Distributed Hypothesis Testing Over Noisy Broadcast Channels

Sadaf Salehkalaibar, *IEEE Member* and Michèle Wigger, *IEEE Senior Member*

Abstract

This paper studies binary hypothesis testing with a single sensor that communicates with two decision centers over a memoryless broadcast channel. The main focus lies on the tradeoff between the two type-II error exponents achievable at the two decision centers. In our proposed scheme, we can partially mitigate this tradeoff when the transmitter has a probability larger than $1/2$ to distinguish the alternate hypotheses at the decision centers, i.e., the hypotheses under which the decision centers wish to maximize their error exponents. In the case where these hypotheses cannot be distinguished at the transmitter (because both decision centers have same alternative hypothesis or because the transmitter's observations have same marginal distribution under both hypotheses) our scheme shows an important tradeoff between the two exponents. The results in this paper thus reinforce the previous conclusions drawn for a setup where communication is over a common noiseless link. Compared to such a noiseless scenario, here however we observe that even when the transmitter can distinguish the two hypotheses, a small exponent-tradeoff can persist, simply because the noise in the channel prevents the transmitter to perfectly describe its guess of the hypothesis to the two decision centers.

I. Introduction

In Internet of Things (IoT) networks, data is collected at sensors and transmitted over a wireless channel to remote decision centers, which decide on one or multiple hypotheses based on the collected information. In this paper, we study simple binary hypothesis testing with a single sensor but two decision centers. The results can be combined with previous studies focusing on multiple sensors and a single decision center to tackle the practically relevant case of multiple sensors and multiple decision centers. In this paper, we consider a single sensor for simplicity and because our main focus is on studying the tradeoff between the performances at the two decision centers. A main theme in our work is that the single sensor has to send information over the channel that can be used by both decision centers. A simple, but highly suboptimal approach would be to time-share communication and serve each of the two decision center only during a part of the transmission. As we will see, better schemes are possible, and in some cases it is even possible to serve each of the two decision centers as if the other center was not present in the system.

In this paper, we follow the information-theoretic framework introduced by [1], [2]. That means, each terminal observes a memoryless sequence and depending on the underlying hypothesis $\mathcal{H} \in \{0, 1\}$. All sequences follow one of two possible joint distributions, which are known to all involved terminals. A priori, the transmitter however ignores the correct hypothesis and has to compute its transmit signal as a function of the observed source symbols only. Decision centers observe outputs of the channel, and combined with their local observations they have to make a decision on whether $\mathcal{H} = 0$ or $\mathcal{H} = 1$. The performance of the decision center is measured by its *type-II error error exponent*, i.e., the expontial decay in the length of the observations of the probability of deciding on $\mathcal{H} = 0$ when the true hypothesis is $\mathcal{H} = 1$. As a constraint on the decision center, we impose that the *type-I error probability*, i.e., the probability of deciding $\mathcal{H} = 1$ when the true hypothesis is $\mathcal{H} = 0$, vanishes (at any desired speed) with increasing observation lengths. The motivation for studying such asymmetric requirements on the two error probabilities stems for example from alert systems, where the *miss-detection* event is much more harmful than the *false-alarm* event, and as a consequence in our systems we require the miss-detection probability to decay much faster than the false-alarm probability.

This problem statement has first been considered for the setup with a single sensor and a single decision center when communication is over a noiseless link of given capacity [1], [2]. For this canonical problem, the optimal error exponent has been identified in the special cases of *testing against independence* [1] and *testing against conditional independence* [3], [4]. The scheme proposed by Shimokawa-Han-Amari in [3], [4] yields an achievable error exponent for all distributed hypothesis testing problems (not only testing against conditional independence) [3], [4], but it might not be optimal in general [5]. The Shimokawa-Han-Amari (SHA) scheme has been extended to various more involved setups such as noiseless networks with multiple sensors and a single decision center [2], [6], [7]; networks where the sensor and the decision center can communicate interactively [8], [9]; multi-hop networks [10], and networks with multiple decision centers [10]–[13].

The works most closely related to the current paper are [10], [12], [13] and [14], [15]. Specifically, [10], [12], [13] consider a single-sensor multi-detector system where communication is over a common noiseless link from the sensor to all decision centers. Focusing on two decision centers, two scenarios can encountered here: 1) the two decision centers have same null and

S. Salehkalaibar is with the Department of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran, s.saleh@ut.ac.ir.
M. Wigger is with LTCI, Telecom Paris, IP Paris, 91120 Palaiseau, Paris, France, michele.wigger@telecom-paristech.fr.
M. Wigger has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 under grant agreement No 715111.

alternate hypotheses and as consequence both aim at maximizing the error exponent under the same hypothesis $\mathcal{H}$; or 2) the two decision centers have opposite null and alternate hypotheses and thus one decision center wishes to maximize the error exponent under hypothesis $\mathcal{H} = 0$ and the other under hypothesis $\mathcal{H} = 1$. The second scenario is motivated by applications where the decision centers have different goals. Hypothesis testing for scenario 1) was studied in [10], [12], [13] and the results showed a tradeoff between the exponents achieved at the two decision centers. Intuitively, the tradeoff comes from the fact that communication from the sensor is serving both decision centers at the same time. Scenario 2) was considered in [12], [13]. In this case, a tradeoff only occurs when the sensor's observation alone provides no advantage in guessing the hypothesis. Otherwise, a tradeoff-free exponents region can be achieved by the following simple scheme: The sensor takes a tentative guess on the hypothesis based only on its local observations. It communicates this tentative guess to both decision centers using a single bit and then dedicates the rest of the communication using a dedicated SHA scheme only to the decision center that wishes to maximize the error exponent under the hypothesis that does *not* correspond to its guess. The other decision center simply keeps the transmitter's tentative guess and ignores the rest of the communication.

In this paper, we extend these previous works to memoryless broadcast channels (BC). Hypothesis testing over BCs was already considered in [10], however only for above scenario 1 and in the special case of testing against conditional independence, in which case the derived error exponents were proved to be optimal. Interestingly, when testing against conditional independence over noisy channels, only the capacity of the channel matters but not the properties, see [14], [15]. General hypothesis testing over noisy channels is much more challenging and requires additional tools, such as joint source-channel coding and *unequal error protection (UEP)* coding [16]. The latter can in particular be used to specially protect the communication of the sensor's tentative guess, which allows to avoid a degradation of the performance of classical hypothesis testing schemes.

We present general distributed hypothesis testing schemes over memoryless BCs, and we analyze the performances of these schemes with a special focus on the tradeoff in exponents they achieve for the two decision centers. We propose two different schemes, depending on whether the sensor can distinguish with error probability $\neq 1/2$ the two null hypotheses at the two decision centers. If a distinction is possible (because the decision centers have different null hypothesis and the sensor's observations follow different marginal distributions under the two hypotheses), then we employ a similar scheme as proposed in [12], [13] over a common noiseless link, but where the SHA scheme is replaced by the UEP-based scheme for DMCs in [14]. That means, the sensor makes a tentative guess about the hypothesis and conveys this guess to both decision centers using an UEP mechanism. Moreover, the joint source-channel coding scheme in [14] with dedicated codebooks is used to communicate to the decision center that aims to maximize the error exponent under the hypothesis that does *not* correspond to the sensor's tentative guess. This scheme shows no tradeoff between the exponents achieved at the two decision centers in various interesting cases. Sometimes however a tradeoff arises because even under UEP the specially protected messages can be in error and because the decision centers can confuse the codwords of the two different sets of codebooks. For the case where the sensor cannot reasonably distinguish the alternate hypotheses at the two decision centers (because both decision centers have same alterate hypotheses or the sensor's observations have same marginal observations under both hypotheses) we present a scheme similar to [10] but again including UEP. In this scheme, a tradeoff between the exponents achieved at the two decision centers naturally arises and mostly stems from the inherent tradeoff in distributed lossy compression systems with multiple decoders having different side-informations.

A. Notation

We mostly follow the notation in [17]. Random variables are denoted by capital letters, e.g., X, Y, and their realizations by lower-case letters, e.g., x, y. Script symbols such as $\mathcal{X}$ and $\mathcal{Y}$ stand for alphabets of random variables, and $\mathcal{X}^n$ and $\mathcal{Y}^n$ for the corresponding n-fold Cartesian products. Sequences of random variables $(X_i, ..., X_j)$ and realizations $(x_i, \dots, x_j)$ are abbreviated by X_i^j and x_i^j. When $i = 1$, then we also use X^j and x^j instead of X_1^j and x_1^j.

We write the probability mass function (pmf) of a discrete random variable X as P_X; to indicate the pmf under hypothesis $\mathcal{H} = 1$, we also use Q_X. The conditional pmf of X given Y is written as $P_{X|Y}$, or as $Q_{X|Y}$ when $\mathcal{H} = 1$. The term $D(P\|Q)$ stands for the Kullback-Leibler (KL) divergence between two pmfs P and Q over the same alphabet. We use $\mathrm{tp}(a^n, b^n)$ to denote the *joint type* of the pair of sequences (a^n, b^n), and $\mathrm{cond\_tp}(a^n|b^n)$ for the conditional type of a^n given b^n. For a joint type π_{ABC} over alphabet $\mathcal{A} \times \mathcal{B} \times \mathcal{C}$, we denote by $I_{\pi_{ABC}}(A; B|C)$ the mutual information assuming that the random triple (A, B, C) has pmf π_{ABC}; similarly for the entropy $H_{\pi_{ABC}}(A)$ and the conditional entropy $H_{\pi_{ABC}}(A|B)$. Sometimes we abbreviate π_{ABC} by π. Also, when π_{ABC} has been defined and is clear from the context, we write π_A or π_{AB} for the corresponding subtypes. When the type π_{ABC} coincides with the actual pmf of a triple (A, B, C), we omit the subscript and simply write $H(A)$, $H(A|B)$, and $I(A; B|C)$.

