} \times \{0,\ldots,\mathsf{W}_2-1\}$ such that:

$$\bar{P}_X^{\otimes n}\Big[X^n \in \mathcal{C}_{\tilde{m}_1}\Big] \geq \frac{1-\epsilon_1}{2}, \tag{110a}$$

$$\bar{P}_{Y_1}^{\otimes n}\Big[Y_1^n \in \mathcal{F}^1_{\tilde{m}_1}\Big] \geq \frac{1-\epsilon_1}{2}, \tag{110b}$$

and

$$P_X^{\otimes n}\Big[X^n \in \mathcal{C}_{m_1^*}\Big] \geq \frac{1-\epsilon_2}{2}, \tag{111a}$$

$$P_{Y_1}^{\otimes n}\Big[Y_1^n \in \mathcal{G}_{m_1^*,m_2^*}\Big] \geq \frac{1-\epsilon_2}{2\mathsf{W}_2}, \tag{111b}$$

$$P_{Y_2}^{\otimes n}\Big[Y_2^n \in \mathcal{F}^2_{m_1^*,m_2^*}\Big] \geq \frac{1-\epsilon_2}{2\mathsf{W}_2}, \tag{111c}$$

Combining (110) with the definition of $\Gamma_{\tilde{m}_1,n}$ in (106) and with [27, Theorem 3] (recall that by assumption $P_{XY_1}(x,y_1) > 0$, for all $(x,y_1) \in \mathcal{X} \times \mathcal{Y}_1$) yields that for sufficiently large n :

$$\Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \geq \max_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_{\tilde{m}_1,n}, \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} e^{-n(D(\tilde{P}_{XY_1} \| P_{XY_1}) + \nu_{1,n})},$$

where $\nu_{1,n} \to 0$ when $n \to \infty$. In the same way, combining (111) with (106) and (107), and extending [27, Theorem 3] to three pmfs (recall that by assumption $P_{XY_1Y_2}(x,y_1,y_2) > 0$, for all $(x,y_1,y_2) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$), for sufficiently large n:

$$\Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1] \geq \max_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_{m_1^*,n}, \\ \tilde{P}_{Y_1} \in \Delta_{m_1^*,m_2^*,n},\ \tilde{P}_{Y_2} = P_{Y_2}}} e^{-n(D(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}) + \nu_{2,n})},$$

where $\nu_{2,n} \to 0$ when $n \to \infty$. Taking $n \to \infty$, by the continuity of the KL divergence, we can conclude that if the exponent pair (θ_1, θ_2) is achievable, then there exist subsets Γ_0 and Γ_1 that cover $\mathcal{P}(\mathcal{X})$, subsets $\Delta_{0,0},\ldots,\Delta_{0,\mathsf{W}_2-1}$ that cover $\mathcal{P}(\mathcal{Y}_1)$, and subsets $\Delta_{1,0},\ldots,\Delta_{1,\mathsf{W}_2-1}$ that cover $\mathcal{P}(\mathcal{Y}_1)$ so that:

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_a, \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big), \tag{112a}$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_c, \\ \tilde{P}_{Y_1} \in \Delta_{c,c_2},\ \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big). \tag{112b}$$

where the indices $a, c \in \{0,1\}$ and $c_2 \in \{0,\ldots,\mathsf{W}_2-1\}$ are such that

$$\bar{P}_X \in \Gamma_a, \tag{113}$$

$$P_X \in \Gamma_c, \tag{114}$$

$$P_{Y_1} \in \Delta_{c,c_2}. \tag{115}$$

We continue to notice that the upper bounds in (112) become looser when elements are removed from the sets Γ_a, Γ_c, and Δ_{c,c_2}. The converse statement thus remains valid by imposing

$$\Delta_{c,c_2} = \{P_{Y_1}\}. \tag{116}$$

If $a = c$, we impose

$$\Gamma_a = \Gamma_c = \{P_X, \bar{P}_X\}. \tag{117}$$

This concludes the proof for the function choice $\mathsf{b}(0) = \mathsf{b}(1) = 0$, because we can associate $\Gamma_a = \Gamma_c$ with $\Gamma_0^{\mathsf{b},r}$ in the theorem.

If $a \neq c$, we impose that Γ_a and Γ_c form a partition and obtain the intermediate result that

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_a, \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big) \tag{118a}$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_c, \\ \tilde{P}_{Y_1} = P_{Y_1},\, \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big) \tag{118b}$$

for two sets Γ_a and Γ_c forming a partition of $\mathcal{P}(\mathcal{X})$ and satisfying (113) and (114).