For a given P_X and a constant $\mu > 0$, let $\mathcal{T}_\mu^n(P_X)$ be the set of *μ-typical sequences* in $\mathcal{X}^n$ as defined in [8, Sec. 2.4]. Similarly, $\mathcal{T}_\mu^n(P_{XY})$ stands for the set of *jointly μ-typical sequences*. The expectation operator is written as $\mathbb{E}[\cdot]$. We abbreviate *independent and identically distributed* by *i.i.d.*. The $\log$ function is taken with base 2. Finally, in our justifications, we use (DP) and (CR) for "data processing inequality" and "chain rule".

II. System Model

Consider the distributed hypothesis testing problem in Fig. 1, where a transmitter observes sequence X^n, Receiver 1 sequence Y_1^n, and Receiver 2 sequence Y_2^n. Under the null hypothesis:

$$\mathcal{H}=0\colon (X^n,Y_1^n,Y_2^n) \quad \text{i.i.d. } \sim P_{XY_1Y_2}, \tag{1}$$

and under the alternative hypothesis:

$$\mathcal{H}=1\colon (X^n,Y_1^n,Y_2^n) \quad \text{i.i.d. } \sim Q_{XY_1Y_2}, \tag{2}$$

for two given pmfs $P_{XY_1Y_2}$ and $Q_{XY_1Y_2}$. The transmitter can communicate with the receivers over n uses of a discrete memoryless broadcast channel $(\mathcal{W},\mathcal{V}_1\times\mathcal{V}_2,P_{V_1V_2|W})$ where $\mathcal{W}$ denotes the finite channel input alphabet and $\mathcal{V}_1$ and $\mathcal{V}_2$ the finite channel output alphabets. Specifically, the transmitter feeds inputs

$$W^n=f^{(n)}(X^n), \tag{3}$$

to the channel, where $f^{(n)}$ denotes the chosen (possibly stochastic) encoding function

$$f^{(n)}:\mathcal{X}^n\to\mathcal{W}^n. \tag{4}$$

Each Receiver $i\in\{1,2\}$ observes the BC ouputs V_i^n, where for a given input $W_t=w_t$,

$$(V_{1,t},V_{2,t})\sim\Gamma_{V_1V_2|W}(\cdot,\cdot|w_t), \qquad t\in\{1,\dots,n\}. \tag{5}$$

Based on the sequence of channel outputs V_i^n and the source sequence Y_i^n, Receiver i decides on the hypothesis $\mathcal{H}$. That means, it produces the guess

$$\hat{\mathcal{H}}_i=g^{(n)}(V_i^n,Y_i^n), \tag{6}$$

for a chosen decoding function

$$g_i^{(n)}\colon \mathcal{V}_i^n\times\mathcal{Y}_i^n\to\{0,1\}. \tag{7}$$

There are different possible scenarios regarding the requirements on error probabilities. We assume that each receiver is interested in only one of the two exponents. For each $i\in\{1,2\}$, let $\mathsf{h}_i\in\{0,1\}$ be the hypothesis whose error exponent Receiver i wishes to maximize, and $\bar{\mathsf{h}}_i$ the other hypothesis, i.e., $\bar{\mathsf{h}}_i\in\{0,1\}$ and $\mathsf{h}_i\neq\bar{\mathsf{h}}_i$. (The values of h_1 and h_2 are fixed and part of the problem statement.) We then have:

Definition 1: An exponent pair (θ_1,θ_2) is said achievable over a BC, if for each $\epsilon_1,\epsilon_2\in(0,1)$ and sufficiently large blocklengths n, there exist encoding and decoding functions $(f^{(n)},g_1^{(n)},g_2^{(n)})$ such that:

$$\alpha_{1,n}\triangleq\Pr[\hat{\mathcal{H}}_1=\mathsf{h}_1|\mathcal{H}=\bar{\mathsf{h}}_1], \qquad \alpha_{2,n}\triangleq\Pr[\hat{\mathcal{H}}_2=\mathsf{h}_2|\mathcal{H}=\bar{\mathsf{h}}_2], \tag{8}$$

$$\beta_{1,n}\triangleq\Pr[\hat{\mathcal{H}}_1=\bar{\mathsf{h}}_1|\mathcal{H}=\mathsf{h}_1], \qquad \beta_{2,n}\triangleq\Pr[\hat{\mathcal{H}}_2=\bar{\mathsf{h}}_2|\mathcal{H}=\mathsf{h}_2], \tag{9}$$

satisfy

$$\alpha_{i,n}\le\epsilon_i, \qquad i\in\{1,2\}, \tag{10}$$

and

$$-\varlimsup_{n\to\infty}\frac{1}{n}\log\beta_{i,n}\ge\theta_i, \qquad i\in\{1,2\}. \tag{11}$$

Definition 2: The *fundamental exponents region* $\mathcal{E}$ is the set of all exponent pairs (θ_1,θ_2) that are achievable.

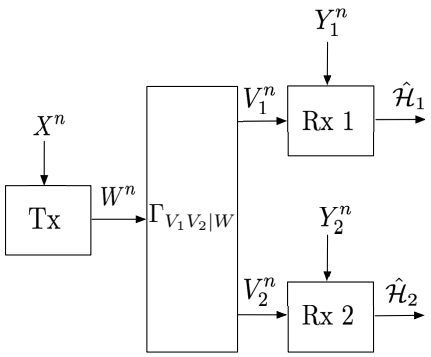

Fig. 1. Hypothesis testing over a noisy BC.

Remark 1: Notice that both $\alpha_{1,n}$ and $\beta_{1,n}$ depend on the BC law $\Gamma_{V_1V_2|W}$ only through the conditional marginal distribution $\Gamma_{V_1|W}$. Similarly, $\alpha_{2,n}$ and $\beta_{2,n}$ only depend on $\Gamma_{V_2|W}$. As a consequence, also the fundamental exponents region $\mathcal{E}$ depends on the joint laws $P_{XY_1Y_2}$ and $Q_{XY_1Y_2}$ only through their marginal laws P_{XY_1}, P_{XY_2}, Q_{XY_1}, and Q_{XY_2}.

Remark 2: As a consequence to the preceding Remark 1, when $P_X = Q_X$ one can restrict attention to a scenario where both receivers aim at maximizing the error exponent under hypothesis $\mathcal{H} = 1$, i.e., $\mathsf{h}_1 = \mathsf{h}_2 = 1$. In fact, under $P_X = Q_X$ the fundamental exponents region $\mathcal{E}$ for arbitrary h_1 and h_2 coincides with the fundamental exponents region $\mathcal{E}$ for $\mathsf{h}_1' = 1$ and $\mathsf{h}_2' = 1$ if one exchanges pmfs P_{XY_1} and Q_{XY_1} in case $\mathsf{h}_1 = 0$ and one exchanges pmfs P_{XY_2} and Q_{XY_2} in case $\mathsf{h}_2 = 0$.

To simplify notation in the sequel, we use the following shorthand notations for the pmfs $P_{XY_1Y_2}$ and $Q_{XY_1Y_2}$. For each $i \in \{1, 2\}$:

$$\text{if } \bar{\mathsf{h}}_i = 0 \quad \Longrightarrow \quad \big(p^i_{XY_1Y_2} := P_{XY_1Y_2} \quad \text{and} \quad q^i_{XY_1Y_2} := Q_{XY_1Y_2}\big) \tag{12a}$$

and

$$\text{if } \bar{\mathsf{h}}_i = 1 \quad \Longrightarrow \quad \big(p^i_{XY_1Y_2} := Q_{XY_1Y_2} \quad \text{and} \quad q^i_{XY_1Y_2} := P_{XY_1Y_2}\big)\,. \tag{12b}$$

We propose two coding schemes yielding two different exponent regions, depending on the whether

$$\forall x \in \mathcal{X}\colon\ p^1_X(x) = p^2_X(x), \tag{13}$$

or

$$\exists x \in \mathcal{X}\colon\ p^1_X(x) \neq p^2_X(x). \tag{14}$$

Notice that (13) always holds when $\mathsf{h}_1 = \mathsf{h}_2$. In contrast, given (14), then obviously $\mathsf{h}_1 \neq \mathsf{h}_2$.

III. Results on Exponents Region

Before presenting our main results, we recall the achievable error exponent over a discrete memoryless channel reported in [14, Theorem 1].

A. Achievable Exponent for Point-to-Point Channels

Consider a single-receiver setup with only Receiver 1 that wishes to maximize the error exponent under hypothesis $\mathsf{h}_1 = 1$. For simplicity then, we drop the user index 1 and simply call the receiver's source observation Y^n and its channel outputs V^n.

Theorem 1 (Theorem 1 in [14]): Any exponent θ satisfying the following condition is achievable:

$$\theta \leq \max \min\{\theta_{\text{standard}}(P_{S|X}),\ \theta_{\text{dec}}(P_{S|X}, P_T, P_{W|T}),\ \theta_{\text{miss}}(P_{S|X}, P_T, P_{W|T})\}, \tag{15}$$

where the maximization is over pmfs $P_{S|X}$, P_T, and $P_{W|T}$ such that the joint law $P_{SWVXY} := P_{XY}P_{S|X}P_TP_{W|T}P_{V|W}$ satisfies

$$I(S; X|Y) \leq I(W; V|T), \tag{16}$$

and where the exponents in (15) are defined as:

$$\theta_{\text{standard}}(P_{S|X}) := \min_{\substack{\tilde{P}_{SXY}\colon \\ \tilde{P}_{SX} = P_{SX} \\ \tilde{P}_{SY} = P_{SY}}} D(\tilde{P}_{SXY} \| P_{S|X}Q_{XY}), \tag{17}$$

$$\theta_{\text{dec}}(P_{S|X}, P_T, P_{W|T}) := \min_{\substack{\tilde{P}_{SXY}\colon \\ \tilde{P}_{SX} = P_{SX} \\ \tilde{P}_{Y} = P_{Y} \\ H_P(S|Y) \leq H_{\tilde{P}}(S|Y)}} D(\tilde{P}_{SXY} \| P_{S|X}Q_{XY}) - I(S; X|Y) + I(W; V|T), \tag{18}$$

$$\theta_{\text{miss}}(P_{S|X}, P_T, P_{W|T}) := D(P_Y \| Q_Y) + \mathbb{E}_{P_T}\left[D\big(P_{V|T} \| \Gamma_{V|W=T}\big)\right] - I(S; X|Y) + I(W; V|T). \tag{19}$$

Here, all mutual information terms are calculated with respect to the joint pmf P_{SWVXY} defined above.

The exponent in Theorem 1 is obtained by the following scheme, also depicted in Figure 2. The transmitter attempts to quantize the source sequence X^n using a random codebook consisting of codewords $\{S^n(m, \ell)\}$. If the quantization fails because no codeword is jointly typical with the source sequence, then the transmitter applies the UEP mechanism in [16] by sending an IID P_T-sequence T^n over the channel. Otherwise it sends the codeword $W^n(m)$ for m indicating the first index of the $S^n(m, \ell)$-codeword that is jointly typical with its source observation X^n. The receiver jointly decodes the channel and source codeword by verifying the existence of indices (m', ℓ') such that $W^n = W^n(m')$ is jointly typical with its channel outputs V^n and there is no other codeword $S^n(m', \tilde{\ell})$ with smaller conditional empirical entropy given Y^n than $S^n(m', \ell')$.

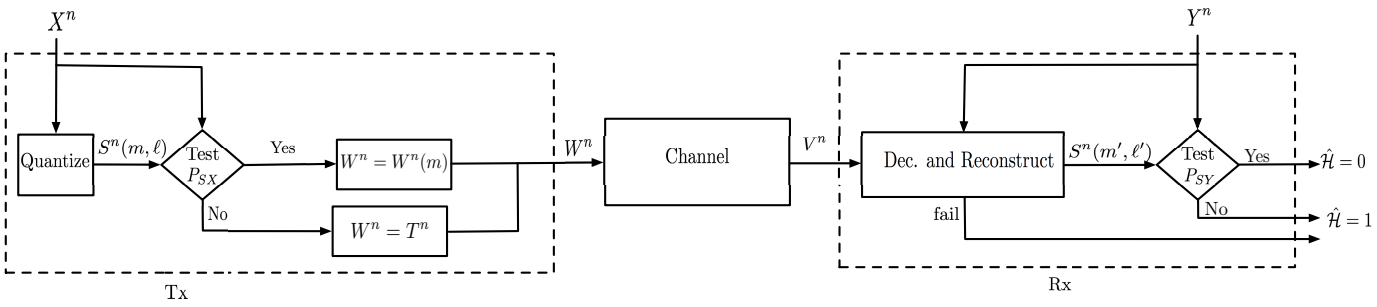

Fig. 2. Coding and testing scheme for hypothesis testing over a DMC.

If the decoded codeword $S^n(m', \ell')$ is jointly typical with the receiver's observation Y^n, then the it produces $\hat{\mathcal{H}} = 0$, and otherwise $\hat{\mathcal{H}} = 1$.

The three competing type-II error exponents in Theorem 1 can be understood in view of this coding scheme as follows. Exponent θ_{standard} indicates the event that a random codeword $S^n(m, \ell)$ is jointly typical with the transmitter's observation X^n and with the receiver's observation Y^n while being under $\mathcal{H} = 1$. This is also the error exponent in Han's scheme [2] over a noise-less communication link and does not depend on the channel law $\Gamma_{V|W}$. Exponent θ_{dec} is related to the joint decoding which checks joint typicality of both the source codeword as well as of the channel codeword and applies a conditional minimum entropy decoder. A similar error exponent is observed in the SHA scheme [3], [4] over a noiseless link if the mutual information $I(W; V|T)$ is replaced by the rate of the link. The third exponent θ_{miss} finally indicates an event where the transmitter sends T^n (so as to indicate the receiver to decide for $\hat{\mathcal{H}} = 1$) but the receiver detects a channel codeword $W^n(m')$ and a corresponding source codeword $S^n(m', \ell')$. This exponent is directly related to the channel transition law $\Gamma_{V|W}$ and not only to the mutual information of the channel and does not occur when transmission is over a noiseless link.

We now explain our achievable exponents region, where we distinguish the two cases 1) $\mathsf{h}_1 \neq \mathsf{h}_2$ and $P_X \neq Q_X$; and 2) $(\mathsf{h}_1 = \mathsf{h}_2)$ or $P_X = Q_X$.

B. Achievable Exponents Region when $\mathsf{h}_1 \neq \mathsf{h}_2$ *and* $P_X \neq Q_X$

Theorem 2: If $\mathsf{h}_1 \neq \mathsf{h}_2$ and $P_X \neq Q_X$, i.e., (14) holds, then all error exponent pairs (θ_1, θ_2) satisfying the following condition are achievable:

$$\theta_i \leq \min\{\theta_{\text{standard},i}\ \theta_{\text{dec},i},\ \theta_{\text{cross},i},\ \theta_{\text{miss},i}\}, \qquad i \in \{1, 2\}, \tag{20}$$

where the union is over pmfs $p^i_{S|X}$, p_T, $p^i_{T_i|T}$, and $p^i_{W|T_i}$, for $i \in \{1, 2\}$, so that the joint pmfs p^1, p^2, q^1, q^2 defined through (12) and

$$p^i_{SXY_1Y_2TT_iWV_1V_2} := p^i_{S|X} \cdot p^i_{XY_1Y_2} \cdot p_T \cdot p^i_{T_i|T} \cdot p^i_{W|TT_i} \cdot \Gamma_{V_1V_2|W}. \tag{21}$$

satisfy constraints

$$I_{p^i}(S; X|Y_i) < I_{p^i}(W; V_i|T, T_i). \tag{22}$$

and where the exponents in (20) are defined as:

$$\theta_{\text{standard},i} := \min_{\substack{\tilde{P}_{SXY_i}: \\ \tilde{P}_{SX}=p^i_{SX} \\ \tilde{P}_{SY_i}=p^i_{SY_i}}} D(\tilde{P}_{SXY_i} \| p^i_{S|X} q^i_{XY_i}), \tag{23}$$

$$\theta_{\text{dec},i} := \min_{\substack{\tilde{P}_{SXY_i}: \\ \tilde{P}_{SX}=p^i_{SX} \\ \tilde{P}_{Y_i}=p^i_{Y_i} \\ H_{p^i}(S|Y_i) \leq H_{\tilde{P}}(S|Y_i)}} D(\tilde{P}_{SXY_i} \| p^i_{S|X} q^i_{XY_i}) - I_{p^i}(S; X|Y_i) + I_{p^i}(W; V_i|T, T_i), \tag{24}$$

$$\theta_{\text{miss},i} := D(p^i_{Y_i} \| q^i_{Y_i}) + \mathbb{E}_{p_T}\left[D\big(p^i_{V_i|T} \| \Gamma_{V_i|W=T}\big)\right] - I_{p^i}(S; X|Y_i) + I_{p^i}(W; V_i|T, T_i), \tag{25}$$

$$\theta_{\text{cross},i} := \min_{\substack{\tilde{P}_{SXY_i}: \\ \tilde{P}_{Y_i}=p^i_{Y_i} \\ H_{p^i}(S|Y_i) \leq H_{\tilde{P}}(S|Y_i)}} \mathbb{E}_{q^i_{XS}}\left[D\left(\tilde{P}_{Y_i|XS} \| q^i_{Y_i|X}\right)\right] - I_{p^i}(S; X|Y_i) + I_{p^i}(W; V_i|T, T_i)$$

$$+ \min_{\substack{\tilde{P}_{TT_iW}: \\ \tilde{P}_{TW}=q^i_{TW} \\ \tilde{P}_{TT_i}=p^i_{TT_i}}} \mathbb{E}_{\tilde{P}_{T_iW}} \left[D(p^i_{V_i|T,T_i} \| \Gamma_{V_i|W}) \right]. \tag{26}$$

Proof: See Section A. ■

In Theorem 2, the exponent triple $\theta_{\text{standard},1}, \theta_{\text{dec},1}, \theta_{\text{miss},1}$ can be optimized over the pmfs $p^1_{S|X}$, $p^1_{T_1|T}$ and $p^1_{W|T_1}$ and independently thereof the exponent triple $\theta_{\text{standard},2}, \theta_{\text{dec},2}, \theta_{\text{miss},2}$ can be optimized over the pmfs $p^2_{S|X}$, $p^2_{T_2|T}$ and $p^2_{W|T,T_2}$. The pmf p^T is common to both optimizations. However, whenever the exponents $\theta_{\text{cross},1}$ and $\theta_{\text{cross},2}$ are not active, Theorem 2 depends only on $p^i_{S|X}$, $p^i_{T_i}$ and $p^i_{W|T_i}$, for $i = 1, 2$, and there is thus no tradeoff between the two exponents θ_1 and θ_2. In other words, the same exponents θ_1 and θ_2 can be attained as in a system where the transmitter communicates over two individual DMCs $\Gamma_{V_1|W}$ and $\Gamma_{V_2|W}$ to the two receivers, or equivalently each receiver achieves the same exponent as if the other receiver was not present in the system.