We now characterize the choice of the sets $\{\Gamma_a, \Gamma_c\}$ that yields the loosest bound in (118). To this end, notice first that by assumption (94), constraints (118) are equivalent to:

$$\theta_1 \leq \min\Bigg\{ \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_a, \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big), \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_c, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big) - r \Bigg\}. \tag{119}$$

Moreover,

$$\begin{aligned}
&\min\Bigg\{ \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_a \backslash \{\bar{P}_X\}, \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big), \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_c \backslash \{P_X\}, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big) - r \Bigg\} \\
&\quad \leq \min_{\pi_X \in \mathcal{P}(\mathcal{X}) \backslash \{P_X, \bar{P}_X\}} \max \Bigg\{ \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X = \pi_X, \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big), \\
&\qquad\qquad \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X = \pi_X, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big) - r \Bigg\},
\end{aligned} \tag{120}$$

and the inequality holds with equality when any type $\pi_X \in \mathcal{P}(\mathcal{X}) \backslash \{P_X, \bar{P}_X\}$ satisfies

$$\begin{gathered}
(\pi_X \in \Gamma_a) \\
\Longleftrightarrow \\
\min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X = \pi_X, \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\Big(\tilde{P}_{XY_1} \| P_{XY_1}\Big) \geq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X = \pi_X, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D\Big(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}\Big) - r.
\end{gathered} \tag{121}$$

If $a \neq c$, one can thus conclude that the bound in (119) holds for sets Γ_a and Γ_c that partition $\mathcal{P}(\mathcal{X})$ and that satisfy $P_X \in \Gamma_c$ and $\bar{P}_X \in \Gamma_a$, and for any other pmf in $\mathcal{P}(\mathcal{X}) \backslash \{P_X, \bar{P}_X\}$ above condition (121) is satisfied. This concludes the proof for the case $a \neq c$, for the choice of the function $\mathsf{b}(0) = 0$ and $\mathsf{b}(1) = 1$ by associating Γ_a with $\Gamma_1^{\mathsf{b},r}$ and Γ_c with $\Gamma_0^{\mathsf{b},r}$.

References

[1] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Transaction on Information Theory*, vol. 32, no. 4, pp. 533–542, July 1986.

[2] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Transaction on Information Theory*, vol. 33, no. 6, pp. 759–772, November 1987.

[3] H. Shimokawa, T. S. Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *Proc. 1994 IEEE International Symposium on Information Theory (ISIT)*, Trondheim, Norway, June 1994, p. 114.

[4] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Transaction on Information 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. 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, Illinois, USA, October 2014, pp. 1246–1251.

[6] A. Zaidi and I. E. Aguerri, "Optimal rate-exponent region for a class of hypothesis testing against conditional independence problems," in *Proc. 2019 IEEE Information Theory Workshop (ITW)*, Visby, Sweden, August 2019, pp. 1–5.

[7] A. Zaidi, "Hypothesis testing against independence under Gaussian noise," in *Proc. 2020 IEEE International Symposium on Information Theory (ISIT)*, Los Angeles (CA), USA, June 2020, pp. 1–5.

[8] Y. Uğur, I. E. Aguerri, and A. Zaidi, "Vector Gaussian CEO problem under logarithmic loss and applications," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4183–4202, July 2020.

[9] S. Salehkalaibar, M. Wigger, and R. Timo, "On hypothesis testing against conditional independence with multiple decision centers," *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2409–2420, June 2018.

[10] C. Tian and J. Chen, "Successive refinement for hypothesis testing and lossless one-helper problem," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4666–4681, October 2008.

[11] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing with communication constraints," in *Proc. 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, Illinois, USA, October 2012, pp. 1065–1072.

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

[13] W. Zhao and L. Lai, "Distributed testing with zero-rate compression," in *Proc. 2015 IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, June 2015, pp. 2792–2796.

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

[15] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4411–4433, July 2019.

[16] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detections," in *Proc. 2018 IEEE International Symposium on Information Theory (ISIT)*, Vail (CO), USA, June 2018, pp. 166–170.

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

[18] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4150–4182, July 2020.

[19] J. Liao, L. Sankar, F. du Pin Calmon, and V. Y. Tan, "Hypothesis testing under maximal leakage privacy constraints," in *Proc. 2017 IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, 2017, pp. 779–783.