The scheme achieving the exponents region in Theorem 2 is described in detail in Section IV and analyzed in Appendix A. The main feature is that the sensor makes a tentative decision on $\mathcal{H}$ and conveys this decision to both receivers through its choice of the codebooks and a special coded time-sharing sequence indicating this choice. The receiver that wishes to maximize the error exponent corresponding to the hypothesis guessed at the sensor directly decides on this hypothesis. The other receiver should compare its own observation to a quantized version of the source sequence observed at the sensor. The sensor uses the quantization and binning scheme presented in [14] tailored to this latter receiver using either codebooks $\{S^n(1; m, \ell)\}$ and $\{W^n(1; m)\}$ or codebooks $\{S^n(2; m, \ell)\}$ and $\{W^n(2; m)\}$, respectively. The overall scheme is illustrated in Fig. 3.

Exponents $\theta_{\text{standard},i}$, $\theta_{\text{dec},i}$, and $\theta_{\text{miss},i}$ have similar explanations have in the single-user case. Exponent $\theta_{\text{cross},i}$ corresponds to the event that the transmitter sends a codeword from $\{W(j; m)\}$, for $j = 3 - i$, but Receiver i decides that a codeword from $\{W(i; m)\}$ was sent and a source codeword $S(i; m, \ell)$ satisfies the minimum conditional entropy condition and the typicality check with the observed source sequence Y^n_i. Notice that setting T_i a constant, decreases error exponent $\theta_{\text{cross},i}$.

For the special case where the BC consists of a common noiseless link, Theorem 2 has been proved in [12], [13]. (More precisely, [12] considers the more general case with $K \geq 2$ receivers and $M \geq K$ hypotheses.) In this case, the exponents $(\theta_{\text{miss},1}, \theta_{\text{cross},1})$ and $(\theta_{\text{miss},2}, \theta_{\text{cross},2})$ are not active.

C. Achievable Exponents Region for $\mathsf{h}_1 = \mathsf{h}_2$ or $P_X = Q_X$

Define for any pmfs P_T, $P_{SU_1U_2|XT}$ and function

$$f \colon \mathcal{S} \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X} \to \mathcal{W} \tag{27}$$

the joint pmfs

$$p^i_{SU_1U_2XY_1Y_1TV_1V_2} := P_{SU_1U_2|XT} \cdot p^i_{XY_1Y_2} \cdot P_T \cdot \Gamma_{V_1V_2|SU_1U_2X}, \tag{28}$$

and

$$\Gamma_{V_1V_2|SU_1U_2X} := \Gamma_{V_1V_2|W}(v_1, v_2|f(s, u_1, u_2, x)), \qquad \forall s \in \mathcal{S}, u_1 \in \mathcal{U}_1, u_2 \in \mathcal{U}_2, x \in \mathcal{X}, \tag{29}$$

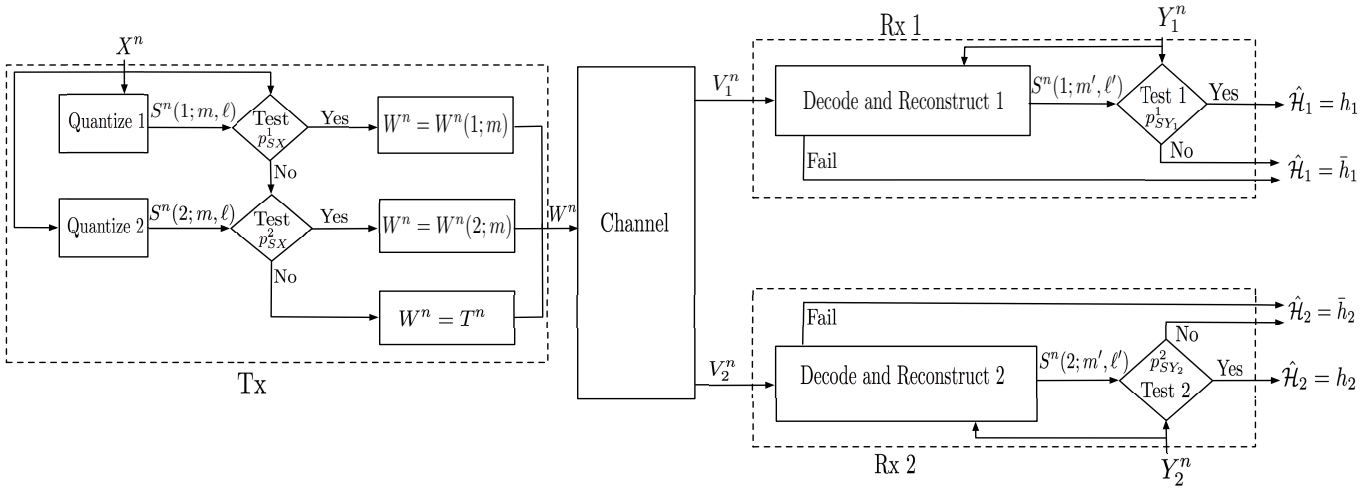

Fig. 3. Coding and testing scheme for hypothesis testing over a BC.

and for each $i \in \{1,2\}$ the four exponents

$$\theta_{\text{standard},i} := \min_{\substack{\tilde{P}_{SU_iXY_iTV_i}: \\ \tilde{P}_{SU_iXT}=p^i_{SU_iXT} \\ \tilde{P}_{SU_iY_iTV_i}=p^i_{SU_iY_iTV_i}}} D\left(\tilde{P}_{SU_iXY_iTV_i}\middle\|p^i_{SU_i|X}q^i_{XY_i}P_T\Gamma_{V_i|SU_1U_2X}\right), \tag{30}$$

$$\theta^a_{\text{dec},i} := \min_{\substack{\tilde{P}_{SU_iXY_iTV_i}: \\ \tilde{P}_{SU_iXT}=p^i_{SU_iXT} \\ \tilde{P}_{Y_iTV_i}=p^i_{Y_iTV_i} \\ H_{p^i}(S,U_i|Y_i,T,V_i)\le H_{\tilde{P}}(S,U_i|Y_i,T,V_i)}} D\left(\tilde{P}_{SU_iXY_iTV_i}\middle\|p^i_{SU_i|X}q^i_{XY_i}P_T\Gamma_{V_i|SU_1U_2X}\right) - I_{p^i}(S,U_i;X|T) + I_{p^i}(S,U_i;Y_i,V_i|T), \tag{31}$$

$$\theta^b_{\text{dec},i} := \min_{\substack{\tilde{P}_{SU_iXY_iTV_i}: \\ \tilde{P}_{SU_iXT}=p^i_{SU_iXT} \\ \tilde{P}_{SY_iTV_i}=p^i_{SY_iTV_i} \\ H_{p^i}(U_i|S,Y_i,T,V_i)\le H_{\tilde{P}}(U_i|S,Y_i,T,V_i)}} D\left(\tilde{P}_{SU_iXY_iTV_i}\middle\|p^i_{SU_i|X}q^i_{XY_i}P_T\Gamma_{V_i|SU_1U_2X}\right) - I_{p^i}(U_i;X|S,T) + I_{p^i}(U_i;Y_i,V_i|S,T), \tag{32}$$

$$\theta_{\text{miss},i} := \mathbb{E}_{P_T}\left[D\left(p^i_{Y_iV_i|T}\middle\|q^i_{Y_i}\Gamma_{V_i|W=T}\right)\right] - I_{p^i}(S,U_i;X|T) + I_{p^i}(S,U_i;Y_i,V_i|T). \tag{33}$$

.

Theorem 3: If $\mathsf{h}_1 = \mathsf{h}_2$ or $P_X = Q_X$, i.e., (13) holds, then the union of all nonnegative error exponent pairs (θ_1, θ_2) satisfying the following conditions are achievable:

$$\theta_i \le \min\left\{\theta_{\text{standard},i},\ \theta^a_{\text{dec},i},\ \theta^b_{\text{dec},i},\ \theta_{\text{miss},i}\right\}, \qquad i \in \{1,2\}, \tag{34a}$$

$$\theta_1+\theta_2 \le \min\Big\{\theta_{\text{standard},1}+\theta_{\text{standard},2},\ \ \theta_{\text{standard},1}+\theta^a_{\text{dec},2},\ \ \theta_{\text{standard},1}+\theta^b_{\text{dec},2},\ \ \theta_{\text{standard},2}+\theta^a_{\text{dec},1},\ \ \theta_{\text{standard},2}+\theta^b_{\text{dec},1},\ \ \theta_{\text{miss},1}+\theta_{\text{miss},2}\Big\} - I_{p^1}(U_1;U_2|S,T), \tag{34b}$$

$$\theta_1+\theta_2 \le \min\left\{\theta^a_{\text{dec},1},\ \theta^b_{\text{dec},1}\right\} + \min\left\{\theta^a_{\text{dec},2},\ \theta^b_{\text{dec},2}\right\} - 2I_{p^1}(U_1;U_2|S,T), \tag{34c}$$

where the union is over pmfs P_T, $P_{SU_1U_2|XT}$ and functions f as in (27) so that the pmfs (28) and (29) satisfy for $i \in \{1,2\}$:

$$I_{p^i}(S,U_i;X|T) \le I_{p^i}(S,U_i;Y_i,V_i|T), \tag{35a}$$

$$I_{p^i}(U_i;X|S,T) \le I_{p^i}(U_i;Y_i,V_i|S,T), \tag{35b}$$

$$I_{p^1}(S,U_1;X|T)+I_{p^1}(S,U_2;X|T)+I_{p^1}(U_1;U_2|S,T) \le I_{p^1}(S,U_1;Y_1,V_1|T)+I_{p^2}(S,U_2;Y_2,V_2|T), \tag{35c}$$

$$I_{p^1}(U_1;X|S,T)+I_{p^1}(U_2;X|S,T)+I_{p^1}(U_1;U_2|S,T) \le I_{p^1}(U_1;Y_1,V_1|S,T)+I_{p^2}(U_2;Y_2,V_2|S,T), \tag{35d}$$

$$I_{p^1}(U_1;X|S,T)+I_{p^1}(S,U_2;X|T)+I_{p^1}(U_1;U_2|S,T) \le I_{p^1}(U_1;Y_1,V_1|S,T)+I_{p^2}(S,U_2;Y_2,V_2|T), \tag{35e}$$

$$I_{p^1}(S,U_1;X|T)+I_{p^1}(U_2;X|S,T)+I_{p^1}(U_1;U_2|S,T) \le I_{p^1}(S,U_1;Y_1,V_1|T)+I_{p^2}(U_2;Y_2,V_2|S,T), \tag{35f}$$

Proof: The coding and testing scheme achieving these exponents is described in Section V. The analysis of the scheme is similar to the proof of [14, Theorem 4] and omitted for brevity. In particular, error exponent $\theta_{\text{standard},i}$ corresponds to the event that Receiver i decodes the correct cloud and satellite codewords but wrongly decides on $\hat{\mathcal{H}}_i = 0$. In contrast, error exponents $\theta^a_{\text{dec},i}$ and $\theta^b_{\text{dec},i}$ correspond to the events that Receiver i wrongly decides on $\hat{\mathcal{H}}_i = 0$ after wrongly decoding both the cloud center and the satellite or only the satellite. Error exponent $\theta_{\text{miss},i}$ corresponds to the miss-detection event. Because of the implicit rate-constraints in (49), the final constraints in (34) are obtained by eliminating the rates R_0, R_1, R_2 by means of Fourier-Motzkin elimination. Notice that in constraint (34c) the mutual information $I_{p^1}(U_1;U_2|S,T)$ is multiplied by a factor 2 whereas in (34c) it appears without a factor. The reason is that the error analysis includes union bounds over the codewords in a bin and when wrongly decoding the satellite codewords (which is the case of exponents $\theta^a_{\text{dec},i}$ and $\theta^b_{\text{dec},i}$) then the union bound is over *pairs* of codewords whereas under correct decoding it is over single codewords. In the former case we have the factor $2^{2nR'_i}$ in the error probability and in the latter case the factor $2^{nR'_i}$. The auxiliary rates R'_1 and R'_2 are then eliminated using the Fourier-Motzkin elimination algorithm. ■

For each $i \in \{1,2\}$, exponents $\theta_{\text{standard},i}, \theta^a_{\text{dec},i}, \theta^b_{\text{dec},i}$, and $\theta_{\text{miss},i}$ have the same form as the three exponents in [14, Theorem 1] for the DMC. There is however a tradeoff between the two exponents θ_1 and θ_2 in above theorem because they share the same choice of the auxiliary pmfs P_T and $P_{SU_1U_2|XT}$ and the function f. In [10], the above setup is studied in the special case of testing against conditional independence, and the mentioned tradeoff is illustrated through a Gaussian example. It is further proved that in some special cases, above theorem yields the optimal exponent.

IV. Coding and Testing Scheme when $p_X^1 \neq p_X^2$

Fix $\mu > 0$, a sufficiently large blocklength n, auxiliary distributions p_T, $p_{T_1|T}^1$ and $p_{T_2|T}^2$ over $\mathcal{W}$, conditional channel input distributions $p_{W|TT_1}^1$ and $p_{W|TT_2}^2$, and conditional pmfs $p_{S|X}^1$ and $p_{S|X}^2$ over a finite auxiliary alphabet $\mathcal{S}$ such that for each $i \in \{1,2\}$:

$$I_{p^i}(S;X|Y_i) < I_{p^i}(W;V_i|T,T_i). \tag{36}$$

The mutual information in (22) is calculated according to the joint distribution:

$$p_{SXY_1Y_2TT_iWV_1V_2}^i = p_{S|X}^i \cdot p_{XY_1Y_2}^i \cdot p_T \cdot p_{T_i|T}^i \cdot p_{W|TT_i}^i \cdot \Gamma_{V_1V_2|W}. \tag{37}$$

For each $i \in \{1,2\}$, if $I_{p^i}(S;X) < I_{p^i}(W;V_i|T,T_i)$, choose rates

$$R_i := I_{p^i}(S;X) + \mu, \tag{38}$$

$$R_i' := 0. \tag{39}$$

If $I_{p^i}(S;X) \geq I_{p^i}(W;V_i|T,T_i)$, then choose rates

$$R_i := I_{p^i}(W;V_i|T,T_i) - \mu, \tag{40}$$

$$R_i' := I_{p^i}(S;X) - I_{p^i}(W;V_i|T,T_i) + 2\mu. \tag{41}$$

Again, all mutual informations in (38)–(41) are calculated with respect to the pmf in (37).

Code Construction: Generate a sequence $T^n = (T_1,\ldots,T_n)$ by independently drawing each component T_k according to p_T. For each $i \in \{1,2\}$, generate a sequence $T_i^n = (T_{i,1},\ldots,T_{i,n})$ by independently drawing each $T_{i,k}$ according to $p_{T_i|T}^i(.|t)$ when $T_k = t$. Also, construct a random codebook

$$\mathcal{C}_W^i = \left\{W^n(i;m)\colon m \in \{1,...,\lfloor 2^{nR_i}\rfloor\}\right\} \tag{42}$$

superpositioned on (T^n, T_i^n) where the k-th symbol $W_k(i;m)$ of codeword $W^n(i;m)$ is drawn independently of all codeword symbols according to $p_{W|TT_i}^i(\cdot|t,t_i)$ when $T_k = t$ and $T_{i,k} = t_i$. Finally, construct a random codebook

$$\mathcal{C}_S^i = \{S^n(i;m,\ell)\colon m \in \{1,\ldots,\lfloor 2^{nR_i}\rfloor\}, \ell \in \{1,\ldots,\lfloor 2^{nR_i'}\rfloor\}\}, \quad i \in \{1,2\}, \tag{43}$$

by independently drawing the k-th component $S_k(i;m,\ell)$ of codeword $S^n(i;m,\ell)$ according to the marginal pmf p_S^i.

Reveal all codebooks and the realizations t^n, t_1^n, t_2^n of the sequences T^n, T_1^n, T_2^n to all terminals.

Transmitter: Given source sequence $X^n = x^n$, the transmitter looks for indices $(i,m,\ell) \in \{1,2\} \times \{1,\ldots,\lfloor 2^{nR_1}\rfloor\} \times \{1,\ldots,\lfloor 2^{nR_i'}\rfloor\}$ such that codeword $s^n(i;m,\ell)$ from codebook $\mathcal{C}_S^i$ satisfies

$$(s^n(i;m,\ell), x^n) \in \mathcal{T}_{\mu/2}^n(p_{SX}^i), \tag{44}$$

and the corresponding codeword $w^n(i;m)$ from codebook $\mathcal{C}_W^i$ satisfies the following:

$$(t^n, t_i^n, w^n(i;m)) \in \mathcal{T}_{\mu/2}^n(p_{TT_iW}^i). \tag{45}$$

(Notice that when μ is sufficiently small, then Condition (44) can be satisfied for at most one value $i \in \{1,2\}$, because $p_X^1 \neq p_X^2$.) If successful, the transmitter picks uniformly at random one of the triples (i,m,ℓ) that satisfy (44), and it sends the sequence $w^n(i;m)$ over the channel. If no triple satisfies Conditions (44) and (45), then the transmitter sends the sequence t^n over the channel.

Receiver $i \in \{1,2\}$: Receives v_i^n and checks whether there exist indices (m',ℓ') such that the following three conditions are satisfied:

1)

$$(t^n, t_i^n, w^n(i;m'), v_i^n) \in \mathcal{T}_\mu^n(p_{TT_iWV_i}^i), \tag{46}$$

2)

$$H_{\mathsf{tp}(s^n(i;m',\ell'),y_i^n)}(S|Y_i) = \min_{\tilde{\ell}} H_{\mathsf{tp}(s^n(i;m',\tilde{\ell}),y_i^n)}(S|Y_i), \tag{47}$$

3)

$$(s^n(i;m',\ell'), y_i^n) \in \mathcal{T}_\mu^n(p_{SY_i}^i). \tag{48}$$

If successful, it declares $\hat{\mathcal{H}}_i = \bar{\mathsf{h}}_i$. Otherwise, it declares $\hat{\mathcal{H}}_i = \mathsf{h}_i$.

Analysis: See Appendix A.