[20] J. Liao, L. Sankar, V. Y. F. Tan, and F. du Pin Calmon, "Hypothesis testing under mutual information privacy constraints in the high privacy regime," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 1058–1071, April 2018.

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

[22] A. Gilani, S. Belhadj Amor, S. Salehkalaibar, and V. Y. F. Tan, "Distributed hypothesis testing with privacy constraints," *Entropy*, vol. 21, no. 5, 2019.

[23] W. Zhao and L. Lai, "Distributed testing with cascaded encoders," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7339–7348, November 2018.

[24] S. Salehkalaibar, M. Wigger, and R. Timo, "On hypothesis testing against independence with multiple decision centers," *IEEE Transactions on Communications*, vol. 66, no. 7, pp. 2409–2420, July 2018.

[25] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over multi-access channels," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, December 2018, pp. 1–6.

[26] P. Escamilla, A. Zaidi, and M. Wigger, "Some results on the vector Gaussian hypothesis testing problem," in *Proc. 2020 IEEE International Symposium on Information Theory (ISIT)*, Los Angeles (CA), USA, June 2020, pp. 1–5.

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

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

[29] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing," *CoRR*, vol. abs/1604.01292, 2016. [Online]. Available: http://arxiv.org/abs/1604.01292

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

[31] T. Cover and J. Thomas, *Elements of Information Theory 2nd Edition*. Wiley-Interscience, 2006.

[32] Y. Uğur, I. E. Aguerri, and A. Zaidi, "A generalization of Blahut-Arimoto algorithm to compute rate-distortion regions of multiterminal source coding under logarithmic loss," in *Proc. 2017 IEEE Information Theory Workshop (ITW)*, Kaohsiung, Taiwan, 2017, pp. 349–353.

[33] M. Ueda and S. Kuzuoka, "The error exponent of zero-rate multiterminal hypothesis testing for sources with common information," in *Proc. 2014 IEEE International Symposium on Information Theory and its Applications*, Melbourne, Australia, October 2014, pp. 546–550.

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

Pierre Escamilla received the State Engineering degrees and the M. Sc. from Supelec, Paris, France in 2010, and the Ph.D. degree in networks, information and communication from Institut Polytechnique de Paris, in 2019. From 2011 to 2016, he worked as a Research Engineer at Safran Identity and Security, on face recognition. Since 2019, he has been a Consultant Researcher at the Mathematics and Algorithmic Sciences Lab., Paris Research Center, Huawei France. His research interests range from information theory to distributed inference and machine learning.

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 Telecom Paris, France, where she is currently a full professor. Dr. Wigger has held visiting professor appointments at the Technion–Israel Institute of Technology and ETH Zurich. Dr. Wigger has previously served as an Associate Editor of the IEEE Communication Letters and 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. Dr. Wigger's research interests are in multi-terminal information theory.

Abdellatif Zaidi received the B.S. degree in Electrical Engineering from ENSTA ParisTech, Paris, in 2002 and the M. Sc. and Ph.D. degrees in Electrical Engineering from TELECOM ParisTech, Paris, France in 2002 and 2006, respectively. From December 2002 to March 2006, he was with the Communications an Electronics Dept., TELECOM ParisTech, Paris, France and the Signals and Sys- tems Lab., CNRS/Suplec, France pursuing his PhD degree. From May 2006 to September 2010, he was at Ecole Polytechnique de Louvain, Universite Catholique de Louvain, Belgium, working as a senior researcher. Dr. Zaidi was Research Visitor at the University of Notre Dame, Indiana, USA, during 2007 and 2008, the Technical University of Munich during Summer 2014, and the Ecole Polytechnique Federale de Lausanne, EFPL, Switzerland. He is currently Associate Professor at Universite Paris-Est, France; and with the Mathematics and Algorithmic Sciences Lab., Paris Research Center, Huawei France. His research interests lie broadly in network information theory and its interactions with other fields, including communication and coding, statistics, security and privacy, as well as learning, with application to diverse problems of data transmission and compression in networks. Dr. Zaidi is IEEE senior member. From 2013 to 2016 he served as Associate Editor for the Eurasip Journal on Wireless Communications and Networking (EURASIP JWCN); and, since 2016, as Associate Editor for the IEEE Transactions on Wireless Communications. He is the recipient of French Excellence in Research Award (in 2011); and co-recipient (with Shlomo Shamai (Shitz)) of the Best Paper Award (in 2014).