V. Coding and Testing Scheme when $p_X^1 = p_X^2$

In this case, the scheme is based on hybrid source-channel coding. Choose a large positive integer n, auxiliary alphabets $\mathcal{S}$, $\mathcal{U}_1$, and $\mathcal{U}_2$, and a function f as in (27). Choose an auxiliary distribution P_T over $\mathcal{W}$, and a conditional distribution $P_{SU_1U_2|XT}$ over $\mathcal{S}\times\mathcal{U}_1\times\mathcal{U}_2$ so that for $i\in\{1,2\}$ inequalities (35) are satisfied with strict inequality. Then, choose a positive μ and rates R_0, R_1, R_2 so that

$$R_0 = I_{p^1}(S;X|T)+\mu, \tag{49a}$$

$$R_i > I_{p^1}(U_i;X|S,T), \qquad i\in\{1,2\}, \tag{49b}$$

$$R_1+R_2 > I_{p^1}(U_1;X|S,T)+I_{p^1}(U_2;X|S,T)+I_{p^1}(U_1;U_2|S,T), \tag{49c}$$

and

$$R_0+R_i \le I_{p^i}(S,U_i;Y_i,V_i|T), \tag{49d}$$

$$R_i \le I_{p^i}(U_i;Y_i,V_i|S,T). \tag{49e}$$

Generate a sequence T^n i.i.d. according to P_T and construct a random codebook

$$\mathcal{C}_S = \left\{S^n(m_0): m_0\in\{1,...,\lfloor 2^{nR_0}\rfloor\}\right\}$$

superpositioned on T^n where each codeword is drawn independently according to $p^1_{S|T}$ conditioned on T^n. Then, for each index m_0 and $i\in\{1,2\}$, randomly generate a codebook

$$\mathcal{C}_{U_i}(m_0) = \left\{U_i^n(m_0,m_i): m_i\in\{1,...,\lfloor 2^{nR_i}\rfloor\}\right\}$$

superpositioned on $(T^n, S^n(m_0))$ by drawing each entry of the n-length codeword $U_i^n(m_0,m_i)$ i.i.d. according to the conditional pmf $p^1_{U_i|ST}(.|S_k(m_0),T)$ where $S_k(m_0)$ denotes the k-th symbol of $S^n(m_0)$. Reveal the realizations of the codebooks and the sequence T^n to all terminals.

Transmitter: Given that it observes the source sequence $X^n=x^n$, the transmitter looks for indices (m_0,m_1,m_2) that satisfy

$$(s^n(m_0),u_1^n(m_0,m_1),u_2^n(m_0,m_2),x^n,t^n)\in\mathcal{T}^n_{\mu/2}\left(p^1_{SU_1U_2XT}\right). \tag{50}$$

If successful, it picks one of these indices uniformly at random and sends the codeword w^n over the channel, where

$$w_k = f\left(s_k(m_0),u_{1,k}(m_0,m_1),u_{2,k}(m_0,m_2),x_k\right), \qquad k\in\{1,\dots,n\}, \tag{51}$$

and where $(s_k(m_0),u_{1,k}(m_0,m_1),u_{2,k}(m_0,m_2))$ denote the k-th components of codewords $(s^n(m_0),u_1^n(m_0,m_1),u_2^n(m_0,m_2))$. Otherwise, it sends the sequence of inputs t^n over the channel.

Receiver $i\in\{1,2\}$: After observing $V_i^n=v_i^n$ and $Y_i^n=y_i^n$, Receiver $i\in\{1,2\}$ looks for indices $m_0'\in\{1,\dots,\lfloor 2^{nR_0}\rfloor\}$ and $m_i'\in\{1,\dots,\lfloor 2^{nR_i}\rfloor\}$ that satisfy the following conditions:

1)

$$(s^n(m_0'),u_i^n(m_0',m_i'),y_i^n,t^n,v_i^n)\in\mathcal{T}^n_\mu(p^i_{SU_iY_iTV_i}). \tag{52}$$

2)

$$H_{\mathsf{tp}\left(s^n(m_0'),u_i^n(m_0',m_i'),y_i^n,t^n,v_i^n\right)}(S,U_i|Y_i,T,V_i) = \min_{\tilde{m}_0,\tilde{m}_i} H_{\mathsf{tp}\left(s^n(\tilde{m}_0),u_i^n(\tilde{m}_0,\tilde{m}_i),y_i^n,t^n,v_i^n\right)}(S,U_i|Y_i,T,V_i), \tag{53}$$

If successful, Receiver i declares $\hat{\mathcal{H}}_i=\bar{\mathsf{h}}_i$. Otherwise, it declares $\hat{\mathcal{H}}_i=\mathsf{h}_i$.

Analysis: Similar to [14, Appendix D] and omitted.

VI. Summary and Conclusions

The paper studies general distributed hypothesis testing over memoryless broadcast channels. A similar phenomenon is observed as for setups with common noisefree communication links from the sensor to all decision centers: while a tradeoff arises when the transmitter cannot distinguish the alternative hypotheses at the two decision centers, such a tradeoff can almost completely be mitigated when such a distinction is possible. In contrast to the noise-free link scenario, under a noisy broadcast channel model, a tradeoff can still arise because receivers can confuse to whom the communication is dedicated. In this paper, we proposed and analyzed general distributed hypothesis testing schemes both for the case where the sensor can distinguish the two null hypotheses and where it cannot not. Our general schemes recover all previously studied special cases.

Interesting directions for future research include information-theoretic converse results and extensions to multiple sensors or more than two decision centers.

APPENDIX A
PROOF OF THEOREM 2

The proof is based on the scheme of Section IV. Fix a choice of blocklength n, the small positive μ and the (conditional) pmfs p_T, $p^1_{T_1|T}$, $p^2_{T_2|T}$, $p^1_{W|TT_1}$, $p^2_{W|TT_2}$, $p^1_{S|X}$ and $p^2_{S|X}$ so that (22) holds. Assume that $I_{p^1}(S;X) \geq I_{p^1}(W;V_1|T,T_1)$ and $I_{p^2}(S;X) \geq I_{p^2}(W;V_2|T,T_2)$ in which case R_1, R_2, R_1', R_2' are given by (40) and (41). Also, set for convenience of notation:

$$p^i_{S'}(s) = p^i_S(s), \qquad \forall s \in \mathcal{S}, \tag{54}$$

$$p^i_{W'|TT_i}(w|t,t_i) = p^i_{W|TT_i}(w|t,t_i), \qquad \forall t, t_i, w \in \mathcal{W}. \tag{55}$$

The analysis of type-I error probability is similar as in [14, Appendix A]. The main novelty is that because $p^1_X(x) \neq p^2_X(x)$ for some $x \in \mathcal{X}$, for sufficiently small values of $\mu > 0$, the source sequence cannot lie in both $\mathcal{T}_{\mu/2}(p^1_X)$ and $\mathcal{T}_{\mu/2}(p^2_X)$. Details are omitted.

Consider the type-II error probability at Receiver 1 averaged over all random codebooks. Define the following events for $i \in \{1,2\}$:

$$\mathcal{E}_{\text{Tx},i}(m,\ell)\colon \quad \{(S^n(i;m,\ell),X^n) \in \mathcal{T}^n_{\mu/2}(p^i_{SX}), \quad (T^n,T^n_i,W^n(i;m)) \in \mathcal{T}^n_{\mu/2}(p^i_{TT_iW}), \quad W^n(i;m)) \text{ is sent}\}, \tag{56}$$

$$\mathcal{E}_{\text{Rx},i}(m',\ell')\colon \quad \{(S^n(i;m',\ell'),Y^n) \in \mathcal{T}^n_{\mu}(p^i_{SY_i}), \quad (T^n,T^n_i,W^n(i;m'),V^n_i) \in \mathcal{T}^n_{\mu}(p^i_{TT_iWV_i}),$$
$$H_{\text{tp}(S^n(i;m',\ell'),Y_1^n)}(S|Y_i) = \min_{\tilde{l}} H_{\text{tp}(S^n(i;m',\tilde{\ell}),Y_i^n)}(S|Y_i)\}. \tag{57}$$

Notice that

$$\mathbb{E}_{\mathcal{C}}[\beta_{1,n}] = \Pr[\hat{\mathcal{H}}_1 = 0|\mathcal{H} = \mathsf{h}_1] \leq \Pr\left[\bigcup_{m',\ell'} \mathcal{E}_{\text{Rx},1}(m',\ell')\middle|\mathcal{H} = \mathsf{h}_1\right]. \tag{58}$$

Above probability is upper bounded as:

$$\begin{aligned}\Pr\left[\bigcup_{m',\ell'} \mathcal{E}_{\text{Rx},1}(m',\ell')\middle|\mathcal{H} = \mathsf{h}_1\right] \leq{}& \Pr\left[\left(\bigcup_{m',\ell'} \mathcal{E}_{\text{Rx},1}(m',\ell')\right) \cap \left(\bigcup_{m,\ell} \mathcal{E}_{\text{Tx},1}(m,\ell)\right)\middle|\mathcal{H} = \mathsf{h}_1\right] \\ &+ \Pr\left[\left(\bigcup_{m',\ell'} \mathcal{E}_{\text{Rx},1}(m',\ell')\right) \cap \left(\bigcap_{m,\ell} \mathcal{E}^c_{\text{Tx},1}(m,\ell)\right) \cap \left(\bigcup_{m,\ell} \mathcal{E}_{\text{Tx},2}(m,\ell)\right)\middle|\mathcal{H} = \mathsf{h}_1\right] \\ &+ \Pr\left[\left(\bigcup_{m',\ell'} \mathcal{E}_{\text{Rx},1}(m',\ell')\right) \cap \left(\bigcap_{m,\ell} \mathcal{E}^c_{\text{Tx},1}(m,\ell)\right) \cap \left(\bigcap_{m,\ell} \mathcal{E}^c_{\text{Tx},2}(m,\ell)\right)\middle|\mathcal{H} = \mathsf{h}_1\right].\end{aligned} \tag{59}$$

The sum of above probabilities can be upper bounded by the sum of the probabilities of the following events:

$$\mathcal{B}_1\colon \{\exists(m,\ell) \quad \text{s.t.} \quad (\mathcal{E}_{\text{Tx},1}(m,\ell) \quad \text{and} \quad \mathcal{E}_{\text{Rx},1}(m,\ell))\}, \tag{60}$$

$$\mathcal{B}_2\colon \{\exists(m,\ell,\ell') \quad \text{with} \quad \ell \neq \ell' \quad \text{s.t.} \quad (\mathcal{E}_{\text{Tx},1}(m,\ell) \quad \text{and} \quad \mathcal{E}_{\text{Rx},1}(m,\ell'))\}, \tag{61}$$

$$\mathcal{B}_3\colon \{\exists(m,m',\ell,\ell') \quad \text{with} \quad \ell \neq \ell' \quad \text{and} \quad m \neq m' \quad \text{s.t.} \quad (\mathcal{E}_{\text{Tx},1}(m,\ell) \quad \text{and} \quad \mathcal{E}_{\text{Rx},1}(m',\ell'))\}, \tag{62}$$

$$\mathcal{B}_4\colon \{\forall(m,\ell) \quad \mathcal{E}^c_{\text{Tx},1}(m,\ell)\} \cap \{\exists(m,m',\ell,\ell') \quad \text{s.t.} \quad \mathcal{E}_{\text{Tx},2}(m,\ell) \cap \mathcal{E}_{\text{Rx},1}(m',\ell')\}, \tag{63}$$

$$\mathcal{B}_5\colon \{\forall(m,\ell) \quad \mathcal{E}^c_{\text{Tx},1}(m,\ell) \quad \text{and} \quad \mathcal{E}^c_{\text{Tx},2}(m,\ell)\} \cap \{\exists(m',\ell') \quad \text{s.t.} \quad \mathcal{E}_{\text{Rx},1}(m',\ell')\}. \tag{64}$$

Thus, we have

$$\mathbb{E}_{\mathcal{C}}\left[\beta_{1,n}\right] \leq \sum_{i=1}^{5} \Pr\left[\mathcal{B}_i\middle|\mathcal{H} = \mathsf{h}_1\right]. \tag{65}$$

The probabilities of events $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ and $\mathcal{B}_5$ can be bounded following similar steps to [14, Appendix A]. This yields:

$$\Pr\left[\mathcal{B}_1\middle|\mathcal{H} = \mathsf{h}_1\right] \leq 2^{-n(\theta_{\mu,\text{standard},1} - \delta_1(\mu))}, \tag{66}$$

$$\Pr\left[\mathcal{B}_2\middle|\mathcal{H} = \mathsf{h}_1\right] \leq 2^{-n(\theta_{\mu,\text{dec},1} - \delta_2(\mu))}, \tag{67}$$

$$\Pr\left[\mathcal{B}_3\middle|\mathcal{H} = \mathsf{h}_1\right] \leq 2^{-n\left(\theta_{\mu,\text{dec},1} - \delta_2'(\mu)\right)}, \tag{68}$$

$$\Pr\left[\mathcal{B}_5\middle|\mathcal{H} = \mathsf{h}_1\right] \leq 2^{-n(\theta_{\mu,\text{miss},1} - \delta_4(\mu))}, \tag{69}$$

for some functions $\delta_1(\mu)$, $\delta_2(\mu)$, $\delta_2'(\mu)$ and $\delta_4(\mu)$ that go to zero as n goes to infinity and $\mu \to 0$, and where we define:

$$\theta_{\text{standard},i} := \min_{\substack{\tilde{P}_{SXY_i}\colon \\ |\pi_{SX} - p^i_{SX}| < \mu/2 \\ |\pi_{SY_i} - p^i_{SY_i}| < \mu}} D(\pi_{SXY_i} \| p^i_{S|X} q^i_{XY_i}), \tag{70}$$

$$\theta_{\text{dec},i} := \min_{\substack{\tilde{P}_{SXY_i}: \\ |\pi_{SX}-p^i_{SX}|<\mu/2 \\ |\pi_{Y_i}-p^i_{Y_i}|<\mu \\ H_{p^i}(S|Y_i)\le H_\pi(S|Y_i)}} D(\pi_{SXY_i}\|p^i_{S|X}q^i_{XY_i}) - I_{p^i}(S;X|Y_i) + I_{p^i}(W;V_i|T,T_i), \tag{71}$$

$$\theta_{\text{miss},i} := D(p^i_{Y_i}\|q^i_{Y_i}) + \mathbb{E}_{p_T}\left[D\big(p^i_{V_i|T}\|\Gamma_{V_i|W=T}\big)\right] - I_{p^i}(S;X|Y_i) + I_{p^i}(W;V_i|T,T_i). \tag{72}$$

Consider event $\mathcal{B}_4$:

$$\begin{aligned}
&\Pr\left[\mathcal{B}_4|\mathcal{H}=\mathsf{h}_1\right] \\
&\le \sum_{m,\ell}\sum_{m',\ell'} \Pr\Big[(S^n(2;m,\ell),X^n)\in\mathcal{T}^n_{\mu/2}(p^2_{SX}),\ \ (T^n,W^n(2;m))\in\mathcal{T}^n_{\mu/2}(p^2_{TW}),\ \ W^n(2;m)\text{ is sent}, \\
&\qquad (S^n(1;m',\ell'),Y_1^n)\in\mathcal{T}^n_\mu(p^1_{SY_1}),\ \ (T^n,T_1^n,W^n(1;m'),V_1^n)\in\mathcal{T}^n_\mu(p^1_{TT_1WV_1}) \\
&\qquad H_{\text{tp}(S^n(1;m',\ell'),Y_1^n)}(S|Y_1)=\min_{\tilde{\ell}} H_{\text{tp}(S^n(1;m',\tilde{\ell}),Y_1^n)}(S|Y_1)\ \ \Big|\mathcal{H}=\mathsf{h}_1\Big] \\
&\overset{(a)}{\le} \sum_{m,\ell}\sum_{m',\ell'} \Pr\Big[(S^n(2;m,\ell),X^n)\in\mathcal{T}^n_{\mu/2}(p^2_{SX}),\ \ (S^n(1;m',\ell'),Y_1^n)\in\mathcal{T}^n_\mu(p^1_{SY_1}), \\
&\qquad H_{\text{tp}(S^n(1;m',\ell'),Y_1^n)}(S|Y_1)=\min_{\tilde{\ell}} H_{\text{tp}(S^n(1;m',\tilde{\ell}),Y_1^n)}(S|Y_1)\big|\mathcal{H}=\mathsf{h}_1\Big] \\
&\qquad \cdot\Pr\Big[(T^n,T_1^n,W^n(1;m'),V_1^n)\in\mathcal{T}^n_\mu(p^1_{TT_1WV_1}),\ (T^n,W^n(2;m))\in\mathcal{T}^n_{\mu/2}(p^2_{TW})\big| \\
&\qquad\qquad W^n(2;m)\text{ is sent},\ \mathcal{H}=\mathsf{h}_1\Big] \\
&\overset{(b)}{\le} 2^{n(R_1+R_1'+R_2+R_2')}\cdot \max_{\substack{\pi_{SS'XY_1}: \\ |\pi_{SX}-p^2_{SX}|<\mu/2 \\ |\pi_{S'Y_1}-p^1_{SY_1}|<\mu \\ H_\pi(S'|Y_1)\le H_\pi(S|Y_1)}} 2^{-n\left(D\left(\pi_{SS'XY_1}\|p^2_Sp^1_{S'}q^1_{XY_1}\right)-\mu\right)} \\
&\qquad\cdot \max_{\substack{\pi_{TT_1W'WV_1}: \\ |\pi_{TW}-p^2_{TW}|<\mu/2 \\ |\pi_{TT_1W'V_1}-p^1_{TT_1WV_1}|<\mu}} 2^{-n\left(D\left(\pi_{TT_1W'WV_1}\|p_Tp^1_{T_1|T}p^1_{W'|TT_1}p^2_{W|T}\Gamma_{V_1|W}\right)-\mu\right)},
\end{aligned} \tag{73}$$

where (a) holds because the channel code is drawn independently of the source code and (b) holds by Sanov's theorem.

Define

$$\begin{aligned}
\tilde{\theta}_{\mu,\text{cross},1} := &\min_{\substack{\pi_{SS'XY_1}: \\ |\pi_{SX}-p^2_{SX}|<\mu/2 \\ |\pi_{S'Y_1}-p^1_{SY_1}|<\mu \\ H_\pi(S'|Y_1)\le H_\pi(S|Y_1)}} D\left(\pi_{SS'XY_1}\|p^2_Sp^1_{S'}q^1_{XY_1}\right) \\
&+ \min_{\substack{\pi_{TT_1W'WV_1}: \\ |\pi_{TW}-p^2_{TW}|<\mu/2 \\ |\pi_{TT_1W'V_1}-p^1_{TT_1WV_1}|<\mu}} D\left(\pi_{TT_1W'WV_1}\|p_Tp^1_{T_1|T}p^1_{W'|TT_1}p^2_{W|T}\Gamma_{V_1|W}\right) - R_1 - R_2 - R_1' - R_2' - 2\mu,
\end{aligned} \tag{74}$$

and notice that

$$\begin{aligned}
\tilde{\theta}_{\mu,\text{cross},1} \overset{((40)\&(41))}{=} &\min_{\substack{\pi_{SS'XY_1}: \\ |\pi_{SX}-p^2_{SX}|<\mu/2 \\ |\pi_{S'Y_1}-p^1_{SY_1}|<\mu \\ H_\pi(S'|Y_1)\le H_\pi(S|Y_1)}} D\left(\pi_{SS'XY_1}\|p^2_Sp^1_{S'}q^1_{XY_1}\right) \\
&+ \min_{\substack{\pi_{TT_1W'WV_1}: \\ |\pi_{TW}-p^2_{TW}|<\mu \\ |\pi_{TT_1W'V_1}-p^1_{TT_1WV_1}|<\mu}} D\left(\pi_{TT_1W'WV_1}\|p_Tp^1_{T_1|T}p^1_{W'|TT_1}p^2_{W|T}\Gamma_{V_1|W}\right) - I_{p^1}(S;X) - I_{p^2}(S;X) - 4\mu
\end{aligned}$$

$$
\begin{aligned}
&\overset{(c)}{=} \min_{\substack{\pi_{SS'XY_1}:\\ |\pi_{SX}-q^1_{SX}|<\mu/2\\ |\pi_{S'Y_1}-p^1_{SY_1}|<\mu\\ H_\pi(S'|Y_1)\le H_\pi(S|Y_1)}} D\left(\pi_{SS'XY_1}\|q^1_S p^1_{S'} q^1_{XY_1}\right)\\
&\quad + \min_{\substack{\pi_{TT_1W'WV_1}:\\ |\pi_{TW}-q^1_{TW}|<\mu\\ |\pi_{TT_1W'V_1}-p^1_{TT_1WV_1}|<\mu}} D\left(\pi_{TT_1W'WV_1}\|p_T p^1_{T_1|T} p^1_{W'|TT_1} q^1_{W|T}\Gamma_{V_1|W}\right) - I_{p^1}(S;X) - I_{q^1}(S;X) - 4\mu\\
&\overset{(CR)}{=} \min_{\substack{\pi_{SS'XY_1}:\\ |\pi_{SX}-q^1_{SX}|<\mu/2\\ |\pi_{S'Y_1}-p^1_{SY_1}|<\mu\\ H_\pi(S'|Y_1)\le H_\pi(S|Y_1)}} \left[D\left(\pi_{SXY_1}\|q^1_{S|X}q^1_{XY_1}\right) + \mathbb{E}_{\pi_{SXY_1}}\left[D(\pi_{S'|SXY_1}\|p^1_{S'})\right]\right] - I_{p^1}(S;X)\\
&\quad + \min_{\substack{\pi_{TT_1W'WV_1}:\\ |\pi_{TW}-q^1_{TW}|<\mu\\ |\pi_{TT_1W'V_1}-p^1_{TT_1WV_1}|<\mu}} \Big[D(\pi_{TT_1W'W}\|p^1_{TT_1}p^1_{W'|TT_1}q^1_{W|T})\\
&\qquad\qquad + \mathbb{E}_{TT_1W'W}\left[D(\pi_{V_1|TT_1W'W}\|\pi_{V_1|TT_1}) + D(\pi_{V_1|TT_1}\|\Gamma_{V_1|W})\right]\Big] - 4\mu\\
&\overset{(DP)}{\ge} \min_{\substack{\pi_{SS'XY_1}:\\ |\pi_{SX}-q^1_{SX}|<\mu/2\\ |\pi_{S'Y_1}-p^1_{SY_1}|<\mu\\ H_\pi(S'|Y_1)\le H_\pi(S|Y_1)}} \left[D\left(\pi_{SXY_1}\|q^1_{S|X}q^1_{XY_1}\right) + \mathbb{E}_{\pi_{Y_1}}\left[D(\pi_{S'|Y_1}\|p^1_{S'})\right]\right] - I_{p^1}(S;X)\\
&\quad + \min_{\substack{\pi_{TT_1W'WV_1}:\\ |\pi_{TW}-q^1_{TW}|<\mu\\ |\pi_{TT_1W'V_1}-p^1_{TT_1WV_1}|<\mu}} \Big[\mathbb{E}_{\pi_{TT_1W'}}\left[D(\pi_{V_1|TT_1W'}\|\pi_{V_1|TT_1}) + D(\pi_{V_1|TT_1}\|\Gamma_{V_1|W})\right] - 4\mu\\
&\overset{(d)}{=} \min_{\substack{\pi_{SXY_1}:\\ |\pi_{Y_1}-p^1_{Y_1}|<\mu\\ H_{p^1}(S|Y_1)\le H_\pi(S|Y_1)}} \mathbb{E}_{q^1_{XS}}\left[D\left(\pi_{Y_1|XS}\|q^1_{Y_1|X}\right)\right] + I_{p^1}(S;Y_1) - I_{p^1}(S;X)\\
&\quad + I_{p^1}(V_1;W|T,T_1) + \min_{\substack{\pi_{TT_1WV_1}:\\ |\pi_{TW}-q^1_{TW}|<\mu\\ |\pi_{TT_1V_1}-p^1_{TT_1V_1}|<\mu}} \mathbb{E}_{\pi_{TT_1W}}\left[D(p^1_{V_1|TT_1}\|\Gamma_{V_1|W})\right] - \delta_3(\mu)\\
&= \theta_{\mu,\text{cross},1} - \delta_3(\mu) \qquad (75)
\end{aligned}
$$

for a function $\delta_3(\mu)$ that goes to zero as $\mu \to 0$ and

$$
\begin{aligned}
\theta_{\mu,\text{cross},1} := &\min_{\substack{\pi_{SXY_1}:\\ |\pi_{Y_1}-p^1_{Y_1}|<\mu\\ H_{p^1}(S|Y_1)\le H_\pi(S|Y_1)}} \mathbb{E}_{q^1_{XS}}\left[D\left(\pi_{Y_1|XS}\|q^1_{Y_1|X}\right)\right] + I_{p^1}(S;Y_1) - I_{p^1}(S;X)\\
&+ I_{p^1}(V_1;W|T,T_1) + \min_{\substack{\pi_{TT_1WV_1}:\\ |\pi_{TW}-q^1_{TW}|<\mu\\ |\pi_{TT_1V_1}-p^1_{TT_1V_1}|<\mu}} \mathbb{E}_{\pi_{TT_1W}}\left[D(p^1_{V_1|TT_1}\|\Gamma_{V_1|W})\right]. \qquad (76)
\end{aligned}
$$

Here (c) holds because the condition $p^1_X \neq p^2_X$ implies that $\mathsf{h}_1 = \bar{\mathsf{h}}_2$ and thus $p^2 = q^1$, and (d) holds by the constraints in the minimizations.

Combining (73), (74) and (75), establishes:

$$\Pr\left[\mathcal{B}_4 \middle| \mathcal{H} = \mathsf{h}_1\right] \leq 2^{-n(\theta_{\mu,\text{cross},1} - \delta_3(\mu))}. \tag{77}$$

Considering (65)–(69) and (77), we get:

$$\mathbb{E}_{\mathcal{C}}\left[\beta_{1,n}\right] \leq \max\{2^{-n(\theta_{\mu,\text{standard},1} - \delta_1(\mu))}, 2^{-n(\theta_{\mu,\text{dec},1} - \delta_2(\mu))}, 2^{-n\left(\theta_{\mu,\text{dec},1} - \delta_2'(\mu)\right)}, 2^{-n(\theta_{\mu,\text{cross},1} - \delta_3(\mu))}, 2^{-n(\theta_{\mu,\text{miss},1} - \delta_4(\mu))}\}. \tag{78}$$

By standard arguments and successively eliminating the worst half of the codewords with respect to $\alpha_{1,n}$ and the exponents $\theta_{\mu,\text{standard},1}$, $\theta_{\mu,\text{dec},1}$, $\theta_{\mu,\text{cross},1}$ and $\theta_{\mu,\text{miss},1}$, it can be shown that there exists at least one codebook for which

$$\alpha_{1,n} < \epsilon, \tag{79}$$

$$\beta_{1,n} \leq 32 \cdot \max\{2^{-n(\theta_{\mu,\text{standard},1} - \delta_1(\mu))}, 2^{-n(\theta_{\mu,\text{dec},1} - \delta_2(\mu))}, 2^{-n\left(\theta_{\mu,\text{dec},1} - \delta_2'(\mu)\right)}, 2^{-n(\theta_{\mu,\text{cross},1} - \delta_3(\mu))}, 2^{-n(\theta_{\mu,\text{miss},1} - \delta_4(\mu))}\}. \tag{80}$$

Letting $\mu \to 0$ and $n \to \infty$, we get $\theta_{\mu,\text{standard},1} \to \theta_{\text{standard},1}$, $\theta_{\mu,\text{dec},1} \to \theta_{\text{dec},1}$, $\theta_{\mu,\text{cross},1} \to \theta_{\text{cross},1}$ and $\theta_{\mu,\text{miss},1} \to \theta_{\text{miss},1}$. A similar bound can be found for θ_2. This concludes the proof.

REFERENCES

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

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

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

[4] H. Shimokawa, *Hypothesis testing with multiterminal data compression*. Master's Thesis, 1994-.

[5] N. Weinberger and Y. Kochman, "On the reliability function of distributed hypothesis testing under optimal detection," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 4940–4965, 2019.

[6] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6282–6303, Oct. 2012.

[7] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *Proc. 52nd Allerton Conf. Comm, Cont. and Comp.*, Monticello, IL, USA, Oct. 2014, pp. 1246–1251.

[8] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing against independence," in *Proc. IEEE Int. Symp. on Info. Theory*, Istanbul, Turkey, Jun. 2013, pp. 2840–2844.

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

[10] S. Salehkalaibar, M. Wigger, and R. Timo, "On hypothesis testing against independence with multiple decision centers," *IEEE Trans. on Communications*, Jan. 2018.

[11] 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, 2019.

[12] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detection," in *Proc. IEEE Int. Symp. on Info. Theory*, Jun. 2018.

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

[14] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal error protection codes," *IEEE Trans. on Info. Theory*, vol. 66, no. 7, pp. 4150–4182, 2020.

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

[16] S. Borade, B. Nakiboglu, and L. Zheng, "Unequal error protection: an information-theoretic perspective," *IEEE Trans. on Info. Theory*, vol. 55, pp. 5511–5539, Nov. 2009.

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